



**UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS**

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS  
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS  
CENTRO DE INVESTIGACIÓN EN CIENCIAS

**“Modelado del transporte de gases en sistemas  
geotérmicos usando herramientas de inteligencia  
artificial”**

**TESIS**

QUE PARA OBTENER EL GRADO DE

**DOCTOR EN CIENCIAS**

PRESENTA

**Mtro. Agustín Acevedo Anicasio**

**DIRECTORES DE TESIS**

**Dr. Edgar R. Santoyo Gutiérrez y**

**Dra. Lorena Díaz González**

**DR. JEAN MICHEL GRÉVY MACQUART  
COORDINADOR DEL POSGRADO EN CIENCIAS  
PRESENTE**

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la TESIS titulada **“Modelado del transporte de gases en sistemas geotérmicos usando herramientas de inteligencia artificial”** que presenta el alumno **Agustin Acevedo Anicasio (9020140108)** para obtener el título de **Doctor en Ciencias**.

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dr. Nodari Vakhania CInC-UAEM	APROBADO	
Dr. Minh Huy Hô Nguyen CIQ-UAEM	APROBADO	
Dr. Juan Manuel Rendón Mancha CInC-UAEM	APROBADO	
Dr. José Alberto Hernández Aguilar FCAel-UAEM	APROBADO	
Dr. Daniel Pérez Zárate Instituto de Geofísica-UNAM	APROBADO	
Dr. Edgar Rolando Santoyo Gutiérrez IER-UNAM	APROBADO	
Dr. Outmane Oubram FCQel-UAEM	APROBADO	



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

### Sello electrónico

**OUTMANE OUBRAM | Fecha:2021-08-16 17:14:43 | Firmante**

rDjCBfjra3zbtHtPoo9xttR+/CGaohillZvbqoVBPXnmTF9EGEF446/7xrddXPRskfj1LYgAeNk7SAB7Ktvpma8Xq/dKdG3Nr0FJIAJHktsriUJHjKOAEYI7A9umWP1pX+GcRbP1bjm54spMqX326wUb7kx6UOheWD+FectlsnsEtjYk/uVTJzUoPuCwBSc/48mp1rHlqHLPJ88P9JuwYZV8W765YK8vpDTodsh27e/O71sz7Tgh9ncF5zYnQJYkec9u97ZZ4Wrk3VRaKlZxy7WC1dFBp+fXi17cpDbZnkHHw/8qAZeLol+LUzakgQw826ZlbqkHhGwCmgx7Jw==

**MINHUY HO NGUYEN | Fecha:2021-08-16 17:57:38 | Firmante**

ZrvK4t2zVhUInCoVwXrWjYHQfQm9x8hG0Dlbpw//8nOliSo45AaNBWh7DkEZHD95wZBGMOkFI9ssiYwDQLzi9e6yFUwRxbYDIoAa1r85Y6tRNxwRGz9DtsTdW6q6P9e+y9+UHyxugetfj8kYjozCUUsN+tnQR0pkYrXV8uRde5sYI5ppGn1+8ArMk5iRH95I3T14HDhikGM2VlvRAR2vcxR513ZBTIIYRBDmlMRomojWog8HWt9PpFk4XQbtaC9eSFHhLRelC0lq1DEhepx4RXQD8NrfmRog5NpwwGwM1C3Z5KfS61b0PI2X4u7M1w07KIWA3/SmQyXIOoQ==

**EDGAR ROLANDO SANTOYO GUTIÉRREZ | Fecha:2021-08-16 18:02:13 | Firmante**

DnDuiCq5IZxGMAuLne4r2+0u/8XAILzE/QcNL1/WBy3fx2fw6//sedwDezlPhlz2e9fi+l2d9jPloVky62k9g3ABaYyl5Fel3X8sPVEIEVMG9ZAU4/GeLaitOLITmEhXytukt463qB/OM/1FQZD5MzUhc0JIA5sa51bfTky3qc3fCVmoFm2/YSAPEz+SxxavxpW4cCHmlshos+fPxYJKyo5EANEykyolGMYTzJbAspouYx1vBzG1ZMPe4yH2cUcWVSmXV92PrPC4ByvibOI93hPnv7BBJbnUs8BkdO8oMBpyEizlfqQPRtFYKRYcKYA6QZ69ySaYxq/twX7iahkQ==

**JUAN MANUEL RENDON MANCHA | Fecha:2021-08-17 01:49:38 | Firmante**

ZRdvgY1Pr8/IAfhwt+AhmFdxw+Wlfxyl33I32zZyBsqVI+Xgk5C+FTABcvUvqfqtMZCVHRSgW+EUen68yTqyOL3Y5F/6aqcF154AjmWmMCOonTZRzV4+IPfO1w5CYM8CK3leWjE38gpjLof/3Mhj8a09GR2np1bylEY60ExMS1g5JKe9TsvzegCRHG/1xcp7ZxjGwcmY8iUFzpa/nh0RHPTfBnzOlmocpvnvP+5LGrzbxzK/L15sbS1N4m/+BKHfVRvA83bZEGPD2y0wvpj1ID1KQVQHg6vMjC9uBkPQgWonpX3cOZ0pq04NwmQzmZDXt8fS5g5SKk+hN/WA==

**DANIEL PÉREZ ZARATE | Fecha:2021-08-17 10:08:55 | Firmante**

GVgy2vjGsM+AJ0icB0t4dLUkeOMjHeld1qlRV45C22tUiZHolzTCXHL0IzP6douxelSmgR8xLJRJmGS8AXFwFi0gzflBWMuF+sF90m101WlystZ1Fj/ocgjifLYI752xowvieU8vu9Hc h7LJ+ptfEPOYfyJYV/VH0M1TFD8YCM6643ffpRHuPxyGLYBYIWRyCblUtpJHtbqwr8K07k5i6fy/OS2ylk2Zm9M/BxFOpuW4wKLOT51WZ8XBI3KNSXdBSJOSvgu22lev70BIK8ZgY7Okc8QUGt3+gnVLFzp1Ik6nrwpWpoyUMNO8UAGS0RpylgOJi3IoNMjAAJprFIQ==

**JOSE ALBERTO HERNANDEZ AGUILAR | Fecha:2021-08-18 11:49:56 | Firmante**

cvFyxV+IGZuT2S5njTjNqkHszs7oQ9V+XZ1J/RZpdIQWMDgnAsyNmlipwyF7ma1jtZVaq+IVT5Ji9G0CPw47TGNsUL0lvzbYwf5Go3ua0rMAvbB/0Sb9f2NxxJijwUY2UbBbr6y5GQ9UGtUAlg0Qk6W5/PUElv6U9DsvROY09yMP8Pkibe6dCAdY7Rq4saRDH5GzNqjNz+GVTVK4SPJ7ZK/+jfdp9QIRIURR6ONeoevmEwpPNhKA711H71FkxRT3u/Hck2DGBPvCaZL+ImOB3Ht9CDT8x8CxoY990jqACV9lwh044yOj3EfGwDzn3xqTkGfDQUngyOZzc5f/Grd7Q==

**NODARI VAKHANIA MAISURADZE | Fecha:2021-08-26 19:15:41 | Firmante**

p5IRXylk6ff8rfMpuab/WQVIQWnWtS1PhE5vt5Jx58y2xyV+0X/GQhArhvcyXC4xoOOVJS6oj2ky4Ju5++AwT86uvygl4CBPh1scHLzMHwR0taS6Q6AeHBqS99NOlcmrVhrVwbnBuD7DgEXVschq4ZHozPX+7Qbh8NYs29SmpVnKWJOehZan54cdAtvHEXnYK8Mxu6PEJjKoWs2fAfjtt+FmShY1Kih9RfuO1P4lCrFngJnxKEUJowxeV5GD6XUrka6SPWvbWPSyT1suypdforHwT895XghJ3nyVEMXUEImcJWV1E+RagoGJ9Kxt88jwJuneaTTsCQI3+ZbMzDA==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



3mg5rC

<https://efirma.uaem.mx/noRepudio/LfLVD1bUKU2md77GhGuxUG7r2pyyy5MI>



## LISTA DEL JURADO

<b>JURADO</b>	<b>NOMBRE</b>	<b>ADSCRIPCIÓN</b>
Presidente	Dr. Nodari Vakhania	CInC-UAEM
Secretario	Dr. Minhuy Hô	CIQ-UAEM
Vocal	Dr. Juan Manuel Rendón Mancha	CInC-UAEM
Vocal	Dr. José Alberto Hernández Aguilar	FCAeI-UAEM
Vocal	Dr. Daniel Pérez Zárata	Instituto de Geofísica-UNAM
Suplente	Dr. Edgar Rolando Santoyo Gutiérrez	IER-UNAM
Suplente	Dr. Outmane Oubram	FCQeI-UAEM

Modelado del transporte de gases en sistemas geotérmicos usando herramientas de inteligencia artificial

Tesis presentada en cumplimiento de un título de PhD en Ciencias

Copyright © 2021 Agustin Acevedo-Anicasio  
Todos los derechos reservados

Centro de Investigación en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas,  
Universidad Autónoma del Estado de Morelos  
Av. Universidad No. 1001, Col. Chamilpa C.P. 62209  
Cuernavaca, Morelos  
México  
Teléfono: +52 777 329 7900

Para citar:

A. Acevedo-Anicasio. (2021). Modelado del transporte de gases en sistemas geotérmicos usando herramientas de inteligencia artificial. Doctor en Ciencias, Centro de Investigación en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, 99 pp.

ORCID de autor: <https://orcid.org/0000-0003-2115-3754>

Google Scholar de autor:

[https://scholar.google.com.mx/citations?view\\_op=search\\_authors&mauthors=acevedo-anicasio&hl=es&oi=ao](https://scholar.google.com.mx/citations?view_op=search_authors&mauthors=acevedo-anicasio&hl=es&oi=ao)

Cuernavaca, Mor. México, septiembre 2021

## PUBLICACIONES

### Manuscritos publicados

#### ANEXO 1.

Pérez-Zárate, D., Santoyo, E., Acevedo-Anicasio, A., Díaz-González, L., García-López, C. (2019). Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. *Computers and Geosciences*, vol. 129, 49-68.

#### ANEXO 2.

Acevedo-Anicasio, A., Santoyo, E., Pérez-Zárate, D., Pandarinath, K., Guevara, M., Díaz-González, L. (2021). GaS\_GeoT: A computer program for an effective use of newly improved gas geothermometers in predicting reliable geothermal reservoir temperatures. *Geotherm Energy*, vol. 9, no. 1, 1-41.

#### ANEXO 3.

Santoyo, E., Acevedo-Anicasio, A., Pérez-Zarate, D., Guevara, M. (2018). Evaluation of artificial neural networks and eddy covariance measurements for modelling the CO<sub>2</sub> flux dynamics in the Acoculco geothermal caldera (Mexico). *International Journal of Environmental Science and Development*, vol. 9, no. 10, 298-302.

### Manuscritos en preparación

#### MS-A

*(To be submitted in November 2021)*. Numerical modelling of the CO<sub>2</sub> flux dynamics in the Acoculco geothermal caldera (Mexico) by using optimised artificial neural networks and Eddy covariance measurements.

#### MS-B

*(To be submitted in February 2022)*. Development of a new and improved software for the determination of inflexion points and the accurate discrimination of statistical geochemical data sets using the Sinclair method.

## RESUMEN

La presente tesis se enmarca en el estudio del transporte de los gases presentes en los sistemas geotérmicos hidrotermales y de roca seca caliente. Ambos sistemas se encuentran distribuidos en el mundo, y su existencia obedece al aprovechamiento del calor que se encuentra en el interior de la Tierra y que puede ser utilizado ya sea de forma directa (manantiales, balnearios, secado agrícola, entre otros) o para la generación de electricidad.

En los sistemas hidrotermales los gases geotérmicos (p. ej. CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub>) brindan información de gran importancia para comprender las condiciones del yacimiento a profundidad, siendo la temperatura la más importante. El conocimiento de esta temperatura permite determinar el potencial geoenergético disponible, es decir, temperaturas superiores a 150°C son utilizadas para la generación de electricidad. Dado que existe una correlación entre los gases y la temperatura del yacimiento, desde los años 80's se han desarrollado diversas ecuaciones analíticas para estimar la temperatura durante la etapa exploratoria. Estas ecuaciones, mejor conocidas como geotermómetros de gases han sido calibradas con las temperaturas medidas en pozos geotérmicos productores y las especies químicas presentes. En la presente investigación, se desarrolló una nueva metodología integral que comprende el uso de redes neuronales artificiales y el análisis de decisión multi-criterio (*MCDA* por sus siglas en inglés: *Multi-Criteria Decision Analysis*). Conceptualmente las redes neuronales son un conjunto de elementos de procesamiento (neuronas) los cuales se conectan entre sí en forma de grafo dirigido con la finalidad de obtener un resultado. Están inspiradas en las neuronas biológicas las cuales reciben información, la procesan y emita una señal de salida.

Cada neurona está asociada a un valor real denominado peso sináptico ( $w_1, w_2, w_3, \dots, w_n$ ), el cual es ajustado iterativamente hasta alcanzar un resultado deseado. El algoritmo utilizado para ajustar dichos pesos sinápticos es conocido como *backpropagation*, que de forma general consta de 2 fases: FASE 1. En la propagación hacia adelante se aplica un patrón a las entradas de la red (variables de entrada) y su efecto se propaga a través de la misma en las capas siguientes, es decir, la capa oculta y de salida. Al llegar a la última capa, la red obtiene un conjunto de salidas (resultados simulados) como respuesta a dicho patrón de entrada. FASE 2. En la propagación hacia atrás los pesos de la red se modifican de

acuerdo con una regla de ajuste del error (por ejemplo, el descenso del gradiente utilizado por el algoritmo *Levenberg-Marquardt* (LM)). Es decir, se calcula el valor de la función de error que compara la salida actual de la red y la salida deseada. Este error se propaga hacia atrás y con base a las derivadas parciales entre el error total de la red y cada peso sináptico, se calcula el NUEVO peso asociado a cada neurona. La contribución que tiene cada peso sináptico para la obtención de un resultado determinado, es ajustada de tal forma que la salida de la red neuronal sea lo más cercano a los valores deseados (*targets*).

El método *MCDA* contribuyó a la evaluación integral de los resultados obtenidos a través de las redes neuronales para identificar las que presentaron mejor desempeño bajo distintos escenarios. Resaltando que, hasta nuestro conocimiento, en esta investigación es la primera vez que se utiliza el *MCDA* para la evaluación integral de arquitecturas de redes neuronales. Se obtuvieron 8 nuevos geotermómetros de gas (GasG<sub>1</sub> a GasG<sub>8</sub>) para estimar la temperatura de yacimientos geotérmicos. Para una práctica aplicación de estos geotermómetros se desarrolló un nuevo programa computacional llamado GaS\_GeoT. La capacidad de predicción de los nuevos geotermómetros fue evaluada con las estimaciones de temperatura de 25 geotermómetros de gas existentes, los cuales utilizan la fase gaseosa de los fluidos geotérmicos. Esta evaluación se realizó para yacimientos de líquido dominante (LIQDR) y vapor dominante (VAPDR) utilizando diferentes métricas estadísticas: DIFF%, RMSE, MAE, MAPE y U de Theil. De esta evaluación se resalta que los 8 nuevos geotermómetros (GasG<sub>1</sub> a GasG<sub>8</sub>) exhibieron estimaciones de temperatura más reales ocupando las primeras 8 posiciones en un ranking de eficiencia para yacimientos de LIQDR. Los geotermómetros GasG<sub>1</sub> y GasG<sub>3</sub> presentaron el mejor desempeño para yacimientos de VAPDR. El GasG<sub>3</sub> exhibió la mayor aplicabilidad para ambos sistemas geotérmicos LIQDR y VAPDR.

Por otro lado, los sistemas de roca seca caliente que no exhiben manifestaciones superficiales debido a la ausencia de un fluido en el interior del yacimiento, requieren la implementación de técnicas no convencionales para su exploración. Dentro de las más utilizadas encontramos a las cámaras de acumulación y el equipo micrometeorológico *Eddy-Covariance*. Estos equipos permiten medir el flujo de gases a nivel del suelo y atmosférico, respectivamente. Así mismo, la presente investigación paralelamente aborda la identificación de anomalías geotérmicas en dos vertientes: (i) modelar el flujo de CO<sub>2</sub> (el gas



mayoritario en ambientes geotérmicos) utilizando redes neuronales artificiales; y, (ii) discriminar su origen con base a un método geoquimiométrico usando la distribución de la probabilidad acumulada, conocido en la literatura como método de Sinclair. En la primera vertiente se desarrollaron alrededor de 16 arquitecturas diferentes para modelar el flujo de CO<sub>2</sub> con fines de complementar la ausencia de datos censados por el equipo *Eddy-Covariance* en periodos de bajas temperaturas y exceso de humedad en la zona bajo estudio. El desempeño de las redes neuronales fue evaluado con la metodología *MCDA*. La arquitectura neuronal utilizada para completar el perfil de mediciones del flujo de CO<sub>2</sub> es la ANN-7 (12-14-1) es decir, 12 variables de entrada (p. ej. la radiación neta, el calor sensible, la temperatura del aire y suelo, la humedad relativa, entre otras), 14 neuronas en la capa oculta y 1 neurona en la capa de salida. La validación estadística de los resultados obtenidos por la ANN-7 se realizó utilizando la prueba F y t-student, las cuales arrojaron que los datos estimados por la red neuronal no exhiben diferencias significativas en las varianzas o medias con respecto a los datos censados por el equipo *Eddy-Covariance*. En la segunda vertiente de desarrollo una nueva metodología computacional apoyada en el método de Sinclair para discriminar el origen del flujo de CO<sub>2</sub> medido en la superficie de la zona bajo estudio. De forma general, este método utiliza la distribución de la frecuencia acumulada para elegir visualmente el punto de inflexión; el cambio de concavidad representa la separación de 2 grupos mezclados en la gráfica de la frecuencia y posteriormente estos grupos son analizados para su interpretación a través de la probabilidad de la mezcla existente. Dado que una de las limitantes de Sinclair es la selección de un punto de inflexión visualmente, se implementó una mejora al método utilizando la prueba de rachas el cual indica un cambio en el signo de la pendiente en una función  $f(x)$ . Puesto que no se tiene una función que represente la frecuencia acumulada graficada con Sinclair, se desarrolló un algoritmo computacional para interpolar las coordenadas desconocidas en la curvatura de la frecuencia utilizando el polinomio de *Lagrange* y al mismo tiempo conocer la ubicación en los cambios de concavidad de la frecuencia acumulada bajo la prueba de rachas. Con el desarrollo de esta mejora matemática al método de Sinclair, fue posible identificar la existencia de 4 orígenes del flujo de CO<sub>2</sub> en la zona bajo estudio (12.2% Magmático, 5.7% Atmosférico, 35.1% Antropogénico y 47% Biogénico).

## ABSTRACT

This thesis is part of the study of the transport of gases present in hydrothermal geothermal systems and hot dry rock. Both systems are distributed in the world, and their existence is due to the use of the heat that is inside the Earth and that can be used either directly (springs, spas, agricultural drying, among others) or for the electricity generation.

In hydrothermal systems, geothermal gases (e.g. CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub>) provide information of great importance for understanding reservoir conditions at depth, temperature being the most important. Knowledge of this temperature allows determining the available geoenergetic potential, that is, temperatures above 150 °C are used to generate electricity. Since there is a correlation between the gases and the reservoir temperature, since the 1980s various analytical equations have been developed to estimate the temperature during the exploratory stage. These equations, better known as gas geothermometers, have been calibrated with the temperatures measured in producing geothermal wells and the chemical species present. In the present investigation, a new comprehensive methodology was developed that comprises the use of artificial neural networks and Multi-Criteria Decision Analysis (*MCDA*). Conceptually, neural networks are a set of processing elements (neurons) which are connected to each other in the form of a directed graph in order to obtain a result. They are inspired by biological neurons which receive information, process it and emit an output signal.

Each neuron is associated with a real value called synaptic weight ( $w_1, w_2, w_3, \dots, w_n$ ), which is iteratively adjusted until a desired result is reached. The algorithm used for adjust these synaptic weights is known as backpropagation, which generally consists of 2 phases: PHASE 1. In forward propagation, a pattern is applied to the network inputs (input variables) and its effect is propagated to through it in subsequent layers, that is, the hidden and output layer. Upon reaching the last layer, the network obtains a set of outputs (simulated results) in response to said input pattern. PHASE 2. In the backward propagation the weights of the network are modified according to an error adjustment rule (for example, the descent of the gradient used by the *Levenberg-Marquardt* (LM) algorithm. That is, the value is calculated of the error function that compares the current output of the network and the desired output. This error is propagated backwards and based on the partial

derivatives between the total error of the network and each synaptic weight, the NEW weight associated with each neuron. The contribution that each synaptic weight has to obtain a determined result is adjusted in such a way that the output of the neural network is as close to the desired values (targets).

The *MCDA* method contributed to the integral evaluation of the results obtained through neural networks to identify those that presented the best performance under different scenarios. Highlighting that, to our knowledge, this research is the first time that the *MCDA* is used for the comprehensive evaluation of neural network architectures. 8 new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) were obtained to estimate the temperature of geothermal reservoirs. For a practical application of these geothermometers, a new computer program called GaS\_GeoT was developed. The predictability of the new geothermometers was evaluated with the temperature estimates of 25 existing gas geothermometers, which use the gas phase of geothermal fluids. This evaluation was carried out for dominant liquid (LIQDR) and dominant vapor (VAPDR) reservoirs using different statistical metrics: DIFF%, RMSE, MAE, MAPE and U of Theil. From this evaluation, it is highlighted that the 8 new geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) exhibited more realistic temperature estimates, occupying the first 8 positions in an efficiency ranking for LIQDR reservoirs. The GasG<sub>1</sub> and GasG<sub>3</sub> geothermometers presented the best performance for VAPDR reservoirs. GasG<sub>3</sub> exhibited the highest applicability for both LIQDR and VAPDR geothermal systems.

On the other hand, hot dry rock systems that do not exhibit surface manifestations due to the absence of a fluid inside the reservoir, require the implementation of unconventional techniques for their exploration. Among the most used we find the accumulation chambers and the Eddy-Covariance micrometeorological equipment. This equipment's allow to measure the flow of gases at ground and atmospheric level, respectively. Likewise, this research simultaneously addresses the identification of geothermal anomalies in two aspects: (i) modeling the flow of CO<sub>2</sub> (the majority gas in geothermal environments) using artificial neural networks; and, (ii) discriminate its origin based on a geochemometric method using the cumulative probability distribution, known in the literature as the Sinclair method. In the first slope, around 16 different architectures were developed to model the flow of CO<sub>2</sub> in order to complement the absence of data registered by the Eddy-Covariance team in periods of low temperatures and excess humidity in

the area under study. The performance of neural networks was evaluated with the MCDA methodology. The neural architecture used to complete the CO<sub>2</sub> flux measurement profile is ANN-7 (12-14-1) that is, 12 input variables (e.g. net radiation, sensible heat, air temperature and soil, relative humidity, among others), 14 neurons in the hidden layer and 1 neuron in the output layer. The statistical validation of the results obtained by the ANN-7 was carried out using the F and t-student tests, which showed that the data estimated by the neural network did not show significant differences in the variances or means with respect to the data censored by the Eddy-Covariance team. In the second aspect of development, a new computational methodology based on the Sinclair method to discriminate the origin of the CO<sub>2</sub> flux measured on the surface of the area under study. Generally, this method uses the cumulative frequency distribution to visually choose the inflection point; the change in concavity represents the separation of 2 mixed groups in the frequency graph and these groups are subsequently analyzed for their interpretation through the probability of the existing mixture. Since one of Sinclair's limitations is the selection of a visually inflection point, an improvement to the method was implemented using the runs test which indicates a change in the sign of the slope in a function  $f(x)$ . Since there is no function that represents the cumulative frequency graphed with Sinclair, a computational algorithm was developed to interpolate the unknown coordinates in the curvature of the frequency using the Lagrange polynomial and at the same time know the location in the concavity changes of the cumulative frequency under the streak test. With the development of this mathematical improvement to the Sinclair method, it was possible to identify the existence of 4 sources of the CO<sub>2</sub> flow in the area under study (12.2% Magmatic, 5.7% Atmospheric, 35.1% Anthropogenic and 47% Biogenic).

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para la realización de mis estudios de Doctorado.

Al Proyecto P09 del Centro Mexicano de Innovación en Energía Geotérmica (CeMIE-Geo) "Desarrollo y Aplicación de Nuevos Métodos Avanzados en Geoquímica de Fluidos y Alteración Hidrotermal para la Exploración de Sistemas Geotérmicos" del Fondo de Sustentabilidad Energética CONACyT-SENER por facilitarme el acceso a datos geoquímicos utilizados para desarrollar este proyecto de investigación.

Al Posgrado en Ciencias del Centro de Investigación en Ciencias (área Modelación Computacional y Cómputo Científico) por admitirme como estudiante de tiempo completo durante mis estudios de doctorado.

Al Dr. Edgar R. Santoyo Gutiérrez, por su disposición, paciencia e innumerables asesorías, que me permitieron finalizar este proyecto con los resultados esperados. Así mismo, por creer en mí y convertirse en un modelo a seguir en el camino de la ciencia.

A la Dra. Lorena Díaz González, por continuar como mi asesora brindándome su apoyo académico, disponibilidad de tiempo y sobre todo compartirme valiosas enseñanzas y conocimientos, que guiaron mis pasos en la dirección correcta "*la búsqueda del conocimiento*".

A los Doctores Daniel Pérez Zárate, Minhuy Hô y Outmane Oubram por su aportación de conocimientos en cada semestre para mejorar mi formación académica y sustentar los resultados obtenidos durante esta investigación.

A los Doctores Dr. Nodari Vakhania, Dr. Juan Manuel Rendón Mancha y Dr. José Alberto Hernández Aguilar por su disposición y valiosos comentarios al formar parte de mi jurado durante la revisión de mi tesis y examen de grado.

## DEDICATORIA

*A mi Dios, que siempre ha estado a mi lado dándome las fuerzas que necesito para cumplir mis objetivos.*

*A mi amada esposa Laura por su apoyo incondicional y a mi hijo Julián por ser el motor que me motiva día a día.*

*A mis padres y hermanos, por creer en mí en todo momento.*

*A familiares y amigos que con unas palabras de aliento me impulsaron a dar lo mejor de mí.*

# CONTENIDO

Índice de Figuras .....	<i>i</i>
Índice de Tablas.....	<i>iii</i>

## **CAPÍTULO 1. Introducción**

1.1 Antecedentes.....	1
1.2 Planteamiento del problema y justificación .....	3
1.3 Objetivo general.....	4
1.4 Objetivos específicos .....	5
<u>Módulo 1.</u> Desarrollar nuevos geotermómetros de gases para la predicción de temperaturas de fondo .....	5
<u>Módulo 2.</u> Desarrollar nuevos modelos de predicción para modelar anomalías o emisiones naturales de CO <sub>2</sub> en sistemas geotérmicos .....	5
<u>Módulo 3.</u> Desarrollar una metodología geoquimiométrica computacional apoyada en el método de Sinclair para la discriminación del transporte de gas en sistemas geotérmicos .....	6
1.5 Preguntas de investigación relacionadas con los trabajos publicados y en preparación.....	6
1.6 Estructura de la tesis .....	7

## **CAPÍTULO 2. Importancia del Transporte de Gases en Sistemas Geotérmicos**

Resumen del Capítulo.....	9
2.1. Sistemas Geotérmicos.....	9
Sistemas hidrotermales (convectivos).....	10
Sistemas de roca seca caliente (conductivos) .....	12
2.2. Aprovechamiento actual de los recursos geotérmicos en el Mundo y en México .....	14
Usos directos .....	15
Generación de electricidad .....	16
2.3. Exploración de recursos geotérmicos – técnicas convencionales .....	17
2.4. Exploración de recursos geotérmicos – técnicas no convencionales .....	18
Método de <i>Eddy-Covariance</i> (EC) .....	19

Método de Cámara de Acumulación (CA).....	21
2.5. El transporte de gases como parámetro geoquímico para identificar anomalías geotérmicas.....	22
2.6. La temperatura como parámetro clave en la evaluación de reservas geoenergéticas.....	24
Geotermómetros de gases.....	26

### **CAPÍTULO 3. Inteligencia Artificial Aplicada en Redes Neuronales**

Resumen del Capítulo.....	28
3.1. Generalidades de la IA.....	28
Aprendizaje máquina.....	30
3.2. Redes Neuronales Artificiales (RNA).....	31
3.3. RNA Perceptrón Multicapa.....	33
Algoritmos de aprendizaje.....	35
3.4. Algoritmo <i>Backpropagation</i> .....	35
3.5. Ejemplo numérico de algoritmo <i>Backpropagation</i> .....	37
Propagación hacia adelante.....	38
Propagación hacia atrás.....	40

### **CAPÍTULO 4. Metodología y Desarrollo del Proyecto de Investigación**

Resumen del Capítulo.....	43
4.1. Metodología del <u>Módulo 1</u> .....	44
4.2. Metodología del <u>Módulo 2</u> .....	50
4.3. Metodología del <u>Módulo 3</u> .....	52
Antecedentes.....	53
Método de Sinclair.....	54
Desarrollo.....	56
Identificación del punto de inflexión.....	60
4.4. Pseudocódigo de los Módulos 1, 2 y 3.....	63

### **CAPÍTULO 5. Resultados del Proyecto de Investigación**

Resumen del Capítulo.....	66
5.1. Resultados del <u>Módulo 1</u> .....	66
Ecuación general geotermométrica.....	68



Programa GaS_GeoT con nuevos y mejorados geotermómetros de gases...	68
5.2. Resultados del <u>Módulo 2</u> .....	70
<i>MCDA</i> para evaluar modelos neuronales .....	72
Resultados del <i>MCDA</i> – Ecuación neuronal .....	77
Mediciones del flujo de CO <sub>2</sub> .....	79
Validación estadística – Prueba F y t-student .....	81
Identificación y relleno de <i>gaps</i> .....	82
5.3. Resultados del <u>Módulo 3</u> .....	84
Resultados con el método de Sinclair mejorado .....	84
Interpretación de resultados .....	85
<b>CAPÍTULO 6. Conclusiones y Trabajos Futuros</b>	
Conclusiones .....	88
Trabajo a futuro.....	89
<b>REFERENCIAS</b> .....	92
<b>ANEXOS</b> .....	99
ANEXO 1 .....	99
ANEXO 2 .....	99
ANEXO 3 .....	99



## Índice de Figuras

Figura 1.1 Características generales de los sistemas Hidrotermales y de Roca Seca Caliente .....	2
Figura 1.2 Descripción general de las preguntas de investigación y la contribución directa en esta investigación.....	7
Figura 2.1 Modelo esquemático de un sistema geotérmico hidrotermal .....	10
Figura 2.2 En relación con la capacidad instalada (1005.8 MWe: Gutiérrez-Negrín, 2020), los recursos geotérmicos en México lo posicionan en el séptimo lugar a nivel mundial.....	11
Figura 2.3 Modelo esquemático de un sistema geotérmico de roca seca caliente.....	12
Figura 2.4 Mapa esquemático de la caldera de Acoculco, con los principales sistemas de fallas, manantiales y los dos pozos exploratorios: EAC-1 y EAC-2.....	14
Figura 2.5 Comparación de los principales usos directos a nivel mundial de energía geotérmica en TJ/año desde 2005, 2010, 2015 y 2020 .....	15
Figura 2.6 Capacidad instalada de 10 países líderes en generación eléctrica en 2015 y 2020 .....	16
Figura 2.7 Resumen de las técnicas convencionales utilizadas en la exploración geotérmica .....	18
Figura 2.8 (A) Medición del viento tridimensional a través de un anemómetro y (B) analizador infrarrojo para medir el número de moléculas de CO <sub>2</sub> .....	20
Figura 2.9 Diagrama esquemático de una CA midiendo el flujo de CO <sub>2</sub> del suelo.....	21
Figura 2.10 Registros de temperatura y zonas convectivas del pozo exploratorio EAC-1 .....	25
Figura 3.1 Interpretación esquemática del proceso de información a través de diferentes estímulos en el cerebro humano .....	28
Figura 3.2 Diagrama de los principales campos de aplicación relacionados con IA .....	29
Figura 3.3 Diagrama general de la rama aprendizaje máquina y la ubicación de las redes neuronales artificiales.....	31
Figura 3.4 Modelo de una neurona artificial .....	32
Figura 3.5 Representación de una topología perceptrón multicapa .....	34
Figura 3.6 Diagrama representativo del entrenamiento de una RNA.....	36
Figura 3.7 Esquema del ejemplo numérico, donde I1 e I2 son las neuronas de entrada; h1 y h2 con las neuronas ocultas; o1 y o2 son las neuronas de salida y w1 a w8 son los pesos sinápticos.....	37
Figura 3.8 Inicialización de los pesos sinápticos con números reales.....	38
Figura 4.1 Diagrama general de los objetivos de la presente investigación.....	43
Figura 4.2 Tres sub-bases de datos a utilizar durante el entrenamiento con 527, 498 y 97 muestras geoquímicas de gases.....	45
Figura 4.3 Diagrama de flujo esquemático utilizado por el método MCDA .....	47
Figura 4.4 (A) Zona de Los Azufres en Acoculco, Puebla; (B) Instalación del equipo <i>Eddy-Covariance</i> ; y, (C) Puesto en operación para medir el flujo de CO <sub>2</sub> atmosférico.....	50
Figura 4.5 Representación esquemática de <i>forward selection</i> .....	51

Figura 4.6 Ejemplo hipotético del manuscrito de Sinclair, 1974 .....	55
Figura 4.7 Diagrama general del método de Sinclair, 1974 .....	56
Figura 4.8 Instrumentos de medición de la Cámara de Acumulación .....	57
Figura 4.9 Frecuencia acumulada en escala logarítmica del flujo de CO <sub>2</sub> medido en Los Azufres .....	59
Figura 4.10 Representación del cambio de pendiente .....	59
Figura 4.11 Diagrama general del procedimiento para calcular el punto de inflexión .....	60
Figura 4.12 Determinación de tres puntos de inflexión con base al cambio del valor de la pendiente .....	61
Figura 4.13 Cuatro grupos identificados en las mediciones del flujo de CO <sub>2</sub> muestreado en Los Azufres.....	62
Figura 4.14 Representación de los Grupos A, B, C, y D por medio de una línea recta.....	62
Figura 5.1 Diagrama general de operación del programa GaS_GeoT .....	69
Figura 5.2 Resultados de la ponderación equilibrada a 10 puntos en el escenario 1.....	74
Figura 5.3 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación global en el escenario 2.....	75
Figura 5.4 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación de entrenamiento en el escenario 3 .....	75
Figura 5.5 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación de prueba en el escenario 4 .....	76
Figura 5.6 Perfil del flujo de CO <sub>2</sub> por hora.....	79
Figura 5.7 Perfil del flujo de CO <sub>2</sub> por día .....	80
Figura 5.8 Regresión lineal entre valores medidos y simulados del flujo de CO <sub>2</sub> .....	81
Figura 5.9 Perfil por día del flujo de CO <sub>2</sub> medido versus simulado .....	82
Figura 5.10 Perfil por día del flujo de CO <sub>2</sub> con relleno de <i>gaps</i> utilizando la ANN-7 .....	83
Figura 5.11 Relleno de huecos en el mes de marzo (A) y abril (B) .....	83
Figura 5.12 Mediciones del flujo de CO <sub>2</sub> con el equipo de medición montado (A) e instalado en el subsuelo (B).....	84
Figura 5.13 Grafica de los resultados finales al aplicar el método de Sinclair mejorado.....	85
Figura 5.14 Distribución porcentual del origen del flujo de CO <sub>2</sub> medido en Los Azufres, Acoculco, Puebla.....	86

## Índice de Tablas

Tabla 2.1 Principales gases en los fluidos geotérmicos y su origen.....	24
Tabla 3.1 Principales campos de aplicación de la IA .....	29
Tabla 3.2 Funciones de activación más utilizadas en RNA.....	32
Tabla 4.1 Extracto de la BDGM con 14 muestras geoquímicas (bases de datos completas: consultar material suplementario de ANEXO 1).....	44
Tabla 4.2 Extracto de la BDEddy con 9 registros del flujo de CO <sub>2</sub> y las variables ambientales que más se correlacionan con él.....	51
Tabla 4.3 Extracto de la BDCA con mediciones del flujo de CO <sub>2</sub> en la zona de Los Azufres ...	57
Tabla 4.4 Frecuencia acumulada de los datos medidos en Acoculco, Puebla (zona de Los Azufres) .....	58
Tabla 5.1 Principales resultados del entrenamiento neuronal para simular el flujo de CO <sub>2</sub> .....	72
Tabla 5.2 Resultados del entrenamiento neuronal para simular el flujo de CO <sub>2</sub> normalizados (0-10) .....	73
Tabla 5.3 Cuatro escenarios creados con ponderaciones diferentes en los coeficientes de correlación.....	74
Tabla 5.4 Resultado final del método <i>MCD</i> A para seleccionar la mejor arquitectura neuronal .....	76
Tabla 5.5 Coeficientes de la ANN-7 asociados a la capa de entrada y la capa oculta de Ec. 5.2.....	77
Tabla 5.6 Coeficientes correspondientes a la ANN-7 asociados a la capa oculta y la capa de salida de Ec. 5.2.....	78
Tabla 5.7 Variables utilizadas para modelar el flujo de CO <sub>2</sub> .....	79
Tabla 5.8 Identificación de las muestras de CO <sub>2</sub> con sus respectivos orígenes y mezclas encontradas .....	87



# CAPÍTULO 1

## Introducción

### 1.1 Antecedentes

Los sistemas geotérmicos son utilizados comercialmente para la generación eléctrica y otros usos directos en más de 50 países alrededor de mundo, incluido México (Bertani, 2016; Kumari & Ranjith, 2019; Lund & Toth, 2020). En comparación con otras energías renovables, representan una fuente geoenergética muy prometedora, dado que tienen la capacidad de funcionar las 24 horas los 365 días de año, aprovechando el calor almacenado en el interior de la Tierra, el cual se estima que es equivalente a 2.5 veces la energía promedio diaria utilizada por los seres humanos (Tomaszewska et al., 2018; Wang et al., 2018; Raine & Reay, 2019).

El origen de los sistemas geotérmicos se da por el calentamiento de grandes extensiones de roca y/o fluidos confinados debajo de la corteza terrestre (Dickson & Fanelli, 2013; Olasolo et al., 2016). Generalmente, están compuestos por tres elementos: una fuente de calor, un reservorio y un fluido. La fuente de calor es una cámara magmática (>600 °C) localizada entre 5 y 10 km de profundidad; el reservorio es una extensión de rocas calientes permeables (de donde los fluidos extraen el calor); y, el fluido que puede estar compuesto por *líquido y/o vapor*, el cual transporta el calor hacia la superficie (Dickson & Fanelli, 2013; Almasi et al., 2015; Ebrahimi et al., 2019). En la actualidad, se han identificado diversos tipos de sistemas geotérmicos, sin embargo, los dos más utilizados para la generación eléctrica son los hidrotermales y los de roca ceca caliente (ver Fig. 1.1).

La explotación de estos recursos geotérmicos no es una tarea fácil, previamente se requiere de una serie de estudios exploratorios (geológicos-geofísicos-geoquímicos) para determinar: el tamaño, forma y estructura del yacimiento, conductividad térmica, composición geoquímica del fluido y temperatura de fondo, entre otros parámetros (Noorollahi et al., 2008; Díaz-González, et al., 2013; Dickson & Fanelli, 2013).

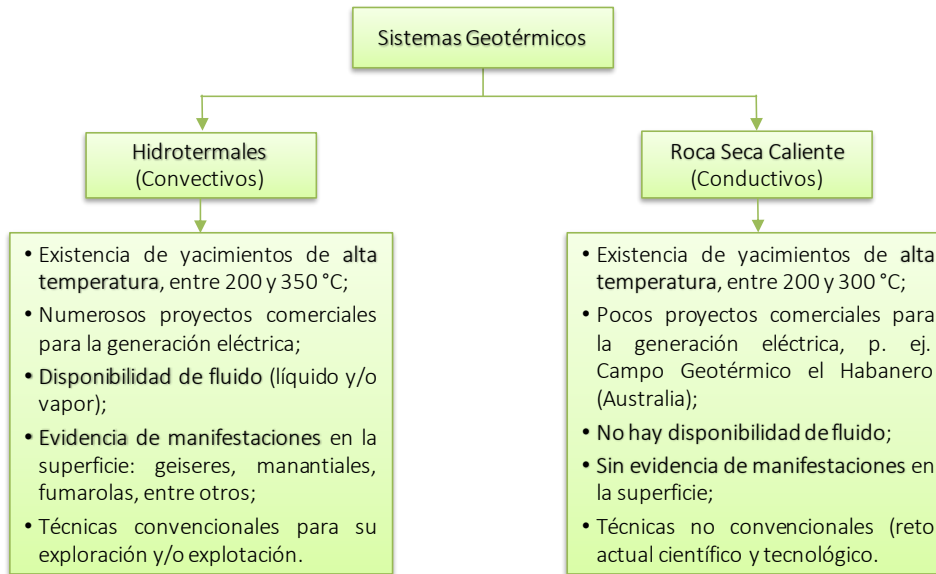


Fig. 1.1 Características generales de los sistemas Hidrotermales y de Roca Seca Caliente (Modificado de Santoyo & Torres-Alvarado, 2010).

Para evaluar una reserva geoenergética virgen, el principal parámetro es la temperatura de fondo de pozo o *BHT* por sus siglas en inglés: *Bottom-Hole Temperature* (Idroes et al., 2019). Esta temperatura se puede medir a través de sondas analógicas, digitales o de fibra óptica, sin embargo, se requiere la perforación de pozos someros lo que representa un costo mayor durante la etapa de exploración (Díaz-González et al., 2013; García-Gil et al., 2015; Espinoza-Ojeda & Santoyo, 2016). Es así como surgen los geotermómetros, los cuales son básicamente ecuaciones empíricas que infieren la temperatura del yacimiento utilizando la composición química de los fluidos geotérmicos, que emergen a la superficie a través de manifestaciones hidrotermales como: manantiales, geiseres, fumarolas, entre otros (García-López et al., 2014; Harvey et al., 2017). En el Capítulo 2, se amplía el fundamento teórico y calibración de los geotermómetros.

A la fecha, se han reportado diversos tipos de geotermómetros en la literatura: químicos (solutos y gases), isotópicos, mineralógicos e inclusiones fluidas (Verma, 2002; Blaise et al., 2014). Particularmente, los geotermómetros de gases utilizan la composición química de los gases presentes en el yacimiento, como son: gases mayores ( $\text{CO}_2$ ,  $\text{H}_2\text{S}$  y  $\text{NH}_3$ ) y gases traza ( $\text{CH}_4$ ,  $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{CO}$ ,  $\text{O}_2$ ). Después del vapor



de agua, el componente más abundante (gas mayoritario) es el dióxido de carbono, en el orden ~70-80% (Portugal et al., 1994; Wallace, 2005). Dado que el CO<sub>2</sub> es uno de los primeros componentes volátiles que son liberados por el magma durante la migración hacia la superficie, las mediciones del flujo de CO<sub>2</sub> permiten identificar una actividad magmática a profundidad o si corresponden a otro tipo de procesos, p. ej. biogénico y/o atmosférico (Wallace, 2005; Giammanco et al., 2007). Por lo que el flujo de CO<sub>2</sub> ayuda a reconocer anomalías geotérmicas, aun cuando las manifestaciones hidrotermales son escasas o nulas ya que no se cuenta con un fluido, como es el caso de los sistemas de roca seca caliente (ver Fig. 1.1).

La presente tesis se enmarca en el modelado del transporte de gases en sistemas geotérmicos hidrotermales y roca seca caliente, utilizando una metodología computacional desarrollada en dos vertientes: (i) herramientas de inteligencia artificial (específicamente, Redes Neuronales Artificiales: RNA); y, (ii) geoquimiometría de gases.

## 1.2 Planteamiento del problema y justificación

Desde los años 80's, se han desarrollado diversos geotermómetros de gases para estimar las temperaturas de yacimientos geotérmicos vírgenes y con ello determinar el potencial geoenergético disponible (D'Amore and Panichi, 1980; Nehring & D'Amore, 1984; Arnórsson & Gunnlaugsson, 1985; Arnórsson, 1987; Giggenbach, 1991; Giggenbach & Glover, 1992; Koga et al., 1995; Supranto et al., 1996; Arnórsson et al., 1998; Blamey, 2006). Generalmente los geotermómetros son utilizados en los sistemas hidrotermales, aunque su aplicación no está restringida a zonas volcánicas con hidrotermalismo. Hasta nuestro conocimiento, alrededor de 31 geotermómetros de gases han sido calibradas fundamentalmente con regresiones lineales simples y múltiples. Sin embargo, en estudios comparativos (Ping & Armannsson, 1996; Powell, 2000; Powel & Cumming, 2010; García-López et al., 2014; Sarmiento et al., 2019) se han observado que las estimaciones de temperatura exhiben diversas discrepancias, como son: (i) con respecto a la *BHT* medida en pozos geotérmicos productores; (ii) entre los geotermómetros; y, (iii) el limitado número de muestras geoquímicas utilizadas para calibrar los

geotermómetros. Es por ello que se justifica el desarrollo de nuevas y mejores herramientas geotermométricas, que provean una estimación de las temperaturas del yacimiento más confiables haciendo uso de las RNA. Resaltando que, antes del presente trabajo, no se había reportado en la literatura internacional ningún geotermómetro de gas basado en RNA (solo en geotermometría de solutos: Bayram, 2001; Can, 2002; Díaz-González et al., 2013), lo cual nos dio un nicho de oportunidad de proponer una investigación original en el campo de la Inteligencia Artificial y Ciencias de la Tierra. Por otro lado, los sistemas de roca seca caliente, los cuales son los más abundantes en la Tierra y que se localizan a profundidades entre 2 - 4 km, representan un reto científico y tecnológico su exploración y explotación (Santoyo & Torres-Alvarado, 2010; Olasolo et al., 2016). En México, este tipo de recursos geotérmicos no han sido aprovechados para la generación eléctrica. Se estima que el potencial de estos recursos susceptible a ser desarrollado es del orden ~24,700 MW a una profundidad máxima de 3 km (Hiriart et al., 2011). Una de las zonas que actualmente se encuentran bajo exploración, es la caldera de Aocolco (Puebla). La cual presenta un importante volumen de masa caliente (~300 °C) en el subsuelo, lo que genera un espectro de posibilidades para su explotación (Viggiano-Guerra et al., 2011).

De acuerdo con la naturaleza de este tipo de recursos geotérmicos, no cuentan con un fluido para extraer el calor. Por lo cual técnicas alternativas para identificar zonas promisorias en la superficie son las más adecuadas, p. ej. medición del flujo de CO<sub>2</sub> en la superficie: cámara de acumulación y, en la atmosfera: *Eddy-Covariance* (Lewicki & Oldenburg, 2004; Melesse & Hanley, 2005; Honglin et al., 2006; Speckman et al., 2015).

### **1.3 Objetivo general**

Desarrollar nuevas herramientas de inteligencia artificial y geoquimiométricas basadas en la composición gaseosa de fluidos para la estimación de temperaturas de fondo y la identificación de anomalías térmicas, así como su aplicación en estudios de prospección y exploración de sistemas geotérmicos.

## 1.4 Objetivos Específicos

El objetivo general se ha dividido en tres módulos:

Módulo 1. Desarrollar nuevos geotermómetros de gases para la predicción de temperaturas de fondo. Este módulo se subdivide en las siguientes etapas:

- a) Actualizar la Base de Datos Geoquímicos Mundial (BDGM) compilada por García-López (2016), que contiene la composición química de los gases CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub> y la *BHT* de pozos geotérmicos productores.
- b) Aplicar diversos criterios geoquimiométricos para asegurar la calidad de las muestras geoquímicas.
- c) Entrenar de forma optimizada múltiples arquitecturas de RNA para predecir la *BHT* utilizando diversas combinaciones de los gases como variables de entrada.
- d) Seleccionar las mejores arquitecturas de RNA utilizando el análisis de decisión multi-criterio o *MCDA* por sus siglas en inglés: *Multi-Criteria Decision Analysis*.
- e) Desarrollar un programa de cómputo que permita estimar la *BHT* utilizando los nuevos geotermómetros de gases (resultados del *MCDA*).
- f) Compilar de la literatura los geotermómetros de gases reportados a la fecha para realizar un análisis comparativo con los nuevos geotermómetros.
- g) Compilar de la literatura la composición gaseosa y *BHT* de pozos geotérmicos productores en una Nueva Base de Datos Geoquímicos Mundial (NBDGM).
- h) Evaluar estadísticamente las temperaturas estimadas por los nuevos geotermómetros de gases y los reportados en la literatura, utilizando la NBDGM.

Módulo 2. Desarrollar nuevos modelos de predicción para modelar anomalías o emisiones naturales de CO<sub>2</sub> en sistemas geotérmicos. Este módulo se subdivide en las siguientes etapas:

- a) Medición de diversos parámetros atmosféricos como son: flujo de CO<sub>2</sub>, flujos de energía y variables ambientales censadas en la caldera de Acoculco (Puebla) por el grupo de investigación en geotermia (con el cual colaboré en el Proyecto P09-CeMIE-Geo), utilizando un equipo micrometeorológico denominado *Eddy-Covariance*.
- b) Crear una Base de Datos con las mediciones del equipo *Eddy-Covariance* (BDEddy).
- c) Entrenar de forma optimizada múltiples arquitecturas de RNA para predecir el flujo de CO<sub>2</sub> utilizando diversas combinaciones de las variables ambientales y los flujos de energía.
- d) Seleccionar la mejor arquitectura de RNA utilizando el *MCDA*.
- e) Predecir el flujo de CO<sub>2</sub> utilizando la RNA seleccionada, en los periodos de ausencia de datos (*gaps*) censados por el equipo *Eddy-Covariance*.

Módulo 3. Desarrollar una nueva metodología geoquimiométrica computacional apoyada en el método de Sinclair para la discriminación del transporte de gas en sistemas geotérmicos. Este módulo se subdivide en las siguientes etapas:

- a) Medir en la superficie terrestre, el flujo de CO<sub>2</sub> emitido en la caldera Acoculco, utilizando el método de cámara de acumulación.
- b) Crear una Base de Datos con las mediciones de la Cámara de Acumulación (BDCA).
- c) Utilizar el método de Sinclair (1974) como procedimiento geoquimiométrico para discriminar las mediciones del flujo de CO<sub>2</sub> medido en Acoculco.

### **1.5 Preguntas de investigación relacionadas con los trabajos publicados y en preparación**

Con base a los objetivos planteados en la presente investigación y el alcance obtenido con los trabajos publicados (ANEXO 1, ANEXO 2 y ANEXO 3) y los trabajos en preparación (MS-A y MS-B), a continuación, se enlistan cinco preguntas de investigación (P1-P5) y su contribución académica (ver Fig. 1.2).

*P1. ¿Las RNA pueden ser utilizadas para modelar el transporte de gases en sistemas geotérmicos?*

*P2. ¿Las RNA pueden modelar procesos multivariados para desarrollar nuevos y mejores geotermómetros de gases?*

*P3. ¿El método MCDA puede ser utilizado para evaluar integralmente el desempeño de múltiples arquitecturas de RNA?*

*P4. ¿Las RNA pueden ser utilizadas para modelar anomalías geotérmicas?*

*P5. ¿Es posible identificar el origen de anomalías geotérmicas con base a la geoquimiometría de gases?*

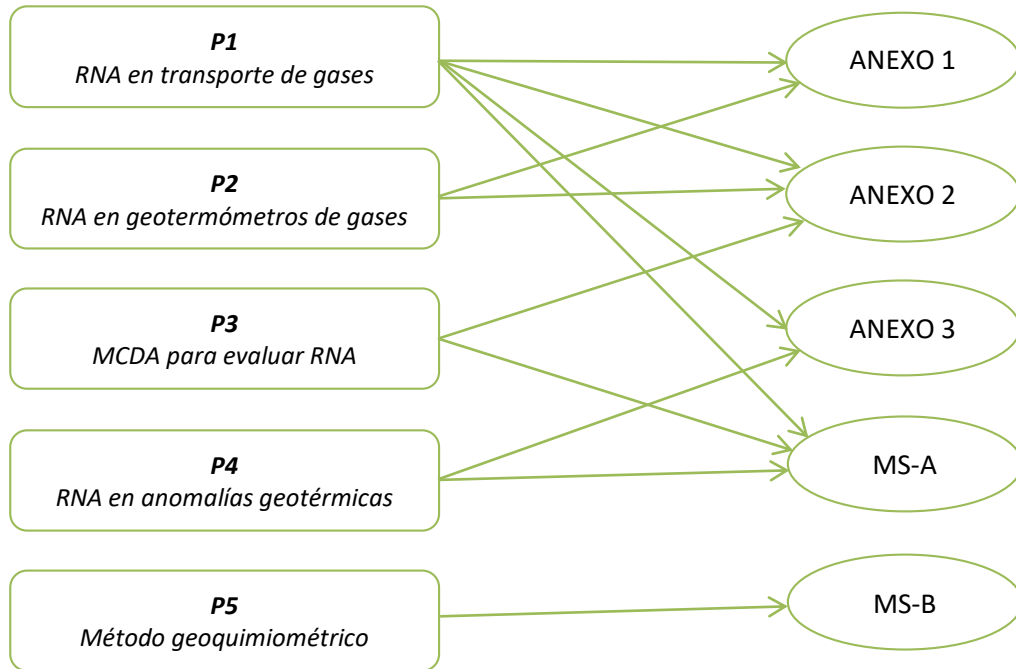


Fig. 1.2 Descripción general de las preguntas de investigación y la contribución directa en esta investigación.

## 1.6 Estructura de la tesis

En el Capítulo 2 se abordan las generalidades de los sistemas geotérmicos hidrotermales y de roca seca caliente. La exploración de estos recursos, la identificación de anomalías geotérmicas utilizando las emisiones de gases, así como la importancia de la estimación de la temperatura para evaluar una reserva geoenergética.

En el Capítulo 3 se describen los fundamentos de la inteligencia artificial, los modelos más utilizados dentro del ámbito de aprendizaje máquina como las RNA. Posteriormente a las generalidades de las RNA, se aborda el algoritmo 'retro-propagación' también conocido como *Backpropagation*. Por último, se presenta un ejemplo numérico del algoritmo *Backpropagation*.

En el Capítulo 4, se muestra la metodología empleada para desarrollar los objetivos específicos descritos en la sección 1.4 del Capítulo 1. La descripción general de las bases de datos utilizadas en cada objetivo, así como el desarrollo de estos.

En el Capítulo 5 se exhiben los resultados obtenidos en cada objetivo específico y las publicaciones derivadas de la presente investigación.

En el Capítulo 6, se presentan las conclusiones de la investigación y los trabajos futuros sobre esta línea de investigación: inteligencia artificial aplicada a la exploración y explotación de recursos geotérmicos.

## CAPÍTULO 2

### Importancia del Transporte de Gases en Sistemas Geotérmicos

**Resumen del capítulo.** En este Capítulo se aborda la importancia del transporte de gases en sistemas geotérmicos. Se describe la zona geotérmica de Acoculco, la cual está en su etapa exploratoria y que promete ser el primer sistema de roca seca caliente en explotación de México. Las generalidades de los estudios geológicos-geofísicos-geoquímicos. Las técnicas convencionales utilizadas en los sistemas hidrotermales y las no convencionales para los de roca seca caliente. Las diferencias fundamentales entre los sistemas “ocultos” y “ciegos”. Finalmente, los indicadores inferidos utilizando la geoquímica de los gases geotérmicos y la importancia de estimar la temperatura mediante uso de las emisiones de gases.

#### 2.1 Sistemas geotérmicos

Los sistemas geotérmicos representan al conjunto de elementos (fuente de calor, reservorio y fluido) que interactúan entre sí, para aprovechar en la superficie el calor proveniente del interior de la Tierra. Se localizan en regiones con un gradiente geotérmico promedio ( $\sim 30$  °C/km) o ligeramente superior, donde las temperaturas de fondo no superan los 100 °C a menos de 2 km de profundidad (Santoyo & Barragán-Reyes, 2010; Dickson & Fanelli, 2013). También pueden encontrarse alrededor de los márgenes de las placas tectónicas, donde las temperaturas pueden ser significativamente más altas, incluso por encima de los 400 °C (Dickson & Fanelli, 2004).

La clasificación de los sistemas geotérmicos con base a la temperatura del fluido es: recursos de alta ( $>200$  °C), mediana (100-200 °C) y baja entalpía ( $<100$  °C). Generalmente, la producción de electricidad se sustenta en los fluidos de alta entalpía (Santoyo & Barragán-Reyes, 2010; Iorio et al., 2020). En el Capítulo 1 se mencionó que existen diversos tipos de sistemas geotérmicos para la generación de electricidad y otros uso directos, tales como: (i) sistemas hidrotermales (convectivos); (ii) sistemas de roca seca caliente (conductivos); (iii) sistemas

geopresurizados; (iv) sistemas conductivos sedimentarios; (v) sistemas radiogénicos; (vi) sistemas marinos; (vii) sistemas magmáticos; (viii) sistemas de agua caliente asociado con yacimientos de petróleo y gas; y (ix) sistemas supercríticos (Chamorro-Camazón, 2009; Santoyo & Torres-Alvarado, 2010). Dado que los sistemas hidrotermales y de roca seca caliente son los más utilizados y que al mismo tiempo forman parte de la presente investigación, solo éstos se describen a continuación:

**Sistemas hidrotermales (convectivos).** Están conformados por los siguientes elementos: una fuente de calor relacionada a una cámara magmática con altas temperaturas  $>600\text{ }^{\circ}\text{C}$ ; el fluido geotérmico es agua meteórica (en fase líquida o vapor, dependiendo de la temperatura y presión); un volumen de rocas permeables donde se almacenan los fluidos y una capa de roca sello que mantiene al fluido confinado (Santoyo & Barragán-Reyes, 2010; Dickson & Fanelli, 2013). En la Fig. 2.1 se representa el modelo esquemático ideal de este tipo de sistemas.

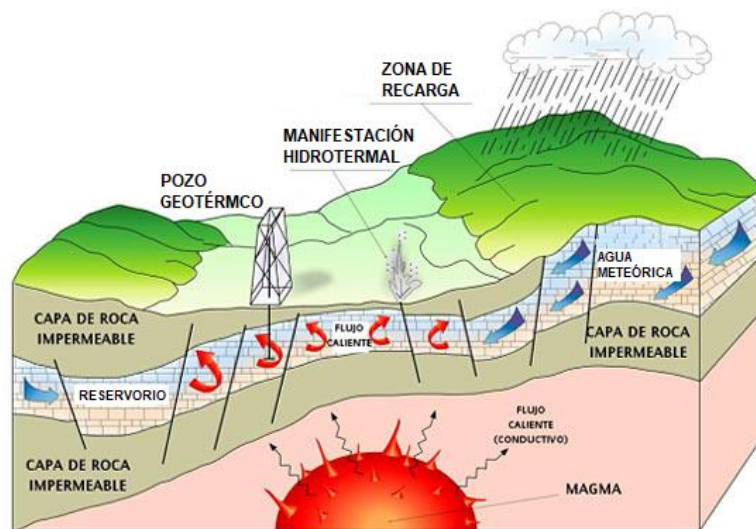


Fig. 2.1 Modelo esquemático de un sistema geotérmico hidrotermal (Modificado de Dickson & Fanelli, 2013).

Una de las características más importantes de este tipo de sistemas es que cuentan con un fluido, el cual asciende a la superficie (a través de fisuras y/o fracturas en la corteza), presentando distintas manifestaciones: manantiales calientes, lagunas de lodo hirviente, fumarolas, géiseres, entre otros (Dickson & Fanelli, 2004; Russell et al., 2005).



Con base al fluido almacenado en los sistemas hidrotermales, se clasifican en yacimientos de: (i) vapor—dominante: los cuales contienen vapor seco de alta entalpía y existen pocos en del mundo, p. ej. Larderello (Italia), Kamojang (Indonesia), The Geysers (EUA) y algunas zonas al sur de Los Azufres (México), entre otros; (ii) liquido—dominante de alta entalpía: contiene salmuera súper caliente (>200 °C) y en comparación con los anteriores, son más abundantes, p. ej. Los Humeros y Cerro Prieto (México), Berlín (El Salvador), Krafla (Islandia), entre otros; y, (iii) liquido—dominante de mediana y baja entalpía: contienen salmuera caliente que puede alcanzar los 200 °C, entre los hidrotermales son los más abundantes a nivel mundial, p. ej. Onikobe y Mori (Japón), Miravalles (Costa Rica), Wairakei-Tauhara (Nueva Zelanda) y decenas de sitios con hidrotermalismo (ver Fig. 2.2) distribuidos en todo México (Santoyo & Barragán-Reyes, 2010; Kaya et al., 2011; Stelling et al., 2016).



Fig. 2.2 En relación con la capacidad instalada (1005.8 MWe: Gutiérrez-Negrín, 2020), los recursos geotérmicos en México lo posicionan en el séptimo lugar a nivel mundial (Modificado de Santoyo & Barragán-Reyes, 2010).

En los sistemas hidrotermales, el fluido (en su estado líquido y/o gaseoso) juega un papel muy importante, así como su correlación con la temperatura de fondo del yacimiento. A continuación, se describen los sistemas geotérmicos que no cuentan con este importante elemento, pero que aun así representan una fuente geoenergética muy prometedora: sistemas de roca seca caliente.

**Sistemas de roca seca caliente (conductivos).** Este tipo de sistemas geotérmicos no cuentan con un fluido debido sus características geológicas, en su mayoría rocas cristalinas y no permeables, con temperaturas en el fondo  $>600\text{ }^{\circ}\text{C}$  (Santoyo & Barragán-Reyes, 2010). También son conocidos como sistemas geotérmicos mejorados o EGS por sus siglas en inglés: *Enhanced Geothermal Systems*. Están disponibles en el subsuelo desde los 2 km de profundidad en todo el mundo, por lo que se les considera los más abundantes (Santoyo & Torres-Alvarado, 2010). Su explotación requiere de técnicas de fracturamiento hidráulico para la creación de un yacimiento artificial y posteriormente la inyección de un fluido de trabajo (generalmente agua a temperatura ambiente) y así poder extraer el calor de la roca, ver Fig. 2.3 (Olasolo et al., 2016; Kumari & Ranjith, 2019).

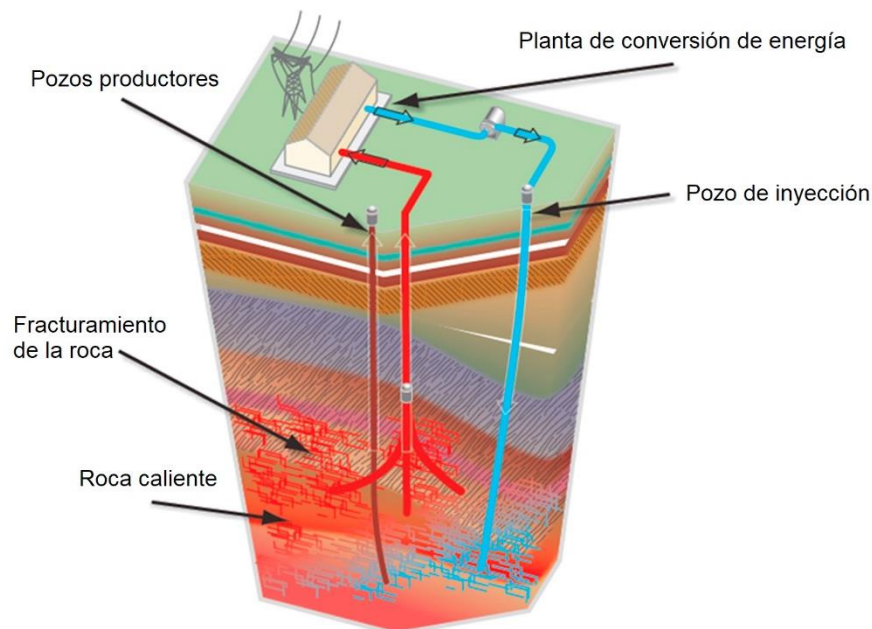


Fig. 2.3 Modelo esquemático de un sistema geotérmico de roca seca caliente (Modificado de Olasolo et al., 2016).

Actualmente, existen pocos proyectos en exploración y desarrollo de sistemas *EGS*, tales como: Fenton Hill (EUA), Soultz-sous-Forêts (Francia), Landau (Alemania), Habenero (Australia), entre otros (Olasolo et al., 2016; Isaka et al., 2019).

De acuerdo con el desarrollo de este tipo de sistemas, y la necesidad de contar con agua dulce para su estimulación, se ha reportado en la literatura (Xu et al., 2008; Pruess & Spycher, 2009; Isaka et al., 2019) que los *EGS* pueden ser operados por CO<sub>2</sub> en lugar de agua como fluido de transmisión de calor. Bajo este escenario se podría combinar la recuperación de la energía geotérmica y como un beneficio secundario el almacenamiento geológico simultáneo del CO<sub>2</sub>.

Se ha estimado que la energía almacenada en los sistemas *EGS* a una profundidad máxima de 10 km, es equivalente a más de 500 veces la energía acumulada en todos los yacimientos de gas y petróleo alrededor del mundo; esto lo convierte en un recurso de mucha importancia en el marco de las energías renovables (Iglesias et al., 2005). No obstante de ser un recurso muy prometedor y disponible en todo el mundo, su exploración y explotación conlleva una serie de retos científicos y tecnológicos aún por resolver, tales como: (i) la implementación de técnicas no convencionales durante su exploración para identificar anomalías geotérmicas en la superficie; (ii) la creación de un volumen apropiado de roca fracturada; (iii) la minimización de pérdidas de agua inyectada hacia la roca fracturada; (iv) la disminución de permeabilidad y porosidad del yacimiento fracturado, debido a la precipitación mineral subsiguiente asociada con la disolución de minerales a través del tiempo (Santoyo & Torres-Alvarado, 2010; Isaka et al., 2019).

En México, uno de los sitios más prometedores para desarrollar el primer sistema *EGS* está ubicado en la caldera de Acoculco. Es una zona geotérmica localizada en el municipio de Chignahuapan en el estado de Puebla (a unos 180 km al noroeste de la Ciudad de México). Como antecedente de esta zona, la caldera de Acoculco, ha sido reconocida por la Comisión Federal de Electricidad (CFE) como un sistema de roca seca caliente en exploración y podría ser candidata para desarrollar un

EGS. En 1995 y 2008 se perforaron dos pozos exploratorios a una profundidad menor a los 2 km (pozo EAC-1: 1810 m y EAC-2: 1900 m, ver Fig. 2.4). En ambos pozos, se midieron temperaturas en el fondo de  $\sim 300$  °C, con permeabilidad literalmente nula (Viggiano-Guerra et al., 2011; Peiffer et al., 2014).

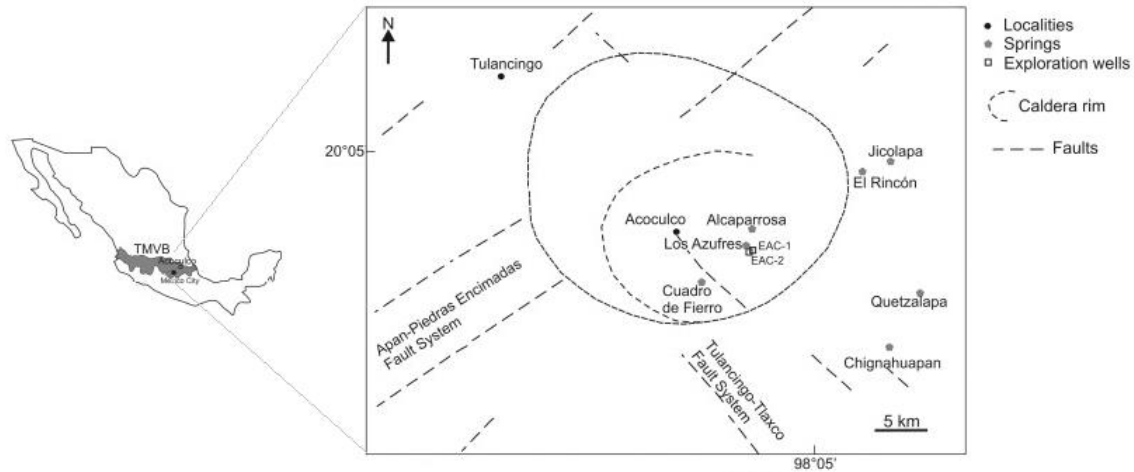


Fig. 2.4 Mapa esquemático de la caldera de Acoculco, con los principales sistemas de fallas, manantiales y los dos pozos exploratorios: EAC-1 y EAC-2 (modificado de Peiffer et al., 2014).

Dado que las manifestaciones geotérmicas dentro de la caldera están limitadas a unos pocos manantiales fríos y algunas emanaciones frías de CO<sub>2</sub> (mofetas), así como estudios realizados en la zona de mineralogía hidrotermal, se asume que Acoculco se encuentra en su etapa hidrotermal senil (Viggiano-Guerra et al., 2011). Estas características conducen al desarrollo de técnicas no convencionales para su exploración y aprovechamiento del recurso térmico ( $\sim 300$  °C a 2 km de profundidad), mismas que se describen en la sección 2.4 del presente Capítulo.

## 2.2 Aprovechamiento actual de los recursos geotérmicos en el Mundo y en México

La energía geotérmica es una fuente de energía renovable aprovechada alrededor del mundo para la generación de electricidad y usos directos (Chamorro-Camazón, 2009). Como antecedente, en 1904 en Larderello, Italia se produjo electricidad por primera vez en el mundo a través del calor proveniente del interior de la Tierra. Por otro lado, la geotermia nace en México en 1959 en la localidad de Pathé, Hidalgo

(Santoyo & Barragán-Reyes, 2010). A continuación, se describe el escenario actual de la geotermia en usos directos y capacidad instalada para generar electricidad:

**Usos directos.** De acuerdo con los datos reportados para el *World Geothermal Congress 2020* por Lund & Toth (2020), la capacidad instalada para usos directos a nivel mundial alcanzó los 107,727 MWt en 88 países, presentando un incremento del 52% con respecto al 2015 creciendo a una tasa compuesta anual del 8.73%. Actualmente, dentro de los principales usos directos encontramos: las bombas de calor, calefacción (viviendas, invernaderos y estanques agricultores), secado agrícola, usos industriales, baño y natación, entre otros (Lund & Toth 2020). El mayor aprovechamiento en este rubro se da a través de las bombas de calor (ver Fig. 2.5), siendo los líderes de esta tecnología países como China, Estados Unidos, Suecia, Alemania y Finlandia.

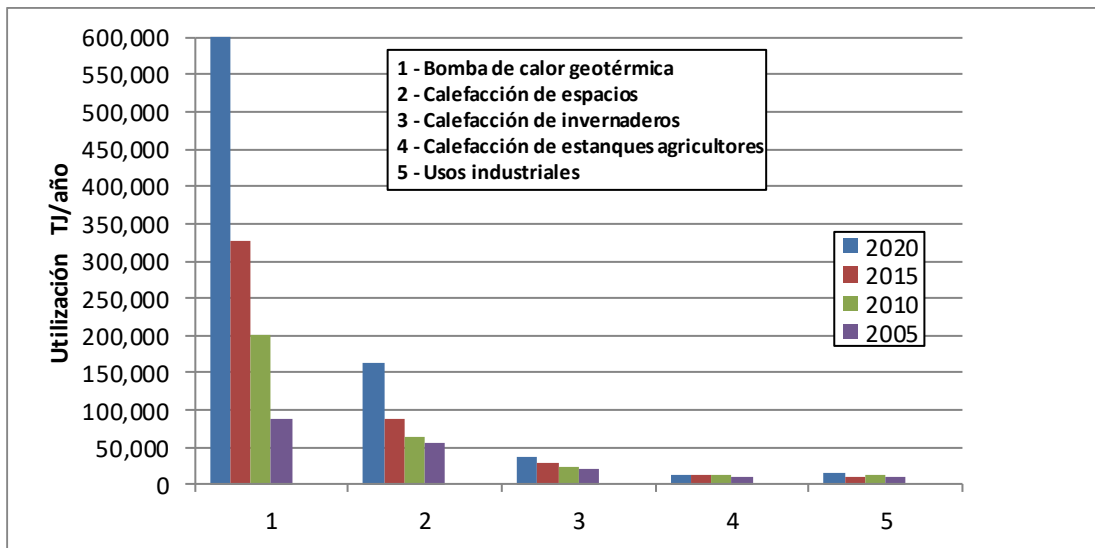


Fig. 2.5 Comparación de los principales usos directos a nivel mundial de energía geotérmica en TJ/año desde 2005, 2010, 2015 y 2020 (modificado de Lund & Toth 2020).

En México, los usos directos están centrados en la calefacción de espacios, secado agrícola, baño-natación y 11 bombas de calor. Estas últimas, se encuentran instaladas como proyectos de demostración en los estados de Puebla, Baja California y Michoacán (en una pequeña escuela y clínica de salud, en invernaderos, en laboratorios, en la Universidad Politécnica de Mexicali y en la Universidad de Michoacán en Morelia), Lund & Toth 2020.

**Generación de electricidad.** Con base a los datos reportados para el *World Geothermal Congress 2020* por Hutterer (2020) la capacidad instalada actualmente en Mega watts y Giga watts eléctricos es de 15,950 MWe mostrando un incremento de 3.6 GWe (~27%) con respecto al 2015. El crecimiento en los últimos 5 años obedece fundamentalmente al desarrollo de nuevos proyectos geotermoeléctricos en países que ya generaban electricidad a partir de recursos geotérmicos y a la incorporación de cinco países que lo hicieron por primera vez. Entre los países nuevos en utilizar la geotermia se encuentran: Bélgica (0,8 MWe), Chile (48 MWe), Croacia (16,5 MWe), Honduras (35 MWe) y Hungría (3 MWe). A continuación, se muestra en la Fig. 2.6 los 10 países líderes en capacidad instalada para la generación eléctrica.

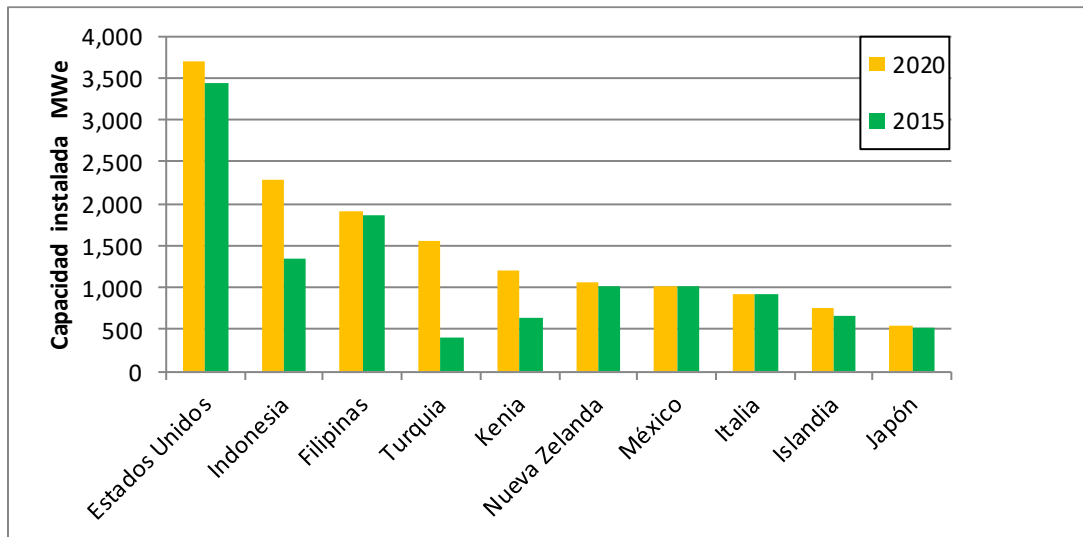


Fig. 2.6 Capacidad instalada de 10 países líderes en generación eléctrica en 2015 y 2020 (modificado de Bertani, 2016 y Hutterer, 2020).

En la Fig. 2.6, se observa que en los últimos 5 años la mayoría de los países ha incrementado su capacidad instalada. El más sobresaliente de ellos es Turquía quien reporto en el 2015 407 MWe y 1549 MWe en el 2020 (con un crecimiento ~280%). Por otro lado, México reporta para el 2020 una capacidad instalada de 1005.8 MWe (Gutiérrez-Negrín et al., 2020) lo que representa un ligero decremento del 1.1 % con respecto al 2015 (1017.4 MWe). Esto se debe a que, en el campo geotérmico de Cerro Prieto cuatro unidades se dieron de baja dado que venían operando desde los años 70's (2 unidades operando desde 1973 y 2 desde 1979).

## 2.3 Exploración de recursos geotérmicos – técnicas convencionales

Durante la etapa exploratoria de un recurso geotérmico, es imprescindible la realización de una serie de estudios multidisciplinarios, tales como: geológicos, geofísicos y geoquímicos (Noorollahi et al., 2008; Dickson & Fanelli, 2013). El objetivo de dichos estudios es determinar la factibilidad de explotación del recurso, fundamentalmente con fines de generación eléctrica y/o usos directos (dependiendo de la temperatura del fluido) tales como: baños medicinales, balnearios, invernaderos, climatización, entre otros (Marzolf, 2014).

En general, los estudios geológicos identifican y catalogan el tipo de rocas que forman en basamento, así como la ubicación de las manifestaciones superficiales, las fracturas, la permeabilidad de las rocas, entre otros (Noorollahi et al., 2008).

Los estudios geofísicos están orientados a la obtención de los parámetros físicos de las formaciones geológicas profundas, tales como: la conductividad eléctrica (métodos eléctricos y electromagnéticos), velocidad de propagación de ondas elásticas (prospección sísmica), densidad (prospección gravimétrica) y susceptibilidad magnética (Dickson & Fanelli, 2013).

Los estudios geoquímicos realizan un análisis químico e isotópico del agua y gas de las manifestaciones termales, como son: manantiales, fumarolas, emisiones de gases, entre otros (Almasi et al., 2015; Ebrahimi et al., 2019). Una de las tareas más importantes dentro de los estudios geoquímicos es la estimación de las temperaturas de fondo del yacimiento (*BHT*) con base a la composición geoquímica de los gases presentes ( $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{NH}_3$ ,  $\text{CH}_4$ ,  $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{CO}$ ,  $\text{O}_2$ ). Esta relación existente entre los gases y la *BHT*, se le conoce como geotermómetros de gases, los cuales serán descritos en la sección 2.6 del presente Capítulo. Es importante resaltar que la metodología integral que incluye los estudios geológicos-geofísicos-geoquímicos, está orientada mayormente a la exploración de los sistemas geotérmicos hidrotermales, es decir, la implementación de técnicas convencionales

donde existe un fluido (líquido y/o gas) en el yacimiento. En la Fig. 2.7 se muestra un esquema general de estas técnicas utilizadas en la etapa exploratoria de un recurso geotérmico.

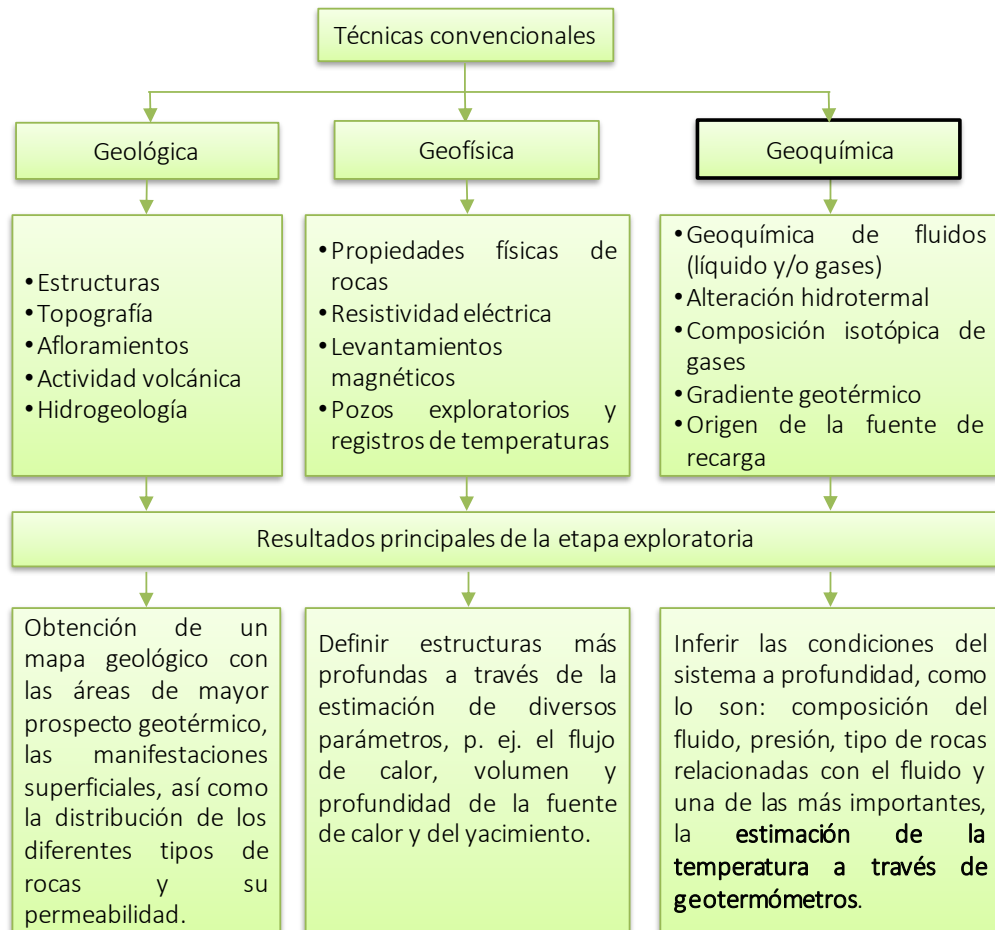


Fig. 2.7 Resumen de las técnicas convencionales utilizadas en la exploración geotérmica (Modificado de Guerrero-Martínez et al., 2020).

## 2.4 Exploración de recursos geotérmicos – técnicas no convencionales

Las técnicas no convencionales surgen a partir de la necesidad de explorar aquellos sistemas geotérmicos que no cuentan con un fluido en el yacimiento (líquido y/o gas), y que representan una importante fuente de energía geotérmica a ser explotada, como lo son los sistemas de roca seca caliente. Dependiendo de la existencia limitada o nula de manifestaciones superficiales, pueden ser denominados sistemas “ocultos” o “ciegos”, respectivamente (Dobson, 2016).



En la sección 2.1 del presente Capítulo, se describió una de las zonas más prometedoras en México para desarrollar el primer EGS. Dado que las manifestaciones superficiales en Acoculco, están limitadas a unas pocas mofetas de CO<sub>2</sub> y manantiales fríos, se puede asumir que es un sistema de roca seca caliente “oculto”. En estos sistemas las emisiones de gases de moderada a baja intensidad (p. ej. CO<sub>2</sub>, CH<sub>4</sub>, He), pueden ser una de las principales señales en la superficie para identificar anomalías geotérmicas (Lewicki & Oldenburg, 2004). Por lo tanto, la identificación de emisiones anómalas principalmente del CO<sub>2</sub>, apunta a ser una herramienta importante para la exploración de estos recursos (Klusman et al., 2000).

Se han reportado en la literatura (Lewicki & Oldenburg, 2004; Dobson, 2016) diversas técnicas no convencionales para la exploración de sistemas geotérmicos ocultos, algunas de ellas son: (i) el analizador de gas infrarrojo para medir concentraciones de CO<sub>2</sub> en la superficie; (ii) el método de *Eddy-Covariance* para medir el flujo neto de distintos gases (p. ej. CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>O) sobre un área determinada; (iii) el método de Cámara de Acumulación para medir los flujos de CO<sub>2</sub> del suelo; (iv) imágenes hiperespectrales de una región para identificar concentraciones elevadas de CO<sub>2</sub>; y, (v) detección y rango de luz que puede medir las concentraciones de gases varios gases en la atmosfera (p. ej. CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>O) con luz láser. Como podemos observar, en todas las técnicas aparece el CO<sub>2</sub> y la razón es que después del agua, es el gas mayoritario en un ambiente geotérmico. De acuerdo con los objetivos planteados en la presente investigación, a continuación, se describen los métodos *Eddy-Covariance* y Cámara de Acumulación.

**Método *Eddy-Covariance* (EC).** Es un método que permite cuantificar intercambios de distintos gases (p. ej. CO<sub>2</sub>) y energía, entre la superficie de la Tierra y la atmosfera. Con ello, se puede realizar un monitoreo continuo del flujo de CO<sub>2</sub> en grandes extensiones, con tamaños que varían desde unos pocos m<sup>2</sup> hasta km<sup>2</sup> (Burba, 2013). El método se basa en mediciones directas y rápidas del transporte

real del CO<sub>2</sub> mediante un viento tridimensional en tiempo real *in-situ*. De esta forma, se puede promediar el producto de la serie de tiempo de la concentración fluctuante de CO<sub>2</sub> ( $c(x, y, z, t)$ ), y la serie de tiempo de la velocidad del viento vertical fluctuante ( $w(x, y, z, t)$ ); Lewicki & Oldenburg (2004).

Las mediciones se realizan por medio de dos aparatos de alta tecnología: el anemómetro y el analizador de gas infrarrojo o *IRGA* por sus siglas en inglés: *Infrared Gas Analyzer*. El primero de ellos mide la velocidad y dirección del viento, y el segundo mide las concentraciones de CO<sub>2</sub> en el aire. Los datos se almacenan simultáneamente decenas de veces por minuto y el promedio puede ser calculado cada 10 min, 30 min, 1 hr., 2 hrs. o 4 hrs. (Burba, 2013). En la Fig. 2.8 se representa el principio de medición de este método.

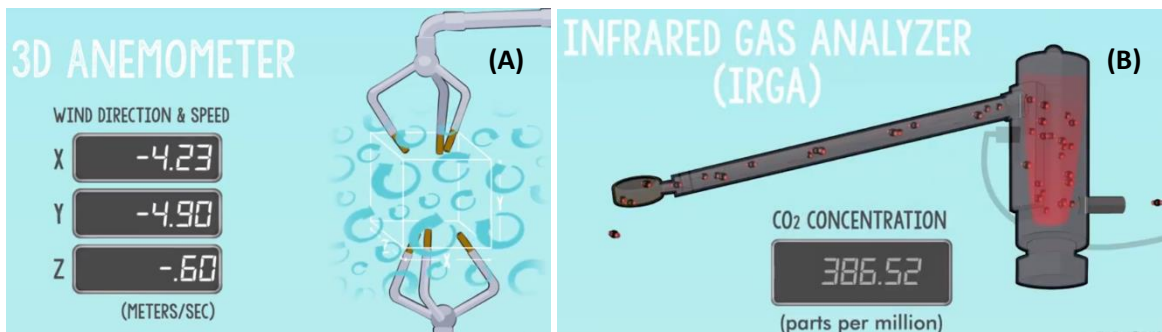


Fig. 2.8 (A) Medición del viento tridimensional a través de un anemómetro y (B) analizador infrarrojo para medir el número de moléculas de CO<sub>2</sub> (Williams et al., 2017).

El flujo de CO<sub>2</sub> describe el movimiento de las moléculas que se trasladan de un lugar a otro sobre un área en un tiempo determinado. Es decir, si en un tiempo  $t_1$  pasan 7 moléculas de CO<sub>2</sub> por el área de medición y en  $t_2$  solo pasan 5 moléculas debido a la fluctuación del viento, entonces existió un flujo neto de 2 moléculas de CO<sub>2</sub>.

Una ventaja del método EC es que proporciona una medición promediada espacial y temporalmente y, por lo tanto, filtra inherentemente la variabilidad espacial y temporal a pequeña escala en el flujo de superficie. Esto, potencialmente proporciona una medida más representativa del flujo de superficie (Lewicki & Oldenburg, 2004; Burba, 2013).

**Método Cámara de Acumulación (CA).** Este método permite medir el flujo de  $\text{CO}_2$  en suelos por unidad de área y unidad de tiempo. Se utiliza un *IRGA* y una cámara hueca con fondo abierto como se muestra en la Fig. 2.9. Cada medición generalmente se realiza en varios minutos y se requiere una persona para operar la instrumentación.

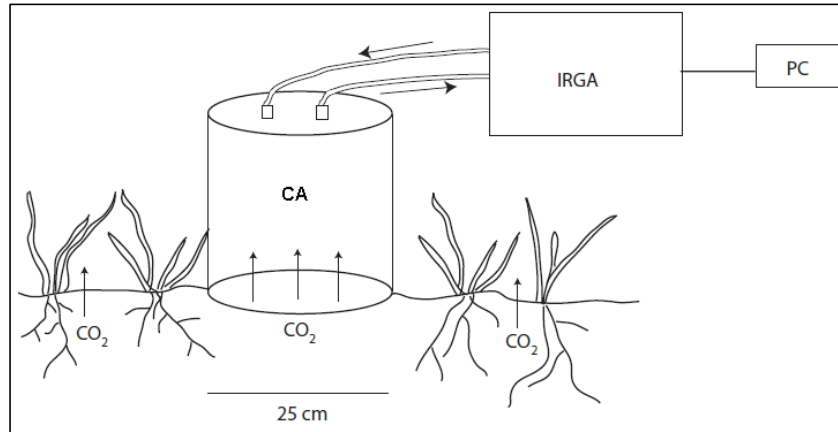


Fig. 2.9 Diagrama esquemático de una CA midiendo el flujo de  $\text{CO}_2$  del suelo (modificado de Lewicki & Oldenburg, 2004).

El procedimiento es colocar la CA con fondo abierto directamente sobre la superficie a medir, posteriormente el aire contenido se hace circular a través de la CA. La tasa de cambio de la concentración de  $\text{CO}_2$  en la cámara ( $d[\text{CO}_2]/dt$ ) se mide con el *IRGA* y se registra en la PC (Lewicki & Oldenburg, 2004). El flujo de  $\text{CO}_2$  ( $F$ ) se calcula luego de acuerdo con la siguiente ecuación:

$$F = \left(\frac{pV}{A}\right) \left(\frac{d[\text{CO}_2]}{dt}\right) \quad \text{Ec. (2.1)}$$

donde  $p$  es la densidad molar del aire;  $V$  es el volumen del sistema de medición y  $A$  es el área de la huella de la CA.

Para evitar mediciones anómalas, se recomienda presionar con firmeza la CA contra el suelo con la finalidad de aislar el volumen interno de la CA del exterior, teniendo cuidado de no alterar la superficie durante la medición dado que esto podría resultar en flujos del gas anormalmente altos (Fridriksson et al., 2016).

## **2.5 El transporte de gases como parámetro geoquímico para identificar anomalías geotérmicas**

La interpretación y el análisis de los datos obtenidos en las mediciones geoquímicas de los gases tienen como objetivo reconstruir las condiciones termodinámicas del yacimiento a profundidad (Nicholson, 1993; Lewicki & Oldenburg, 2004; Lowenstern et al., 2015). A continuación, se enlistan algunos indicadores importantes inferidos a través del transporte de gases geotérmicos:

- a) Anomalías geotérmicas en la superficie tales como: fumarolas, los suelos vaporizantes, las pozas calientes, mofetas de gas frío, entre otras.
- b) El  $H_2$  es un gas altamente reactivo, por lo cual su concentración varía con el tiempo y a partir de esto podemos inferir direcciones de flujo ascendente.
- c) Las altas concentraciones de  $N_2$  suelen encontrarse en sistemas geotérmicos de baja entalpía.
- d) Generalmente, la presencia de altas concentraciones de oxígeno en el muestreo de los gases en sistemas geotérmicos denota contaminación de aire atmosférico.
- e) Las altas concentraciones de los gases geotérmicos indican una alta permeabilidad por la cual ascienden a la superficie desde el yacimiento.
- f) Cuando existen elevadas concentraciones de  $H_2$  y  $H_2S$  se puede inferir que la zona es de alta temperatura tanto en el yacimiento como en la superficie.
- g) Con base a la composición geoquímica se puede inferir la temperatura del yacimiento.

De acuerdo con su naturaleza, los gases geotérmicos pueden dividirse en dos grupos: gases reactivos ( $H_2O$ ,  $CO_2$ ,  $H_2S$ ,  $NH_3$ ,  $H_2$ ,  $N_2$  y  $CH_4$ ) y gases inertes (Ar, He). Como su nombre lo indica los gases reactivos son aquellos que producen diferentes reacciones químicas con las rocas, y los gases inertes no participan en estas reacciones (Nicholson, 1993).

La migración de los gases geotérmicos hacia la superficie es a través de fisuras y/o fracturas en la corteza terrestre siguiendo patrones de flujo vertical, mientras que

emanaciones líquidas siguen patrones de flujo horizontal (Nicholson, 1993). El mecanismo del transporte de los gases puede ser explicado a través de dos procesos: flujos difusivos y advectivos. Generalmente, los flujos bajos de un gas (p. ej. CO<sub>2</sub>) están relacionados por procesos difusivos y los flujos altos por procesos advectivos (Chiodini et al., 2001). Una analogía simple para entender ambos mecanismos del transporte de gas es considerar el transporte de la sal vertida en un río. Si el río es al principio de agua dulce y fluye rápidamente, la forma predominante del transporte de la sal en el agua será advectiva, cuando el flujo de agua en sí mismo transportaría la sal. Si el río no estuviera fluyendo, la sal simplemente se dispersaría hacia fuera desde su fuente en una manera difusiva.

Matemáticamente el transporte difusivo está representado por la primera ley de Fick, la cual dice que existe un flujo proporcional a un gradiente y que se expresa con la siguiente ecuación:

$$F_D = -D \frac{dC}{dX} \quad \text{Ec. (2.2)}$$

donde  $D$  es el coeficiente de difusión;  $\frac{dC}{dX}$  es el gradiente de concentración en la dirección  $X$  y el signo negativo indica que el flujo va de las zonas de mayor a las de menor concentración (Gomez, 2016).

El flujo advectivo a través de un medio poroso se expresa matemáticamente por la ley de Darcy:

$$F_A = -\frac{k}{\mu} \frac{dP}{dX} \quad \text{Ec. (2.3)}$$

donde  $k$  es la permeabilidad absoluta del medio poroso;  $\mu$  es la viscosidad del fluido;  $\frac{dP}{dX}$  es el gradiente de presión en la dirección  $X$  del flujo. El signo negativo indica si  $dX$  es medido en la dirección de flujo, la presión  $P$  declina conforme  $X$  se incrementa. Esta declinación dará como resultado un valor negativo para  $\frac{dP}{dX}$ . En consecuencia,

el signo negativo deberá ser incluido en la ecuación para asegurar que el flujo advectivo sea positivo (Sánchez-Robles, 2010). En la Tabla 2.1, se enlistan los gases más comúnmente medidos en ambientes geotérmicos y el origen de estos.

Tabla 2.1. Principales gases en los fluidos geotérmicos y su origen (Nicholson, 1993).

Gas	Origen
CO <sub>2</sub> Dióxido de carbono	Después del H <sub>2</sub> O, es el gas mayoritario, su origen puede ser magmático o por reacciones químicas con las rocas.
H <sub>2</sub> S Ácido sulfhídrico	Es un gas más soluble que el CO <sub>2</sub> y su origen puede ser magmático o por reacciones químicas con las rocas.
CH <sub>4</sub> Metano	Su origen está basado en la reacción Fischer-Tropsch: $CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$
H <sub>2</sub> Hidrogeno	Se origina por reacciones químicas a altas temperaturas del agua con óxidos ferrosos y silicatos contenidos en las rocas,
N <sub>2</sub> Nitrógeno	Proviene de la disolución de este gas en aguas meteóricas.

## 2.6 La temperatura como parámetro clave en la evaluación de reservas geoenergéticas

El transporte de calor desde el interior de la Tierra hacia la superficie puede ser explicado a través de tres mecanismos: conducción, convección y radiación. La conducción transmite el calor de una entidad más caliente a una más fría estando ambas en contacto. La convección sucede cuando un fluido (líquido y/o gas) se encuentra en movimiento. La radiación es el calor emitido por los cuerpos debido a su temperatura y para ello no requiere un medio material (Ledesma, 1993; Incropera & DeWitt, 1999; Llopis & Rodrigo, 2008).

Independientemente del mecanismo de transporte, cuando el calor es transferido a un fluido y este llega a la superficie manifestándose como fumarolas y/o manantiales, la medición de temperaturas nos provee información de importancia para calibrar el modelo numérico del yacimiento (Hiriart et al., 2011). Otra manera

de conocer la temperatura es mediante la perforación de pozos exploratorios, midiendo los perfiles de temperatura y presión conforme se incrementa la profundidad. En la Fig. 2.10 se muestra un perfil de temperaturas medidas en un pozo exploratorio ubicado en la zona geotérmica de Acoculco.

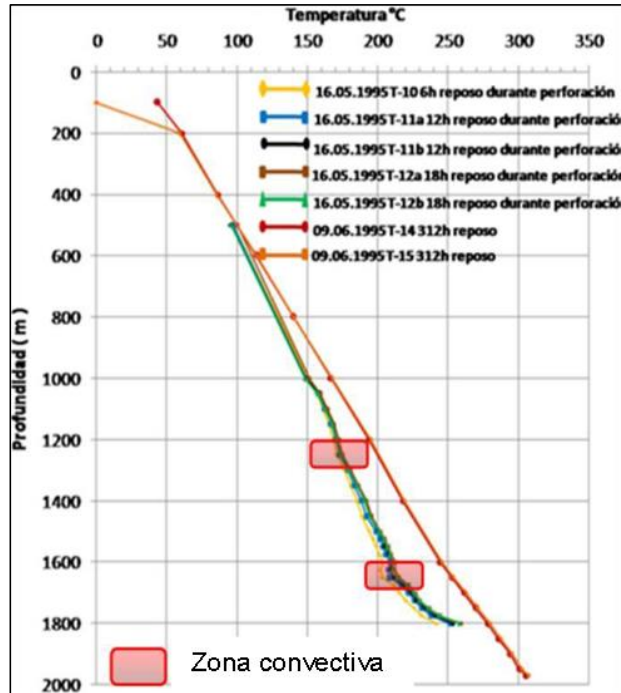


Fig. 2.10 Registros de temperatura y zonas convectivas del pozo exploratorio EAC-1 (Pulido et al., 2011).

Es importante resaltar que la perforación de pozos exploratorios nos provee de temperaturas medidas, sin embargo, la desventaja es el costo de perforación. Una alternativa más económica y que es muy utilizada en la exploración es la estimación de las temperaturas del yacimiento a través de geotermómetros. Los geotermómetros proporcionan temperaturas utilizadas durante la exploración, perforación, construcción y explotación de sistemas geotérmicos (Díaz-González et al., 2013). Con estas temperaturas podemos conocer el potencial de energía geotérmica existente del yacimiento (García-López et al., 2014). Existen diferentes tipos de geotermómetros, tales como: solutos, gases, isotópicos, mineralógicos e inclusiones fluidas. Particularmente, los geotermómetros de gases son mayormente utilizados en la predicción de altas temperaturas (Arnórsson a& Gunnlaugsson, 1985).

**Geotermómetros de gases.** Son herramientas de bajo costo que proveen una estimación de la *BHT* usando la composición química de los gases geotérmicos que emergen a la superficie (Wetzel et al., 2006). Esencialmente, son ecuaciones analíticas que utilizan la composición química de los gases muestreados *in-situ*.

En general, la calibración de los geotermómetros está desarrollada a partir de dos metodologías:

- a) Fundamentadas en bases de datos de mediciones *in-situ*, de los gases muestreados y las temperaturas de fondo medidas en pozos geotérmicos productores, bajo la suposición teórica de un equilibrio químico existente entre la interacción roca-fluido-gas en el yacimiento geotérmico. Es decir, que la temperatura depende del equilibrio químico que predomina en el yacimiento (Pang, 2001; Díaz-González et al., 2013).
- b) Experimentos de solubilidad de minerales, de datos calorimétricos o de estudios de interacción agua-roca en condiciones de equilibrio químico y termodinámico (Nicholson, 1993).

A la fecha, se han reportado alrededor de 31 geotermómetros de gases analíticos (ver Tabla 1, ANEXO 2), los cuales estiman la *BHT* como una función directa de la concentración de los gases presentes en el yacimiento.

La composición de los gases en ambientes geotérmicos está dada por: H<sub>2</sub>O, gases mayores (CO<sub>2</sub>, H<sub>2</sub>S y NH<sub>3</sub>) y gases traza (CH<sub>4</sub>, H<sub>2</sub>, N<sub>2</sub>, CO, O<sub>2</sub>). En el Capítulo 1 se resaltó la importancia del CO<sub>2</sub> dado que después del vapor de agua es el componente mayoritario ~70-80% (Giammanco et al., 2007). Así mismo, existen otros gases traza presentes (p. ej. He, Ne, Ar) que durante el muestreo pueden ser fácilmente contaminados por el aire, por lo cual las mediciones se deben interpretar con ciertas reservas (Powell & Cumming, 2010). Debido a esto, la mayoría de los geotermómetros de gases están calibrados fundamentalmente con la concentración de CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub>, N<sub>2</sub>.



Por ejemplo, D'Amore & Panichi (1980) propusieron un geotermómetro empírico calibrado en unidades de % volumen de la concentración de los gases CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub> muestreados en 34 sitios geotérmicos. Las ecuaciones geotermométricas obtenidas son:

si %CO<sub>2</sub> ≥ 75

$$T(^{\circ}C) = \frac{24775}{2\log\left(\frac{CH_4}{CO_2}\right) - 6\log\left(\frac{H_2}{CO_2}\right) - 3\log\left(\frac{H_2S}{CO_2}\right) + 36.05} - 273.15 \quad \text{Ec. (2.4)}$$

si %CO<sub>2</sub> < 75

$$T(^{\circ}C) = \frac{24775}{2\log\left(\frac{CH_4}{CO_2}\right) - 6\log\left(\frac{H_2}{CO_2}\right) - 3\log\left(\frac{H_2S}{CO_2}\right) - 7\log(0.1) + 36.05} - 273.15 \quad \text{Ec. (2.5)}$$

Para más detalles de este y otros geotermómetros de gas, consultar Tabla 3, ANEXO 2.

## CAPÍTULO 3

### Inteligencia Artificial Aplicada en Redes Neuronales

**Resumen del Capítulo.** En este Capítulo se abordan los conceptos fundamentales en el marco de la Inteligencia Artificial (IA) con sus principales vertientes. Se amplía el concepto de aprendizaje máquina y como a través de él se sitúan las redes neuronales artificiales. Se describen las generalidades de estas últimas, la forma en que están inspiradas en la neurona biológica y como el conjunto de ellas dan lugar al modelo perceptrón multicapa. Al final del Capítulo se describe un ejemplo número del algoritmo *backpropagation*, el cual ha generado múltiples aplicaciones alrededor de las redes neuronales, entre ellas la presente investigación.

#### 3.1 Generalidades de la IA

La IA es la parte de las ciencias computacionales que desarrolla sistemas que emulan características asociadas al comportamiento humano como son: la comprensión del lenguaje, el aprendizaje, el razonamiento, la resolución de problemas, entre otros (Russell & Norvig, 2002), ver Fig. 3.1. Algunos de los rasgos que se buscan obtener a través de la IA son: aprender de la experiencia, darles sentido a mensajes confusos, capacidad de responder rápidamente a situaciones desconocidas, adquirir y aplicar nuevos conocimientos y pensar-razonar.

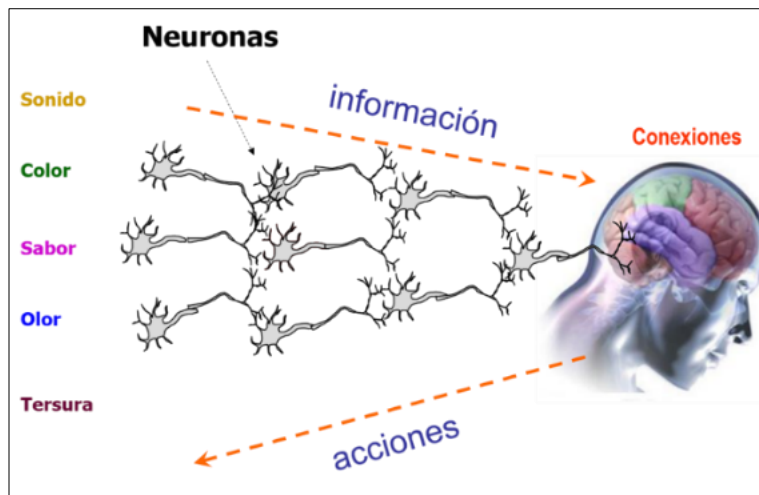


Fig. 3.1 Interpretación esquemática del proceso de información a través de diferentes estímulos en el cerebro humano (modificado de Oporto, 2006).

La IA trata de representar el pensamiento humano basándose en el desarrollo de algoritmos. Estos algoritmos permiten combinar distintos campos de la ciencia, tales como: robótica, sistema de visión, algoritmos genéticos, agentes inteligentes, sistemas expertos, aprendizaje máquina (p. ej. redes neuronales), entre otros (Jackson, 2019), ver Fig. 3.2.

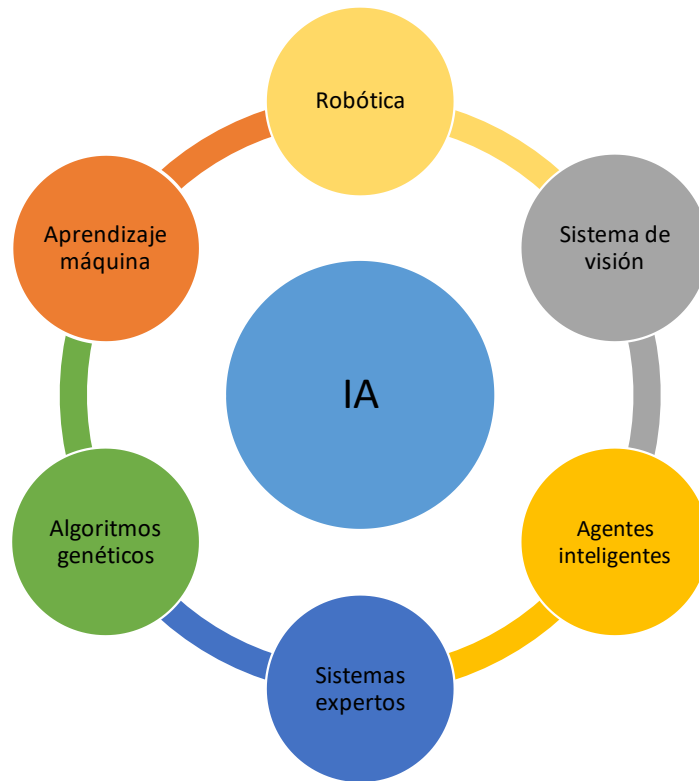


Fig. 3.2 Diagrama de los principales campos de aplicación relacionados con IA (modificado de Ponce et al., 2014).

El objetivo de la IA es crear modelos (físicos y/o virtuales) que por sí solos puedan “pensar” y tomar decisiones, lo que da como resultado un gran número de estudios y aplicaciones, dentro de las que destacan las redes neuronales (Méndez & Morales, 2008), ver Tabla 3.1.

Tabla 3.1. Principales campos de aplicación de la IA (Méndez & Morales, 2008; Jackson, 2019).

Campo	Descripción
Robótica	Dispositivo físico cuya característica es la capacidad de realizar funciones y/o tareas con un alto grado de precisión.

Campo	Descripción
Sistema de visión	Dispositivo físico y software que logra captar, almacenar y procesar imágenes. Se puede usar en conjunto con robots para proveerles de "visión" y tomar decisiones con base a lo que visualiza.
Agentes inteligentes	Software que revisa grandes cantidades de datos y la información más adecuada para el usuario.
Sistemas expertos	Software que está basado en un conjunto de reglas y procedimientos de inferencia para resolver problemas.
Algoritmos genéticos	Es un método iterativo de búsqueda basado en la probabilidad, utiliza modelos inspirados en la naturaleza para resolver problemas de optimización.
Aprendizaje máquina	Es una serie de metodologías y algoritmos tienen la capacidad de aprender a partir de la experiencia.

**Aprendizaje máquina.** Se refiere a la implementación de algoritmos que utilizan un conjunto de datos con el objetivo de identificar patrones para la predicción y optimización. Estos datos son utilizados en una etapa posterior como retroalimentación del algoritmo para mejorar las predicciones y emular un tipo de aprendizaje de las computadoras sin ser programadas directamente (Pulido et al., 2018; Bi et al., 2019). Por ejemplo, si contamos con datos acerca del comportamiento de compras de nuestros clientes, podemos hacer predicciones sobre sus futuras compras. Con base al tipo de procesamiento de los datos de entrada, estos algoritmos se subdividen en tres categorías (ver Fig. 3.3):

1. Aprendizaje supervisado: partiendo de un conjunto de datos de entrenamiento, el algoritmo produce una función de salida para hacer predicciones de los valores esperados;
2. Aprendizaje no supervisado: estos algoritmos son utilizados cuando no se tiene una salida deseada (*target*); y,
3. Aprendizaje por refuerzo: con los datos de entrenamiento, produce salidas evaluando el error o certeza de la predicción. La característica más relevante es la búsqueda de prueba-error y la recompensa si es que aplica (Méndez & Morales, 2008; Jackson, 2019).

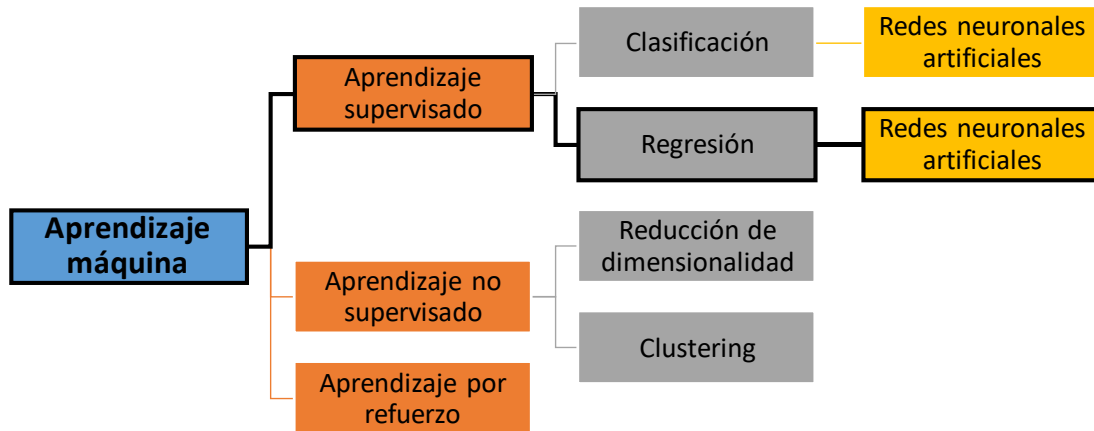


Fig. 3.3 Diagrama general de la rama aprendizaje máquina resaltando la ubicación de las redes neuronales artificiales en esta investigación.

### 3.2 Redes Neuronales Artificiales (RNA)

Las RNA son un conjunto de elementos de procesamiento (neuronas) los cuales se conectan entre sí en forma de grafo dirigido. Las conexiones entre las neuronas establecen una estructura jerárquica similar a la forma en que funciona el sistema nervioso (Zhang et al., 2019).

La analogía con las neuronas biológicas está dada por la información que recibe una neurona a través de las conexiones precedentes, procesa la información y emite una señal de salida. El primer modelo de una neurona artificial fue propuesto por McCulloch & Pitts (1943), el cual compone de varios elementos, como son:

- Un conjunto de datos de entrada  $X = (x_1, x_2, \dots, x_n)$ ;
- Cada entrada está asociada a un valor ajustable  $W = (w_1, w_2, \dots, w_n)$  denominado peso sináptico;
- Un peso sináptico adicional que posee cada neurona, conocido como umbral de acción  $\theta$  o bias;
- Una salida  $y$  cuyo valor está en función de los datos de entrada  $X$  y los pesos sinápticos  $W$ ; y,
- Una función de activación  $f$  que limita la amplitud de la salida neuronal.

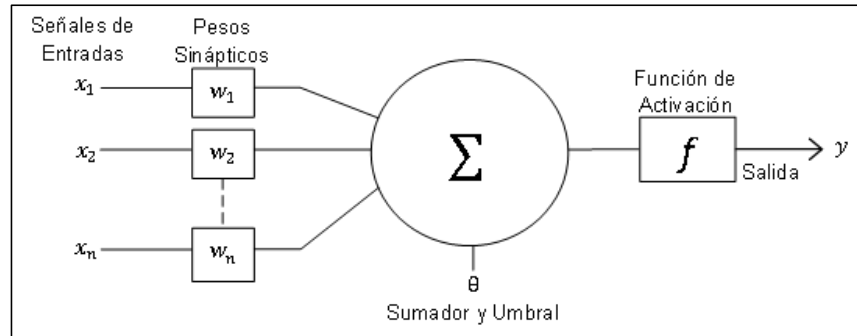


Fig. 3.4 Modelo de una neurona artificial (Modificado de Zhang et al., 2019).

En la Fig. 3.4 se está representado el modelo propuesto por McCulloch & Pitts (1943) y que matemáticamente se expresa por la Ec. 3.1.

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i + \theta\right) \quad \text{Ec. (3.1)}$$

donde  $n$  es el número de datos de entrada.

La función de transferencia  $f$  producirá siempre la misma salida si los valores precedentes no cambian, es decir, los pesos sinápticos (López & Fernández, 2008). Existen diferentes tipos de funciones de activación, cada una de ellas tiene un comportamiento específico (activada o desactivada) acorde al problema a resolver. Las más utilizadas se enlistan en la Tabla 3.2.

Tabla 3.2. Funciones de activación más utilizadas en RNA (López & Fernández, 2008).

Función	Expresión matemática	Figura
Escalón	$f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases}$	

Función	Expresión matemática	Figura
Lineal	$f(u) = u$	
Saturación	$f(u) = \begin{cases} 1, & u > 1 \\ u, & 0 \leq u \leq 1 \\ 0, & u < 0 \end{cases}$	
Logística	$f(u) = \frac{1}{1 + e^{-u}}$	
Tangente hiperbólica	$f(u) = \frac{2}{1 + e^{-2u}} - 1$	

La función de escalón y saturación tienen 2 variantes adicionales, con rangos de -1 a 1.

### 3.3 RNA Perceptrón Multicapa

De acuerdo con la topología las RNA se clasifican en: perceptrón simple, perceptrón multicapa, convolución, recurrentes y base radial. Una de las más utilizadas en Ciencias de la Tierra y otras áreas es la red neuronal perceptrón multicapa (Zhang et al., 2019).

Este tipo de RNA está conformada por diversas neuronas agrupadas en capas: capa de entrada, oculta y de salida (ver Fig. 3.5).

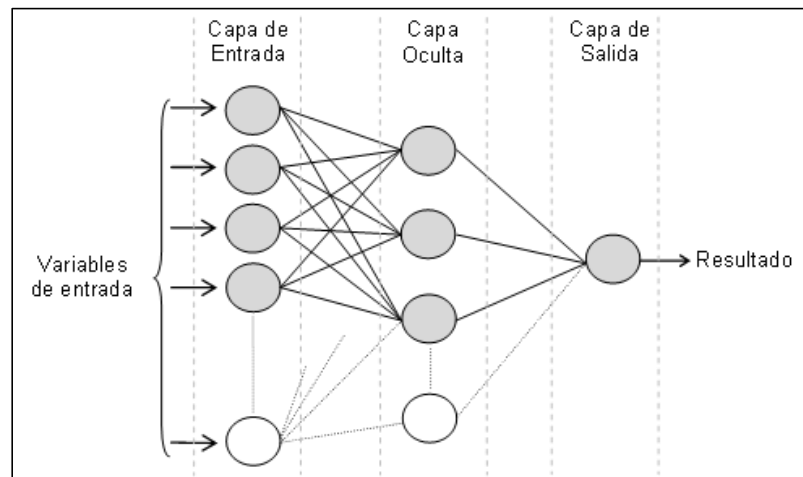


Fig. 3.5 Representación de una topología perceptrón multicapa.

La capa de entrada provee al modelo neuronal de los datos externos a ser procesados. La capa oculta realiza un primer procesamiento asociado a los pesos sinápticos de cada neurona oculta y el resultado sirve de entrada a la capa de salida. La capa de salida transfiere el resultado computado por la RNA al usuario en un entorno exterior (López & Fernández, 2008). Es importante resaltar que solo puede existir una capa de entrada y una de salida, sin embargo, puede haber un número mayor de capas ocultas. Algunas de las ventajas más importantes son:

- Aprendizaje Adaptativo. Los pesos sinápticos de la red se ajustan con datos de entrada (muestras o medidas) relacionados con el proceso a modelar, con lo que se hace posible la adquisición de conocimiento por experiencia (Da Silva et al., 2017);
- Capacidad de aprendizaje. A través de un método de aprendizaje, la red neuronal puede extraer información con respecto a la relación existente entre las variables bajo estudio (Da Silva et al., 2017); y
- Capacidad de generalización. Posterior al proceso de aprendizaje, la red neuronal es capaz de generalizar el conocimiento adquirido haciendo posible la estimación y/o predicción de soluciones del objeto bajo estudio (Da Silva et al., 2017); entre otras.



**Algoritmos de aprendizaje.** El aprendizaje es el proceso por el que una RNA modifica el valor numérico de las conexiones (pesos sinápticos) en respuesta a una información de entrada. Con ello, la red puede generalizar un problema utilizando un conjunto de datos de entrada como ejemplos (López & Fernández, 2008). Simplificando el esquema de aprendizaje de una neurona (topología perceptrón simple), los pesos se pueden determinar con la siguiente expresión: “*Peso Nuevo = Peso viejo + Cambio de Peso*”. Matemáticamente:

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}(t) \quad \text{Ec. (3.2)}$$

donde  $t$  hace referencia a la etapa de aprendizaje;  $w_{ij}(t + 1)$  al peso nuevo y  $w_{ij}(t)$  al peso viejo. Los subíndices  $ij$  es el  $i$ -ésimo peso de entrada a la  $j$ -ésima neurona.

$$\Delta w_{ij} = \alpha \cdot x_i(yd_j - y_j) \quad \text{Ec. (3.3)}$$

donde  $\Delta w_{ij}$  es la variación en el peso entre las neuronas  $i$  y  $j$ ;  $x_i$  es la  $i$ -ésima entrada a la  $j$ -ésima neurona;  $yd_j$  es el valor de la salida deseada para la neurona  $j$ ;  $y_j$  el valor de salida de la  $j$ -ésima neurona;  $\alpha$  es un factor o tasa de aprendizaje que regula su el ajuste del peso sináptico.

El aprendizaje de las RNA se subdivide en supervisado y no supervisado. En la siguiente sección se describe el algoritmo supervisado *backpropagation*, el cual sirvió de base para futuros modelos de optimización en el aprendizaje de las RNA.

### 3.4 Algoritmo *Backpropagation*

El creador de este algoritmo fue Paul J. Werbos quien en 1974 publicó la idea básica de propagación hacia atrás o mejor conocido como *Backpropagation* en su tesis denominada “*Beyond regression: new tools for prediction and analysis in the behavioral sciences*”. Sin embargo, no fue sino hasta más de una década después que este algoritmo fue redescubierto por Rumelhart et al. (1986).

Al igual que el perceptrón simple (ver Ecs. 3.2 y 3.3), este algoritmo ajusta los pesos sinápticos con base al error obtenido a la salida de la RNA al ser comparada con la salida deseada. El ajuste de los pesos lo realiza con el método de descenso del gradiente comenzando por la capa de salida (acorde al error cometido), y continúa propagando este error a las capas anteriores hasta llegar a la capa de entrada (López & Fernández, 2008). En general, este algoritmo consta de 2 fases:

1. En la propagación hacia adelante se aplica un patrón a las entradas de la red y su efecto se propaga a través de esta, capa a capa. Al llegar a la última capa, la red presenta un conjunto de salidas como respuesta a dicho patrón de entrada. Normalmente, en la primera iteración el algoritmo parte de un conjunto de pesos con un valor inicial pequeño aleatorio (Méndez & Morales, 2008).
2. En la propagación hacia atrás los pesos de la red se modifican de acuerdo con una regla de ajuste del error. Es decir, se calcula el valor de la función de error que compara la salida actual de la red y la salida deseada. Este error se propaga hacia atrás (Méndez & Morales, 2008).

Esquemáticamente, el algoritmo backpropagation está representado en la Fig. 3.6 donde se aprecia el inicio del ciclo e iterativamente continúa hasta lograr la convergencia.

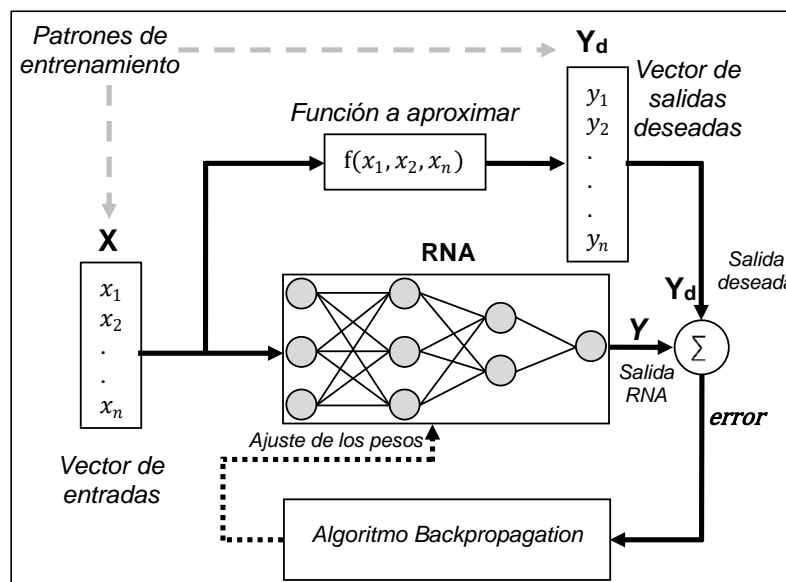
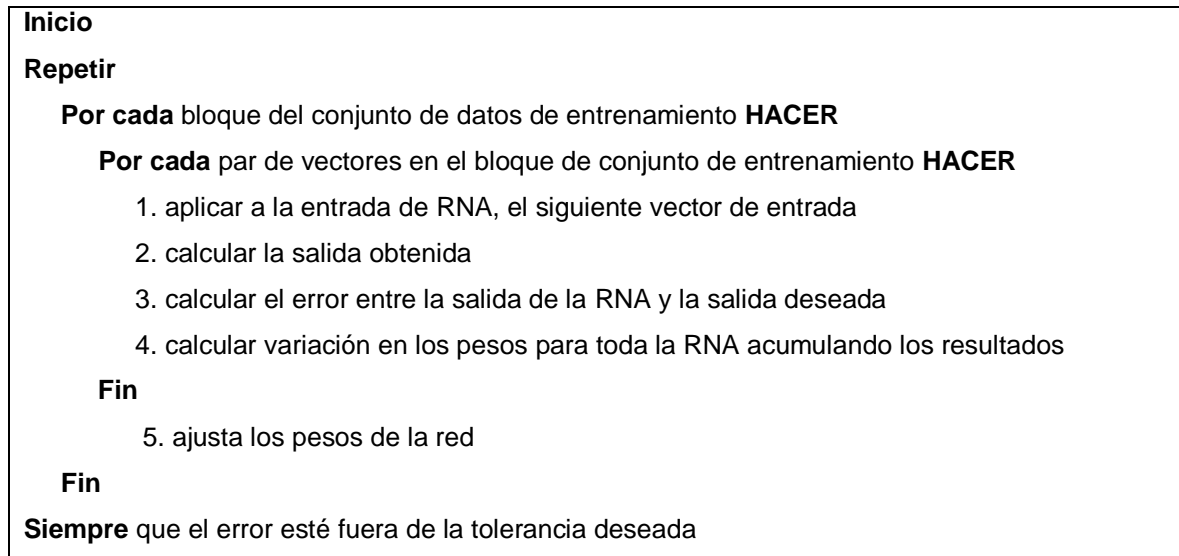


Fig. 3.6 Diagrama representativo del entrenamiento de una RNA (Modificado de Díaz-González et al., 2013).

El pseudocódigo del algoritmo *backpropagation* (de Lima, 2005) es:



### 3.5 Ejemplo numérico de algoritmo *Backpropagation*

Para comprender el efecto del procesamiento de cada neurona propagándose hasta calcular la salida de la red, en el presente trabajo se muestra un ejemplo completo tomado de Mazur (2018). La arquitectura utilizada en el ejemplo es de 2 neuronas en la capa de entrada, 2 en la capa oculta y 2 en la salida (2-2-2). Las 2 neuronas de entrada están asociadas a 2 variables con valores de 0.05 y 0.1. Estos valores servirán de insumo a la red neuronal para obtener dos *targets* con valores de 0.01 y 0.99. Esquemáticamente el ejemplo se representa con la Fig. 3.7.

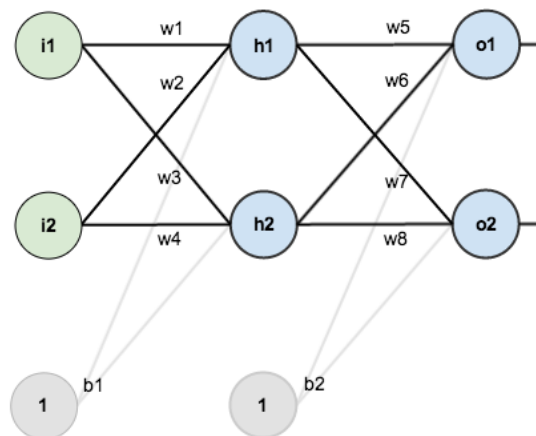


Fig. 3.7 Esquema del ejemplo numérico, donde I1 e I2 son las neuronas de entrada; h1 y h2 con las neuronas ocultas; o1 y o2 son las neuronas de salida y w1 a w8 son los pesos sinápticos.

Como se mencionó anteriormente en la sección 3.4, los pesos sinápticos (conexiones entre las neuronas) se inicializan con valores aleatorios como se muestra en la Fig. 3.8.

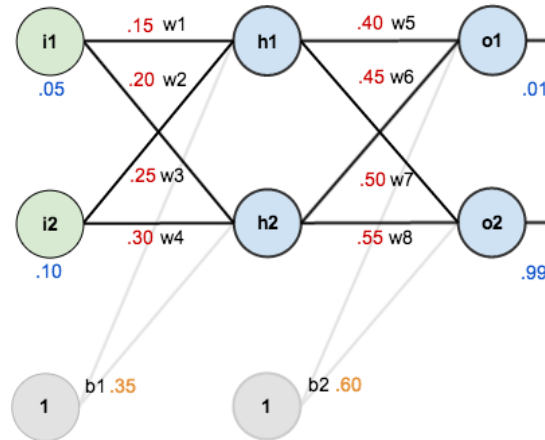


Fig. 3.8 Inicialización de los pesos sinápticos con números reales.

El objetivo del algoritmo es ajustar los pesos para que la red neuronal pueda aprender a predecir la salida deseada, es decir, dadas las entradas 0.05 y 0.10, queremos que la red neuronal produzca 0.01 y 0.99.

**Propagación hacia adelante.** Inicialmente calculemos que valores predice actualmente la red neuronal dados los pesos de la Fig. 3.8 en conjunto con las entradas de 0.05 y 0.10. Para ello, calculamos la entrada neta total a cada neurona de la capa oculta, posteriormente limitamos la entrada neta total usando una función de activación (función logística), luego repetimos el proceso con las neuronas de la capa de salida.

Para calcular la entrada neta de la neurona oculta  $h_1$ :

$$net_{h_1} = w_1 * i_1 + w_2 * i_2 + b_1 * 1 \quad \text{Ec. (3.4)}$$

Sustituyendo tenemos:

$$net_{h_1} = (0.15 * 0.05) + (0.2 * 0.10) + (0.35 * 1) = 0.3775$$

Aplicando la función logística (ver Tabla 3.2) a la salida de la neurona oculta  $h_1$ :

$$out_{h1} = \frac{1}{1 + e^{-net_{h1}}} = \frac{1}{1 + e^{-0.3775}} = 0.593269992$$

Para calcular la entrada neta de la neurona oculta h2:

$$net_{h2} = w_3 * i_1 + w_4 * i_2 + b_1 * 1 \quad \text{Ec. (3.5)}$$

Sustituyendo tenemos:

$$net_{h2} = (0.25 * 0.05) + (0.3 * 0.10) + (0.35 * 1) = 0.3925$$

Aplicando la función logística:

$$out_{h2} = \frac{1}{1 + e^{-net_{h2}}} = \frac{1}{1 + e^{-0.3925}} = 0.596884378$$

Para calcular la entrada neta de la neurona de salida o1 Ec. (3.6)

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

Sustituyendo tenemos:

$$net_{o1} = (0.4 * 0.5932) + (0.45 * 0.5968) + (0.6 * 1) = 1.105905967$$

Aplicando la función logística:

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}} = \frac{1}{1 + e^{-1.1059}} = 0.751365069$$

Para calcular la entrada neta de la neurona de salida o2 Ec. (3.7)

$$net_{o2} = w_7 * out_{h1} + w_8 * out_{h2} + b_2 * 1$$

Sustituyendo tenemos:

$$net_{o2} = (0.5 * 0.5932) + (0.55 * 0.5968) + (0.6 * 1) = 1.224921404$$

Aplicando la función logística:

$$out_{o2} = \frac{1}{1 + e^{-net_{o2}}} = \frac{1}{1 + e^{-1.2249}} = 0.772928465$$

Para calcular el error total (usando la función del error al cuadrado) cometido por cada neurona de salida

$$E_{total} = \sum \frac{1}{2} (target - output)^2 \quad \text{Ec. (3.8)}$$

Sustituyendo tenemos:

$$E_{o1} = \frac{1}{2} (target_{o1} - out_{o1})^2 = \frac{1}{2} (0.01 - 0.7513)^2 = 0.274811082$$

$$E_{o2} = \frac{1}{2} (target_{o2} - out_{o2})^2 = \frac{1}{2} (0.99 - 0.7729)^2 = 0.023560025$$

$$E_{total} = E_{o1} + E_{o2} = 0.2748 + 0.02356 = 0.298371107$$

**Propagación hacia atrás.** El objetivo es ajustar cada uno de los pesos en la red para que la salida sea más cercana al valor deseado. Para ello, minimizamos el error cometido en cada neurona. Por ejemplo, ¿cuánto afecta un cambio al peso en  $w_5$  con respecto al error total  $\frac{\partial E_{total}}{\partial w_5}$  ?

Aplicando la regla de la cadena para conocer que tanto  $w_5$  afecta al error total

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5} \quad \text{Ec. (3.9)}$$

Resolviendo primero para el error total  $E_{total}$  con respecto a la neurona de salida o1

$\partial out_{o1}$  tenemos:

$$E_{total} = \frac{1}{2} (target_{o1} - out_{o1})^2 + \frac{1}{2} (target_{o2} - out_{o2})^2$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2} (target_{o1} - out_{o1})^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.7513) = 0.741365069$$

Cuando tomamos la derivada parcial del error total con respecto a  $out_{o1}$  el término  $\frac{1}{2} (target_{o2} - out_{o2})^2$  se convierte en cero porque  $out_{o1}$  no lo afecta, lo que significa que estamos tomando la derivada de una constante que es cero.

Ahora bien, ¿cuánto cambia la salida de la neurona o1 con respecto a su entrada neta total? La derivada parcial de la función logística es la salida multiplicada por 1 menos la salida:

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}}$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1} (1 - out_{o1}) = 0.7513 (1 - 0.7513) = 0.186815602$$

Continuando, debemos evaluar ¿cuánto cambia la entrada neta total de la neurona o1 con respecto al peso  $w_5$ ? Tenemos que la Ec. (3.7) es:

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$\frac{\partial net_{o1}}{\partial w_5} = 1 * out_{h1} * w_5^{(1-1)} + 0 + 0 = out_{h1} = 0.593269992$$

Con todos los valores calculados, sustituimos en Ec. (3.9)

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5}$$

$$\frac{\partial E_{total}}{\partial w_5} = 0.7413 * 0.1868 * 0.5932 = 0.08216704$$

Finalmente, para ajustar el valor del peso actual de  $w_5$  y calcular el peso siguiente  $w_5^+$ , restamos el valor obtenido por la regla de la cadena. Opcionalmente multiplicado por una tasa de aprendizaje, p. ej.  $\alpha = 0.5$

$$w_5^+ = w_5 - \alpha * \frac{\partial E_{total}}{\partial w_5} = 0.4 * 0.5 * 0.08216 = 0.35891648$$

Al repetir el procedimiento que inicia con la Ec. (3.9) y calcular todos los pesos asociados a las neuronas de salida ( $w_6, w_7, w_8, b_2$ ) obtenemos los siguientes valores:

$$w_6^+ = 0.408668997; w_7^+ = 0.51130127; w_8^+ = 0.561370121;$$

$$b_2^+ = \{0.530750719, 0.619049094\}$$

Para ajustar los pesos asociados a las neuronas de la capa oculta se realiza un procedimiento similar al de la Ec. (3.9). Considerando que la salida de cada neurona oculta contribuye al valor calculado por las neuronas de salida ( $o1$  y  $o2$ ), tenemos que para el peso  $w_1$

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1}$$

Ec. (3.10)

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

Con los valores previamente calculados y comenzando por  $\frac{\partial E_{o1}}{\partial out_{h1}}$  tenemos:

$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}}$$

$$\frac{\partial E_{o1}}{\partial net_{o1}} = \frac{\partial E_{o1}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} = 0.741365069 * 0.186815602 = 0.138498561$$

$$\frac{\partial net_{o1}}{\partial out_{h1}} = w_5 = 0.4$$

Sustituyendo tenemos:

$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}} = 0.138498561 * 0.4 = 0.055399424$$

Con los valores previamente calculados y continuando por  $\frac{\partial E_{o2}}{\partial out_{h1}}$  tenemos:

$$\frac{\partial E_{o2}}{\partial out_{h1}} = \frac{\partial E_{o2}}{\partial net_{o2}} * \frac{\partial net_{o2}}{\partial out_{h1}}$$

$$\frac{\partial E_{o2}}{\partial net_{o2}} = \frac{\partial E_{o2}}{\partial out_{o2}} * \frac{\partial out_{o2}}{\partial net_{o2}} = -0.217071535 * 0.1755100529 = -0.038098236$$

$$\frac{\partial net_{o2}}{\partial out_{h1}} = w_7 = 0.5$$

Sustituyendo tenemos:

$$\frac{\partial E_{o2}}{\partial out_{h1}} = \frac{\partial E_{o2}}{\partial net_{o2}} * \frac{\partial net_{o2}}{\partial out_{h1}} = -0.038098236 * 0.5 = -0.019049118$$

El error calculado por la neurona oculta  $h_1$  en la Ec. (3.10) es:

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}} = 0.055399424 + (-0.019049118) = 0.036350306$$

Continuando con los términos  $\frac{\partial out_{h1}}{\partial net_{h1}}$  y  $\frac{\partial net_{h1}}{\partial w_1}$  de la Ec. (3.10) tenemos:

$$\frac{\partial out_{h1}}{\partial net_{h1}} = out_{h1}(1 - out_{h1}) = 0.5932 * (1 - 0.5932) = 0.241300708$$

$$\frac{\partial net_{h1}}{\partial w_1} = i_1 = 0.05$$

Sustituyendo todos los valores en Ec. (3.10) tenemos:

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1} = 0.03635 * 0.2413 * 0.05 = 0.0004385677$$

Para ajustar el valor del peso actual de  $w_1$  y calcular el peso siguiente  $w_1^+$ , repetimos el procedimiento utilizado en  $w_5^+$

$$w_1^+ = w_1 - \alpha * \frac{\partial E_{total}}{\partial w_1} = 0.15 - 0.5 * 0.0004385 = 0.149780716$$

Al repetir el procedimiento que inicia con la Ec. (3.10) y calcular todos los pesos asociados a las neuronas ocultas ( $w_2, w_3, w_4, b_1$ ) obtenemos los siguientes valores:

$$w_2^+ = 0.199561432; w_3^+ = 0.249751143; w_4^+ = 0.299502287;$$

$$b_1^+ = \{0.345614323, 0.345022873\}$$

A través del algoritmo *backpropagation* se han actualizado todos los pesos sinápticos en este ejemplo. En la propagación hacia adelante el error obtenido por la red neuronal con los pesos aleatorios iniciales (ver Fig. 3.8) fue igual a 0.298371107. Después de ajustar los pesos en una época, el error disminuyó ligeramente obteniendo un valor de 0.291027924.

Así mismo, al ajustar los pesos durante 50 mil épocas el error disminuye a 0.0000000001 y la salida de la red neuronal genera valores de 0.010010 y 0.989991 siendo los valores objetivo de 0.01 y 0.99 respectivamente.



## CAPÍTULO 4

### Metodología y Desarrollo del Proyecto de Investigación

**Resumen del Capítulo.** En este Capítulo se describe la metodología y las bases de datos utilizadas para desarrollar los objetivos específicos planteados en el Capítulo 1, mismos que se subdividieron en tres módulos: Módulo 1. Desarrollar nuevos geotermómetros de gases para la predicción de temperaturas de fondo; Módulo 2. Desarrollar nuevos modelos de predicción para modelar anomalías o emisiones naturales de CO<sub>2</sub> en sistemas geotérmicos y, Módulo 3. Desarrollar una nueva metodología geoquimiométrica computacional apoyada en el método de Sinclair para la discriminación del transporte de gas en sistemas geotérmicos.

En la Fig. 4.1 se muestra un diagrama general de los problemas específicos que se abordaron en la presente investigación y de los cuales se derivaron los tres módulos a desarrollar.

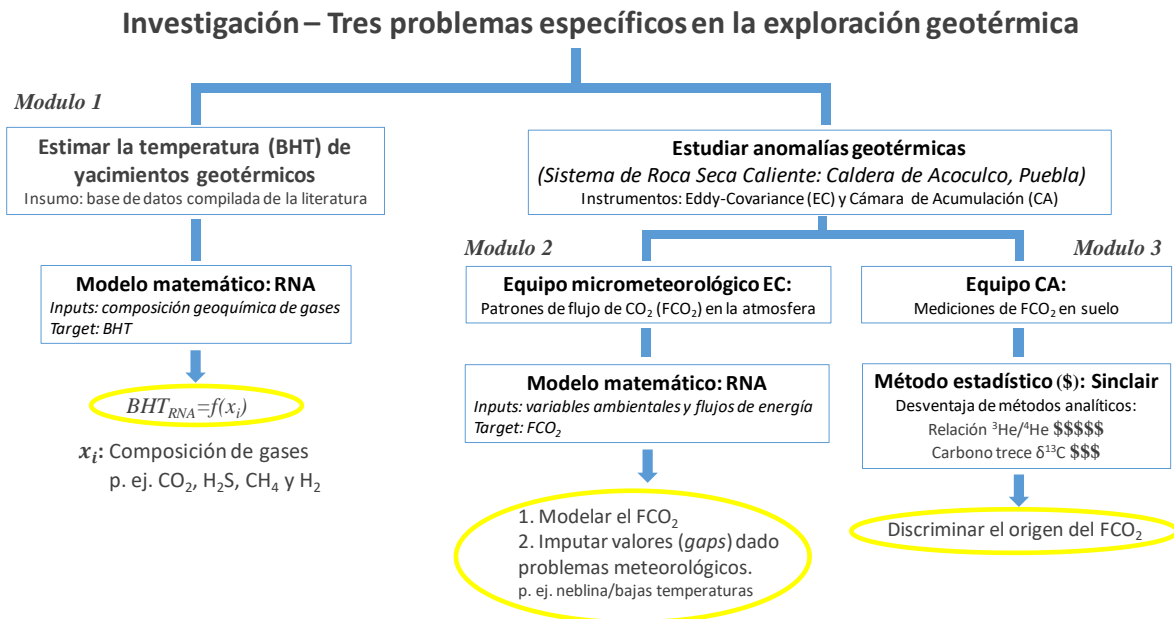


Fig. 4.1 Diagrama general de los objetivos de la presente investigación.

## 4.1 Metodología del Módulo 1

El desarrollo de nuevos geotermómetros de gases para la predicción de temperaturas de fondo comprende una serie de pasos secuenciales, los cuales se describen a detalle en el ANEXO 1 y 2. A continuación se enlistan de forma general cada uno de los procesos utilizados.

1. Compilación de datos de la literatura: revistas indizadas, congresos, reportes técnicos, entre otros. La información compilada comprende la composición geoquímica de gases geotérmicos medidos en pozos productores. Así como la temperatura de fondo (*BHT*). Dentro de los gases mínimos requeridos para agregar un registro completo a la Base de Datos Geoquímicos Mundial (BDGM), deberá contener las concentraciones de CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub>, ver Tabla 4.1.

Tabla 4.1 Extracto de la BDGM con 14 muestras geoquímicas (bases de datos completas: consultar material suplementario de ANEXO 1).

<i>n</i>	País	Pozo geotérmico	Campo geotérmico	Fecha	<i>BHT<sub>m</sub></i>	CO <sub>2</sub>	H <sub>2</sub> S	NH <sub>3</sub>	CH <sub>4</sub>	H <sub>2</sub>	N <sub>2</sub>
1	Iceland	HE-03	Hellisheidi	n.r.	275	680	220	n.r.	3.5	39.0	58.1
2	Iceland	HE-05	Hellisheidi	n.r.	275	846	65	n.r.	5.9	5.3	78.4
3	Iceland	HE-06	Hellisheidi	n.r.	270	828	102	n.r.	1.0	61.3	7.3
4	Italy	1	Larderello	1970-1976	250	941	19	n.r.	14.2	19.6	7.4
5	Italy	2	Larderello	1970-1976	250	940	20	n.r.	13.2	20.6	6.9
6	Italy	3	Larderello	1971-1977	260	958	14	n.r.	6.9	15.3	6.0
7	Japan	TT-2	Takigami	n.r.	205	757	70	n.r.	9.0	2.0	159.0
8	Japan	TT-7	Takigami	n.r.	217	847	90	n.r.	7.7	0.6	53.0
9	Japan	TT-8	Takigami	n.r.	215	718	59	n.r.	5.1	1.0	210.0
10	Kenya	OW-901	Olkaria	2008	338	864	90	n.r.	2.0	24.5	19.6
11	Kenya	OW-903	Olkaria	2008	343	904	7	n.r.	7.1	1.7	79.6
12	México	M-105	Cerro Prieto	1982	325	828	59	18.8	36.9	52.2	7.9
13	México	M-105	Cerro Prieto	1982	325	818	53	11.2	44.3	63.4	9.5
14	México	M-11	Cerro Prieto	1977	281	713	57	39.8	42.6	17.7	110.0

Las unidades de *BHT* son en °C y las concentraciones de los gases en mmol/mol base seca.

2. En la BDGM compilada por García-López (2016) se recolectaron 591 muestras provenientes de 11 campos geotérmicos en 6 países a las cuales se les aplicó un procedimiento de validación a través de una serie de controles de calidad a los datos, tales como: (i) de acuerdo a las unidades de concentración utilizadas

mmol/mol base seca, considerar que la suma total de los gases no exceda el  $\pm 5\%$ , es decir,  $1000 \pm 50$  mmol/mol; (ii) la temperatura de la muestra no exceda el intervalo  $170 \text{ }^\circ\text{C} \leq BHT \leq 365 \text{ }^\circ\text{C}$ ; (iii) la concentración de  $\text{CO}_2$  deberá ser  $\geq 70\%$  de la concentración de todos los gases; y, (iv) verificar existencia de registros duplicados.

Dado que la medición de la *BHT* es muy costosa durante la exploración y explotación, existen diversas muestras geoquímicas provenientes de un mismo campo geotérmico y pozo, donde la concentración de los gases era variable, pero compartían la misma *BHT*. Esto generó una evaluación de todas esas muestras mediante pruebas estadísticas para detectar datos discordantes (*outliers*) a través de pruebas univariadas y multivariadas (detalles en sección 2.2.3, ANEXO 1). Como resultado se obtuvo una sub-base de datos muy pequeña con 97 muestras geoquímicas y las otras dos sub-bases de datos con 527 y 498 muestras (ver Fig. 4.2).

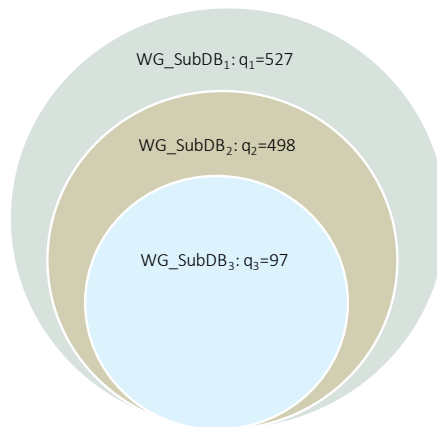


Fig. 4.2 Tres sub-bases de datos a utilizar durante el entrenamiento con 527, 498 y 97 muestras geoquímicas de gases.

- Se desarrolló un algoritmo de entrenamiento optimizado para evaluar múltiples arquitecturas de RNA con las siguientes consideraciones: (i) se diseñaron 13 combinaciones diferentes de las concentraciones de los gases geotérmicos para servir como variables de entrada (ver Tabla 7, ANEXO 1). Por ejemplo, combinación no. 1: variable de entrada  $\ln(\text{CO}_2)$ , variable objetivo (*target*  $\rightarrow$  *BHT*); (ii) la normalización de los datos de entrada y del *target* se hizo entre -1

y 1 de acuerdo a la función de activación tangente hiperbólica; (iii) los datos se dividieron en tres conjuntos 80% para entrenamiento, 10% para validación (evitar el sobre-ajuste de la RNA) y 10% para prueba; (iv) se utilizó un modelo de RNA con 3 capas: una capa de entrada, una oculta y una de salida. La capa de entrada fue variando el número de neuronas conforme a las 13 combinaciones descritas en la Tabla 7 del ANEXO 1 siendo el mínimo 1 neurona y el máximo 5, las neuronas de la capa oculta fueron variando desde 1 hasta 35 y en la capa de salida solo se utilizó una neurona en todas las arquitecturas para estimar la *BHT*; (v) se reguló la tasa de aprendizaje con 4 valores diferentes: 0.01, 0.001, 0.0001 y 0.00001; (vi) para disminuir el error se utilizó el algoritmo Levenberg-Marquardt (LM), más detalles en Ecs. 2-10 del ANEXO 1; (vii) para evaluar el desempeño de cada arquitectura entrenada, se calcularon diferentes métricas estadísticas tales como: el coeficiente de correlación  $r$ , RMSE, MAE (ver Tabla 9, ANEXO 1); y, (viii) se calculó la contribución relativa de cada variable de entrada con base al método de Garson (1991). Este algoritmo de entrenamiento se encuentra disponible en forma de *scripts* de Matlab en el siguiente repositorio: <https://github.com/ANNGroup/GasG-Scripts-ANNs.git>

4. Posteriormente al entrenamiento de las RNA, se obtuvieron 39 arquitecturas (ver Tablas 10-12, ANEXO 1) provenientes de las 3 sub-bases de datos, es decir, 13 arquitecturas por cada sub-base. Si bien es cierto que los resultados obtenidos durante el entrenamiento neuronal fueron aceptables, se identificaron algunas limitaciones de aplicabilidad fundamentalmente en las RNA que utilizaron un conjunto de datos muy pequeño durante el proceso de aprendizaje ( $WG\_SubDB_3: q_3 = 97$ ). Esto se debe a la baja representatividad de esta sub-base y se refleja en rangos de aplicabilidad muy limitados. Para abordar estas restricciones y poder seleccionar las RNA más confiables de entre las 39 arquitecturas provenientes de las 3 sub-bases de datos (ver Tablas 10-12, ANEXO 1) se aplicó la metodología Análisis de Decisión Multi-Criterio o (*MCDA* por siglas en inglés: *Multi-Criteria Decision Analysis*).

5. La evaluación integral del desempeño y capacidad de predicción de las 39 arquitecturas con nuevos datos geoquímicos (no utilizados durante el entrenamiento), fue desarrollada con la metodología *MCDA*. Con ello, se busca obtener los mejores modelos predictivos y proponer los nuevos geotermómetros de gases basados en RNA. Esta metodología permite crear diferentes escenarios con base a métricas estadísticas y evaluar los mejores modelos predictivos como se muestra en la Fig. 4.3.

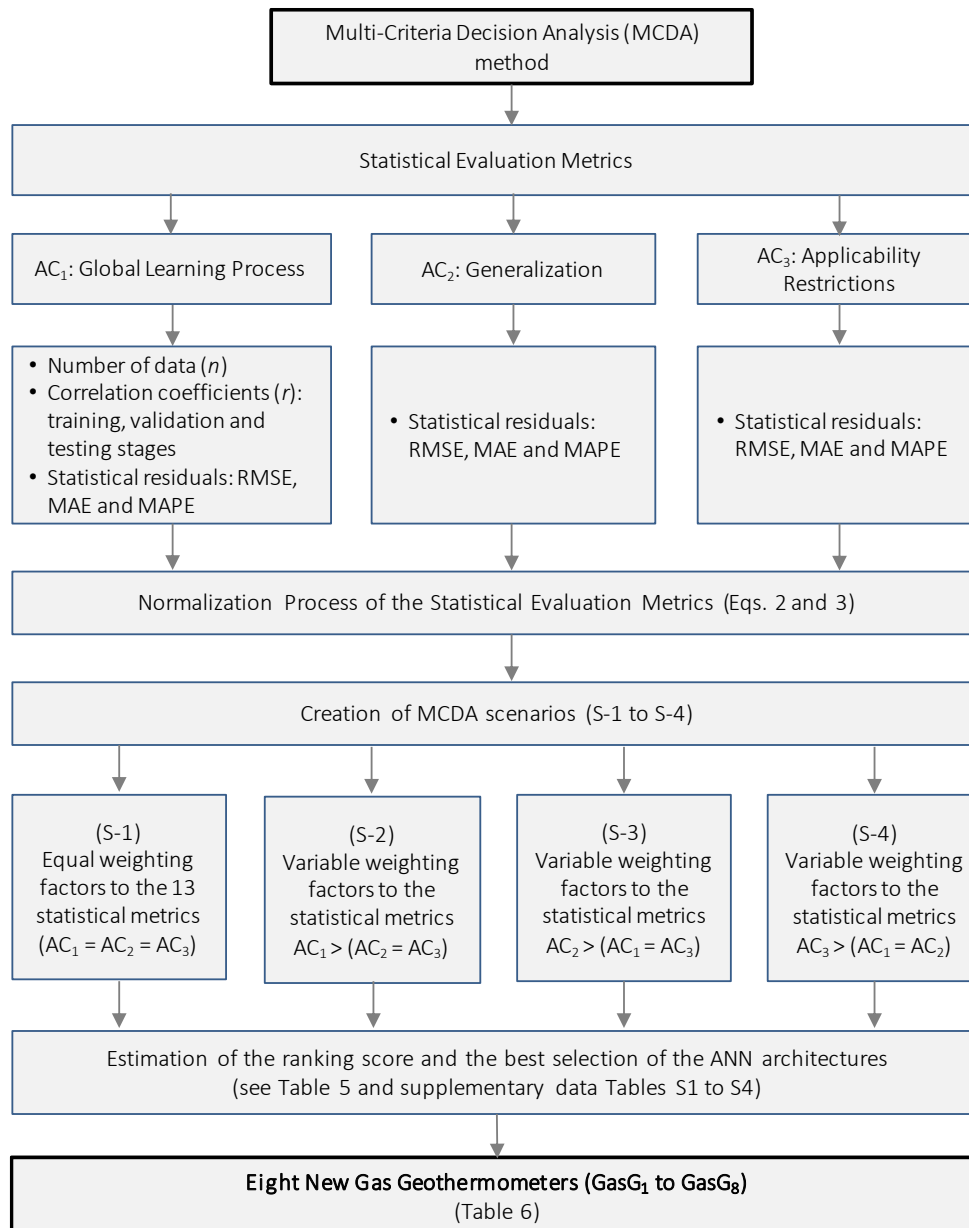


Fig. 4.3 Diagrama de flujo esquemático utilizado por el método *MCDA*.  
(Fuente: Fig. 1b ANEXO 2; más detalles en material suplementario del mismo ANEXO en Tablas S1-S4).

Hasta nuestro conocimiento, en la presente investigación se utilizó el método *MCDA* por primera vez en la literatura de las RNA para evaluar la eficiencia de predicción de distintas arquitecturas neuronales. Los resultados de las métricas estadísticas (coeficiente de correlación ( $r$ ), RMSE, MAE y MAPE) se dividieron en 3 casos de análisis representados en la Fig. 3.4 como AC<sub>1</sub> - AC<sub>3</sub>: AC<sub>1</sub> el proceso global de aprendizaje, AC<sub>2</sub> generalización de las RNA y AC<sub>3</sub> restricciones de aplicabilidad:

AC<sub>1</sub>. Implica los procesos de aprendizaje global de las tres sub-bases de datos (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub> y WG\_SubDB<sub>3</sub>) aplicando las 39 RNA utilizando las siguientes métricas estadísticas: el número de datos utilizados por cada sub-base de datos ( $q_1 = 527$ ;  $q_2 = 498$  y  $q_3 = 97$ ), los coeficientes de correlación ( $r$ ) obtenidos entre las temperaturas medidas (BHT<sub>m</sub>) y simuladas (BHT<sub>ANN</sub>) para las etapas de entrenamiento, validación y prueba, y los residuales calculados entre temperaturas medidas y simuladas (RMSE, MAE y MAPE);

AC<sub>2</sub>. Implica el cálculo de las temperaturas de las 39 RNA a la sub-base de datos más representativa ( $q_1 = 527$ ), como un caso de generalización utilizando los residuales obtenidos (RMSE, MAE y MAPE) entre temperaturas medidas y predichas; y

AC<sub>3</sub>. La aplicación de las 39 RNA a la misma sub-base de datos ( $q_1 = 527$ ) pero altamente restringido por las condiciones de aplicabilidad de cada arquitectura neuronal utilizando los mismos residuales entre temperaturas medidas y predichas.

Previo a la aplicación del *MCDA*, es necesario normalizar las métricas estadísticas (Tabla 4, ANEXO 2) con base a su representatividad, por ejemplo: en el caso del RMSE, MAE y MAPE la mejor eficiencia de predicción de las arquitecturas neuronales ocurre cuando los valores de estas métricas son bajos. Por otro lado, en el caso del coeficiente de correlación lineal ( $r$ ) los valores más altos cercanos a la unidad son los ideales. En este contexto, la normalización para minimizar y maximizar las métricas estadísticas se realizó con las Ec. 2 y 3 del ANEXO 2.

Continuando con el método *MCDA* y evaluar integralmente las 39 RNA, se crearon 4 distintos escenarios (S-1 a S-4, ver Fig. 4.3) asignando valores de ponderación variable a las métricas estadísticas para posteriormente determinar el puntaje parcial promedio de cada arquitectura (ver Tabla 5, ANEXO 2). Detalles del procedimiento publicado en material suplementario de ANEXO 2 Tablas S1 a S4.

6. Se compiló de la literatura 25 geotermómetros de gases analíticos simples cuya estimación de la temperatura es una función directa de la concentración de los gases geotérmicos (ver Tabla 3, ANEXO 2).
7. Para evaluar la capacidad de predicción de los nuevos geotermómetros de gases y de los 25 reportados en la literatura, se compiló una Nueva Base de Datos Geoquímica Mundial (NGWDB = 265 muestras geoquímicas de gases y *BHT*).
8. De acuerdo con su procedencia, las muestras geoquímicas de la NGWDB fueron separadas en sistemas de líquido— y vapor—dominante, con 178 y 87 muestras, respectivamente.
9. Considerando las características de cada geotermómetro de gas, se realizaron las conversiones necesarias de la composición gaseosa acorde a su calibración, ver unidades de concentración en Tabla 3, ANEXO 2.
10. Para el uso de los nuevos geotermómetros de gas basados en RNA, se desarrolló una aplicación en el lenguaje de programación Java denominado *GaS\_GeoT*. Este programa utiliza la composición gaseosa de CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub> en unidades de mmol/mol base seca concentrados en una plantilla de Microsoft Excel, disponible el programa y manual de usuario en: [https://github.com/ANNGroup/GaS\\_GeoT.git](https://github.com/ANNGroup/GaS_GeoT.git)
11. Con las temperaturas calculadas por *GaS\_GeoT* y los 25 geotermómetros de la literatura, se realizó un análisis geoquimiométrico utilizando diferentes parámetros, tales como: DIFF%, RMSE, MAE, MAPE y Theil's U (ver Tabla 9, ANEXO 2).

Los resultados obtenidos durante el desarrollo metodológico del Módulo 1, se presentarán en el Capítulo 5.

## 4.2 Metodología del Módulo 2

De forma similar a la sección 4.1, el desarrollo de este módulo se describe en el ANEXO 3. A continuación se enlistan de forma general los pasos para desarrollar nuevos modelos de predicción para modelar anomalías o emisiones naturales de CO<sub>2</sub> en sistemas geotérmicos. Resaltando que estas actividades fueron desempeñadas por el grupo de investigación en geotermia (con el cual colaboré en el Proyecto P09-CeMIE-Geo).

1. Se seleccionó la zona conocida como Los Azufres dentro de la Caldera de Acoculco (Puebla) para medir los flujos de CO<sub>2</sub>. Esta locación fue elegida dadas las condiciones geológicas del lugar, como son: emisión de mofetas de gas frío, manantiales burbujeantes, evidencia hidrogeológica en las rocas de la zona, entre otras.
2. Se instaló un equipo micrometeorológico denominado *Eddy-Covariance* (EC) el cual mide además del flujo de CO<sub>2</sub>, otras variables ambientales, así como los flujos de energía, p. ej. la temperatura del aire, la humedad, la radiación solar, entre otras (ver Fig. 4.4).

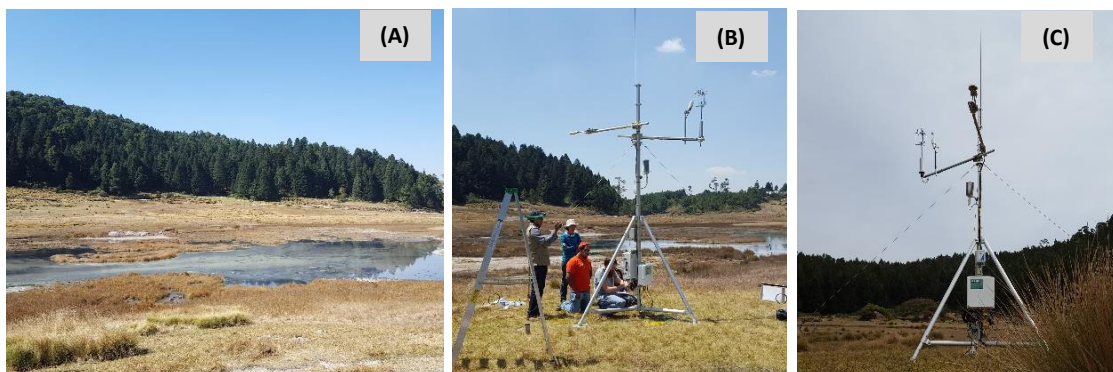


Fig. 4.4 (A) Zona de Los Azufres en Acoculco, Puebla; (B) Instalación del equipo *Eddy-Covariance*; y, (C) Puesto en operación para medir el flujo de CO<sub>2</sub> atmosférico.

3. Se creó una Base de Datos con las mediciones del equipo micrometeorológico (BDEddy), la cual contiene más de 100 variables medidas ver Tabla 4.2



Tabla 4.2 Extracto de la BDEddy con 9 registros del flujo de CO<sub>2</sub> y las variables ambientales que más se correlacionan con él de acuerdo con Melesse & Hanley (2005).

n	Date	Time	H	LE	RN	TA	TS	Tdew	CO <sub>2</sub> flux
1	15/03/2016	21:30:00	-12.650	5.840	-62.703	8.875	16.445	3.399	4.057
2	15/03/2016	22:00:00	-9.891	2.450	-60.459	8.066	16.136	1.122	2.093
3	15/03/2016	22:30:00	-9.023	5.768	-60.735	8.485	15.817	2.285	2.937
4	15/03/2016	23:00:00	-17.992	-0.599	-57.521	7.483	15.525	3.252	4.071
5	15/03/2016	23:30:00	-7.068	1.906	-56.955	7.680	15.237	1.868	1.867
6	16/03/2016	00:00:00	-5.176	-2.999	-56.982	7.525	14.949	3.023	1.296
7	16/03/2016	00:30:00	-12.898	-3.696	-50.455	7.318	14.700	3.330	4.597
8	16/03/2016	01:00:00	-8.228	-3.925	-50.035	6.910	14.499	3.196	4.592
9	16/03/2016	01:30:00	-12.211	-2.305	-44.888	6.980	14.270	2.224	5.786

H: Calor sensible, LE: Calor latente, RN: Radiación neta, TA: Temperatura del aire, TS: Temperatura del suelo, Tdew: Temperatura del punto de rocío y CO<sub>2</sub> flux: flujo de CO<sub>2</sub>. Fuente de datos: Proyecto P09-CeMIE-Geo.

4. Seleccionar las variables más adecuadas para modelar el flujo de CO<sub>2</sub> utilizando RNA. Este procedimiento se realizó con 3 vertientes: (i) seleccionar las variables que se reportan en la literatura como las más adecuadas para modelar el flujo de CO<sub>2</sub>; (ii) utilizar las variables de acuerdo con una matriz de correlación con respecto a la variable a modelar; y, (iii) utilizar la técnica *Stepwise Selection (forward selection)*, ver Fig. 4.5.

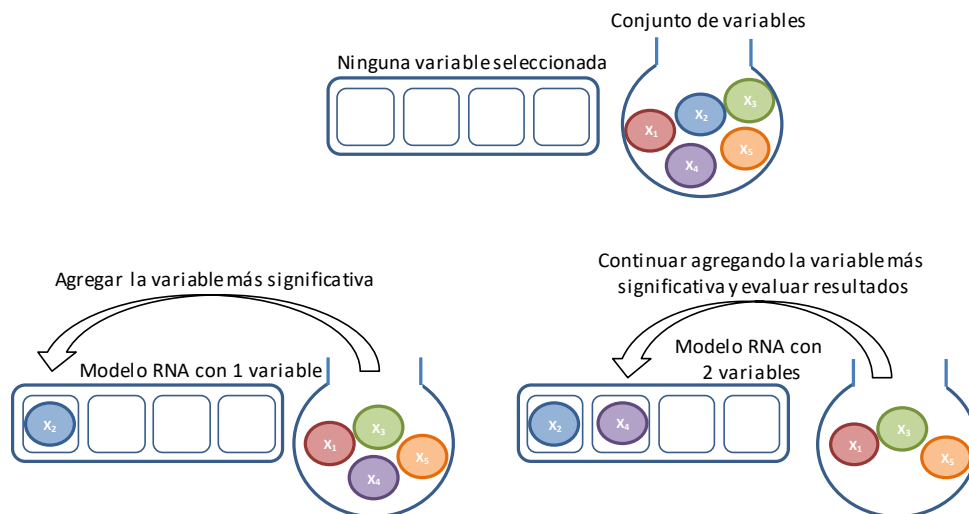


Fig. 4.5 Representación esquemática de *forward selection*.

5. Se entrenaron distintas arquitecturas de RNA utilizando el algoritmo propuesto en el ANEXO 1 (scripts de Matlab) disponible en <https://github.com/ANNGroup/GasG-Scripts-ANNs.git>. El modelo neuronal

fue entrenado con las siguientes consideraciones: (i) se diseñaron alrededor de 50 combinaciones diferentes de las variables de entrada para modelar el flujo de CO<sub>2</sub>, las arquitecturas más sobresalientes aparecen en la Tabla 1, ANEXO 3; (ii) la normalización de los datos de entrada y del target se hizo entre -1 y 1 de acuerdo a la función de activación tangente hiperbólica; (iii) los datos se dividieron en tres conjuntos 50% para entrenamiento, 25% para validación (evitar el sobreajuste de la RNA) y 25% para prueba; (iv) se utilizó un modelo de RNA con 3 capas: una capa de entrada, una oculta y una de salida. La capa de entrada fue variando el número de neuronas conforme al procedimiento *forward selection* (ver Fig. 4.5), las neuronas de la capa oculta fueron variando de 1 - 30 y en la capa de salida se utilizó una neurona para estimar el flujo de CO<sub>2</sub>; (v) similar al procedimiento del Módulo 1, se reguló la tasa de aprendizaje o con 4 valores diferentes: 0.01, 0.001, 0.0001 y 0.00001 y utilizando el método Levenberg-Marquardt (LM); (vi) para el desempeño de cada arquitectura entrenada, se calcularon diferentes métricas estadísticas tales como: el coeficiente de correlación  $r$ , RMSE, MAE (ver Tabla 2, ANEXO 3); y, (viii) la contribución relativa de cada variable de entrada fue calculada con base al algoritmo de Garson (1991).

6. La evaluación de las arquitecturas obtenidas durante el entrenamiento se realizó con base al método *MCD*A descrito brevemente en el punto 5 del Módulo 1 y detallado en el ANEXO 2.
7. Con base a la arquitectura seleccionada, se calculó el flujo de CO<sub>2</sub> utilizando las variables ambientales y flujos de energía en los periodos de ausencia de medición del flujo de CO<sub>2</sub>, es decir, rellenar los huecos evaluando el perfil obtenido por el modelo neuronal.

Los resultados obtenidos durante el desarrollo metodológico del Módulo 2, se presentarán en el Capítulo 5.

### **4.3 Metodología del Módulo 3**

A diferencia de los Módulos 1 y 2 donde los antecedentes, desarrollo metodológico y los resultados se encuentran detallados en ANEXOS (artículos publicados en

revistas internacionales bajo arbitraje de expertos), el desarrollo de nuevas herramientas para la discriminación de datos de transporte de gas en sistemas geotérmicos se describe en esta sección.

**Antecedentes.** La composición de los fluidos geotérmicos en su fase vapor está dada por H<sub>2</sub>O, gases mayores (CO<sub>2</sub>, H<sub>2</sub>S y NH<sub>3</sub>) y gases traza (CH<sub>4</sub>, H<sub>2</sub>, N<sub>2</sub>, CO y O<sub>2</sub>). El componente más abundante después del agua es el dióxido de carbono ~70-80% y el restante ~30-20% corresponde a los gases traza (Tabaco, et. al., 1991; Portugal et al., 1994, Viggiano-Guerra et al., 2011). En los últimos años, la comunidad geocientífica ha centrado sus estudios en los gases provenientes del suelo. Los objetivos de éstos son la cuantificación del gas y el flujo del calor proveniente de volcanes y sistemas geotérmicos (Fridriksson et al., 2016).

La medición de las emisiones de CO<sub>2</sub> que emergen a la superficie a través de fisuras y/o fracturas en las placas tectónicas, es una práctica muy común y de gran importancia en zonas volcánicas y geotérmicas (Chiodini et al., 1998). Dado que el CO<sub>2</sub> es uno de los primeros componentes volátiles, que son liberados por el magma durante la migración hacia la superficie, el conocimiento del flujo de CO<sub>2</sub> permite identificar geoquímicamente la existencia de una actividad magmática a profundidad (Giammanco et al., 2007) o si el valor observado corresponde a otro tipo de procesos (p. ej. biogénico y/o atmosférico). El flujo de CO<sub>2</sub> es variable y depende del ecosistema que se estudie, así como la época del año. Las mediciones de CO<sub>2</sub> en la superficie consideradas biogénicas, dependen de factores como la porosidad del suelo, la humedad, la temperatura, la materia orgánica, entre otras (Welles et al., 2001). Particularmente, en áreas volcánicas y geotérmicas la existencia de una fuente biogénica de CO<sub>2</sub> es el resultado de un fondo regional y un grupo de muestras consideradas como “anómalas” proveniente de la profundidad (Chiodini et al., 2008; Peiffer et al., 2014). Generalmente, los valores de flujo de CO<sub>2</sub> que rondan en las decenas (g m<sup>-2</sup> d<sup>-1</sup>), están asociados a una fuente biogénica y los flujos de hasta 10 veces más altos se consideran representativos de una fuente hidrotermal volcánica (Chiodini et al., 2008).

Un método para identificar los límites entre datos geoquímicos de fondo y datos "anómalos" basado en la partición de la distribución acumulada en un gráfico es el propuesto por Sinclair (1974). En la industria de la exploración mineral, el término de umbral o límite es utilizado para darle un significado a la identificación de un valor específico que permita separar los valores altos y bajos, conforme a la presencia de características diferentes en las mediciones. Es por ello que elegir un valor límite tiene un impacto directo al momento de dirigir la exploración de un lugar con concentraciones altas, donde las probabilidades de descubrir un depósito natural es económicamente más factible (Sinclair, 1974). Particularmente en la presente investigación, el enfoque propuesto para el método de Sinclair es la identificación de los posibles/probables orígenes del flujo de CO<sub>2</sub> medido a nivel superficial como identificador de anomalías térmicas en estudios de exploración y prospección geotérmica.

**Método de Sinclair.** Para discriminar el origen del flujo de CO<sub>2</sub>, Sinclair (1974) propone un método para elegir valores de umbral entre datos anómalos y de fondo geoquímico, utilizando la partición de la frecuencia acumulada en un gráfico. Este método no está limitado a un mineral específico dado que, en su manuscrito Sinclair describe varios ejemplos con Cu, Zn y Ni.

Es un método muy utilizado en estudios geoquímicos. Así mismo, existen algunos trabajos donde ha sido empleado satisfactoriamente para discriminar el flujo CO<sub>2</sub> de suelos (p.ej. Parks et al., 2013; Peiffer et al., 2014). El método se basa en graficar la frecuencia acumulada [%] del flujo de CO<sub>2</sub> en las abscisas y una escala aritmética en las ordenadas. Los valores de las abscisas se trazan partiendo de la parte superior o valor más alto (Lepeltier, 1969). El cambio de concavidad en la curva denota un punto de inflexión. Por ejemplo, en la Fig. 4.6 aparece un punto de inflexión en el percentil 20 de la frecuencia acumulada, lo que representa el 20% de un grupo A y el restante 80% pertenece a un grupo B. La forma de la curva es característica de dos grupos superpuestos. En la Fig. 4.6 aparecen dos grupos de muestras ideales A y B que se combinan en las proporciones A/B = 20/80. La curva es dibujada a través de puntos sólidos. Un punto de inflexión se muestra por la punta

de flecha. Los umbrales arbitrarios al nivel del 1% del grupo B y el 99% del grupo A corresponden a 78 y 44 ppm, respectivamente.

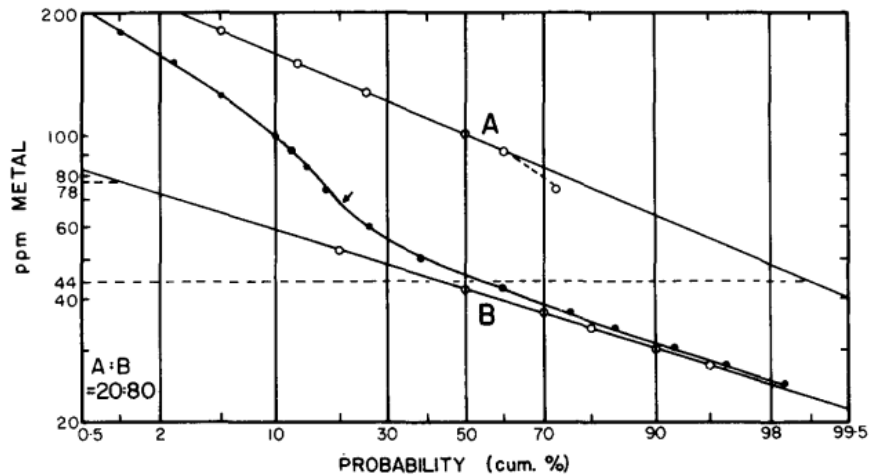


Fig. 4.6 Ejemplo hipotético del manuscrito de Sinclair, 1974.

El primer punto sólido trazado en la parte superior de la curva al nivel de ordenadas de 180 ppm representa el 1% del total de datos. Sin embargo, también representa  $(1/20 \times 100) = 5$  por ciento del grupo de muestras A.

Por lo tanto, un punto en el grupo de A se define en 5 por ciento de la frecuencia acumulada en el nivel de 180 ppm. Este procedimiento se repite análogamente hasta que se obtienen suficientes puntos para definir el grupo A por una línea recta. Si los puntos comienzan a apartarse de un patrón lineal, significa que el grupo B está presente en cantidades significativas. Los puntos del grupo B se obtienen como valores complementarios, p. ej. 90 por ciento acumulativo se lee como  $(100 - 90) = 10$  por ciento. Con este método y la selección arbitraria de dos umbrales (1% y 99%), la interpretación general de la Fig. 4.6 indica la existencia definida de tres grupos de datos: uno superior (A) por encima de 78 ppm, uno inferior (B) por debajo de 44 ppm y una mezcla de ambos grupos (entre 44-78 ppm). Suponiendo que los grupos A y B corresponden a datos anómalos y de fondo geoquímicos respectivamente, el método de Sinclair ha dividido de manera muy eficaz los datos. A pesar de ser una alternativa para discriminar el origen de ciertos gases (p. ej. flujo de  $\text{CO}_2$ ), el método de Sinclair presenta un obstáculo que el mismo autor menciona en el trabajo original, es decir, la selección arbitraria del punto de inflexión para

determinar los grupos de muestras presentes. Esta es la razón por la cual es imprescindible proponer una alternativa a la selección del punto de inflexión que tenga un sustento matemático. El esquema general del método de Sinclair se muestra en la Fig. 4.7, en la cual se observan siete procesos secuenciales, resaltando que la mayor aportación en la presente investigación está centrada en el proceso número (5), es decir, la selección no arbitraria del punto de inflexión en la curva de la frecuencia acumulada.

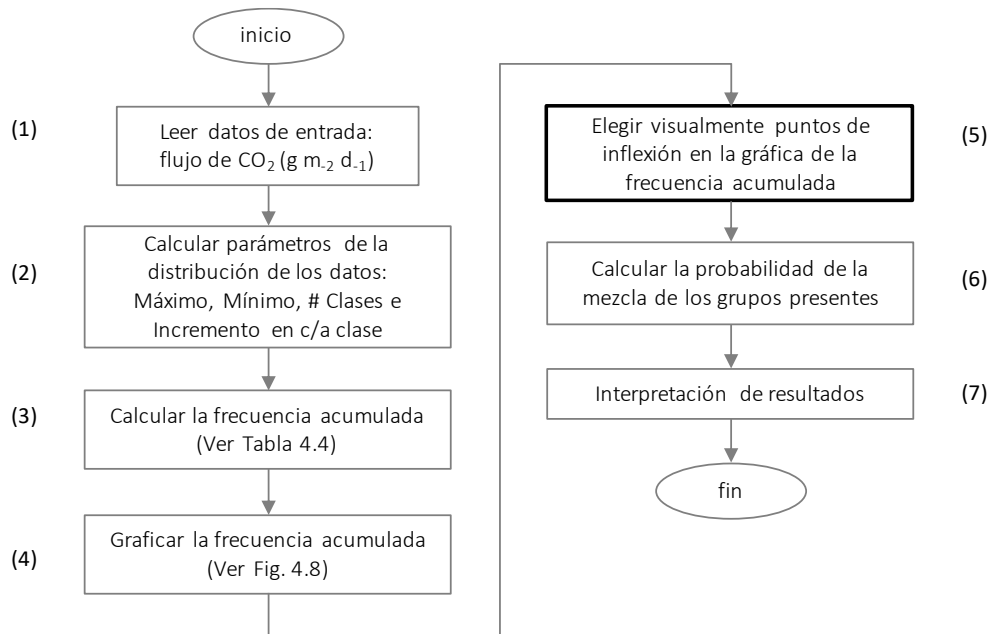


Fig. 4.7 Diagrama general del método de Sinclair, 1974.

**Desarrollo.** El objetivo principal de este estudio es el desarrollo de una herramienta geotermométrica computacional para la identificación del origen del flujo de CO<sub>2</sub> en estudios de exploración y prospección geotérmica. El método geotermométrico utilizado es el método de Sinclair y la zona de estudio para identificar/discriminar el flujo de CO<sub>2</sub> medido a nivel de la superficie terrestre es la caldera de Acoculco, ubicada en el estado de Puebla. Dado que es un método estadístico y económicamente más viable en comparación con métodos analíticos (p. ej. 3He/4He), a continuación, se describe el procedimiento del método:

- (1) En colaboración con el grupo de investigación en geotermia (Proyecto P09-CeMIE-Geo) se compilaron los datos del flujo de CO<sub>2</sub> fueron obtenidos de la zona de estudio conocida como Los Azufres por exhibir pocas anomalías

geotérmicas en la superficie como son: mofetas de CO<sub>2</sub> y evidencia de alteración hidrotermal en las rocas superficiales. Para la creación de la Base de Datos (BDCA), las mediciones del flujo de CO<sub>2</sub> fueron realizadas mediante el método de Cámara de Acumulación (CA), ver Fig. 4.8. Este método permite medir el flujo de CO<sub>2</sub> en suelos por unidad de área y unidad de tiempo. Consiste en determinar la tasa de aumento de concentración de CO<sub>2</sub> dentro de la cámara, a lo largo del tiempo mientras esté situada en el suelo.

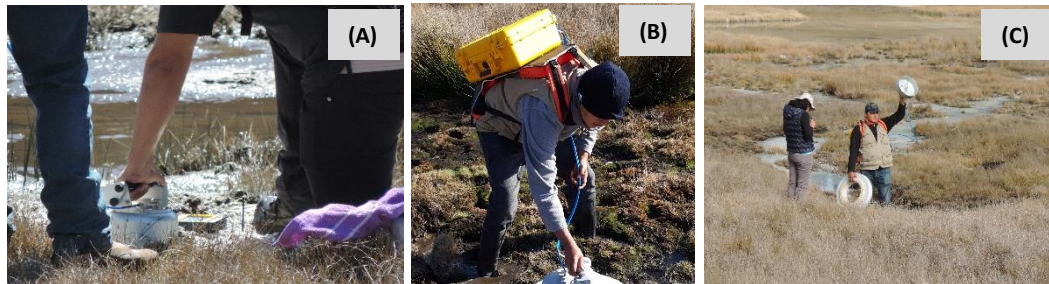


Fig. 4.8 Instrumentos de medición de la Cámara de Acumulación.

En la Fig. 4.8, La sección (A) es la CA portátil en contacto directo con el suelo. Después de ~10 minutos, en la sección (B) se toma la lectura de flujo de CO<sub>2</sub> en una palm conectada al equipo, así como las coordenadas del punto de medición a través de un GPS integrado. En la sección (C), se recomienda que posterior a cada medición, la campana de metal se levanta y se deja que el aire apoye a limpiar la huella de la medición anterior. En la Tabla 4.3 se muestra un extracto de 10 registros de la BDCA.

Tabla 4.3 Extracto de la BDCA con mediciones del flujo de CO<sub>2</sub> en la zona de Los Azufres.

n	Zona	Longitud	Latitud	Elevación	Flujo de CO <sub>2</sub>
1	AZ	-98.144129	19.922493	2842.8	5.32
2	AZ	-98.14411	19.923179	2848.28	13.4
3	AZ	-98.14421	19.922504	2845.8	7.83
4	AZ	-98.144249	19.922519	2847.31	801
5	AZ	-98.144282	19.922568	2845.26	0.559
6	AZ	-98.144158	19.923228	2848.92	7.13
7	AZ	-98.144221	19.922571	2845.46	3.43
8	AZ	-98.144127	19.922587	2843.44	246
9	AZ	-98.144044	19.923016	2851.26	46.2
10	AZ	-98.144129	19.922493	2842.8	5.32

Las unidades de concentración del flujo de CO<sub>2</sub> son: g m<sup>-2</sup> dia<sup>-1</sup>. Fuente de datos: Proyecto P09-CeMIE-Geo.

- (2) La BDCA creada a partir de las mediciones del flujo de CO<sub>2</sub> contiene 279 registros, con valores máximos y mínimos de 79200 y 0.559 (g m<sup>-2</sup> día<sup>-1</sup>), respectivamente. El # de clases en la frecuencia acumulada fue determinado por la raíz cuadrada de 279, es decir, 17 clases y el incremento para cada clase fue el resultado de la diferencia entre el valor máximo y valor mínimo dividido entre el # clases.
- (3) Los datos de la frecuencia acumulada se muestran en la Tabla 4.4. Como se puede observar, la segunda y tercera columna muestran los valores iniciales y finales para cada clase. Y en la cuarta columna se contabilizan el # de muestras encontradas en ese rango específico.

Tabla 4.4 Frecuencia acumulada de los datos medidos en Aocolco, Puebla (zona de Los Azufres)

# Clase	Valor inicial (g m <sup>-2</sup> d <sup>-1</sup> )	Valor final (g m <sup>-2</sup> d <sup>-1</sup> )	# Muestras en la clase
1	-0.26	0.02	1
2	0.02	0.3	1
3	0.3	0.58	9
4	0.58	0.86	21
5	0.86	1.14	37
6	1.14	1.42	62
7	1.42	1.7	79
8	1.7	1.98	8
9	1.98	2.26	4
10	2.26	2.54	4
11	2.54	2.82	3
12	2.82	3.1	4
13	3.1	3.38	5
14	3.38	3.66	7
15	3.66	3.94	12
16	3.94	4.22	10
17	4.22	5	12
<b>Total</b>			<b>279</b>

Los valores originales del flujo de CO<sub>2</sub> fueron transformados a escala logarítmica.

- (4) La gráfica de la frecuencia acumulada se muestran en la Fig. 4.9 de acuerdo con el método de Sinclair. Es notable que visualmente se puede apreciar más de un punto de inflexión en la curva de la frecuencia acumulada.



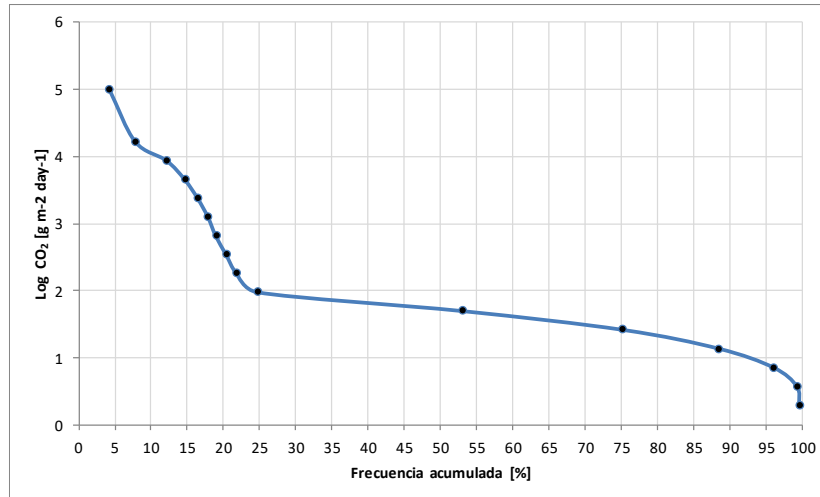


Fig. 4.9 Frecuencia acumulada en escala logarítmica del flujo de CO<sub>2</sub> medido en Los Azufres.

- (5) Para determinar el punto de inflexión no visualmente (mejora al método de Sinclair), se propuso utilizar el principio de la prueba de rachas el cual indica un cambio en el signo de la pendiente de una función (Espinoza-Ojeda et al., 2011), por ej. +++ --- +++, ver Fig. 4.10.

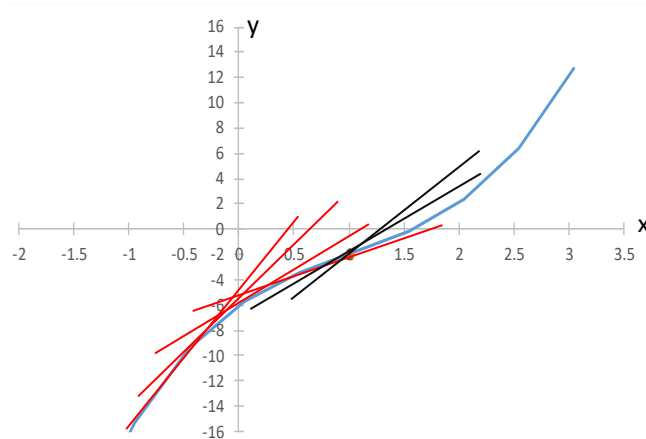


Fig. 4.10 Representación del cambio de pendiente.

En el ejemplo de la Fig. 4.10 aparece graficada la función  $f(x) = x^3 - 3x^2 + 6x - 6$  en color azul. Las líneas rojas indican el decremento del valor de la pendiente calculada con subconjuntos de 4 datos; y las líneas negras muestran un incremento en el valor de la pendiente. Considerando la naturaleza del método de Sinclair, donde no se cuenta con una función  $f(x)$  que represente la curvatura de la frecuencia acumulada, se utilizó la

interpolación de *Lagrange* para las coordenadas desconocidas de la curva y con ello poder calcular el valor de la pendiente en subconjuntos de 4 datos. El polinomio de *Lagrange* interpola valores dado un conjunto de puntos conocidos en la forma:

$$P(x) = y_1L_1(x) + y_2L_2(x) + \dots + y_nL_n(x) \quad \text{Ec. (4.1)}$$

donde:

$$L_1 = \frac{(x - x_2)(x - x_3) \dots (x - x_n)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_n)} ; \quad L_1 = \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_2 - x_1)(x_2 - x_3) \dots (x_n - x_n)} ;$$

$$L_1 = \frac{(x - x_1)(x - x_2) \dots (x - x_{n-1})}{(x_n - x_1)(x_n - x_2) \dots (x_n - x_{n-1})}$$

**Identificación del punto de inflexión.** El diagrama general para encontrar el punto de inflexión a través del cambio de pendiente e interpolando con *Lagrange* las coordenadas desconocidas de la frecuencia acumulada se muestra en la Fig. 4.11.

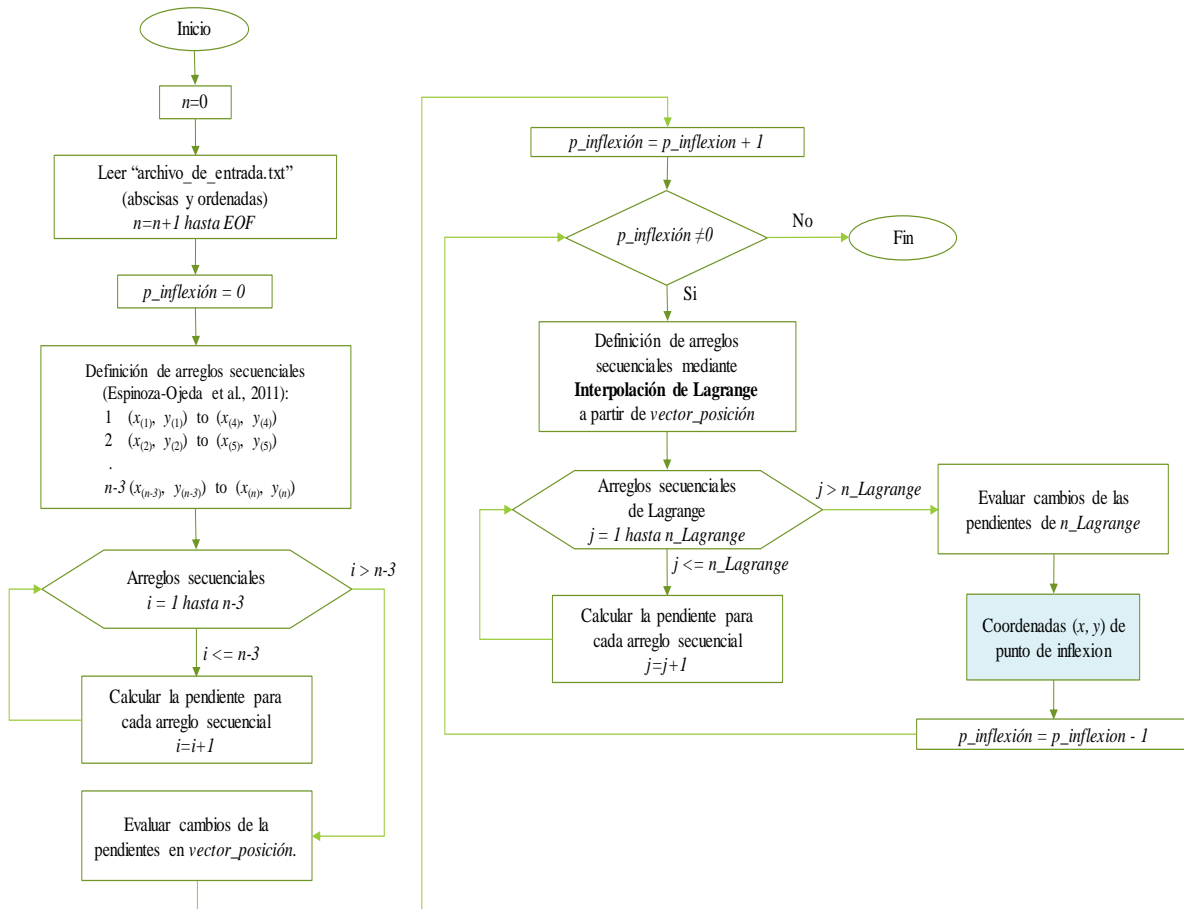


Fig. 4.11 Diagrama general del procedimiento para calcular el punto de inflexión.

- (6) Con base al procedimiento anterior para la determinación del punto de inflexión en la gráfica de la frecuencia acumulada, se encontraron tres cambios en la concavidad de la curva en las siguientes ubicaciones: 12.2%, 17.9% y 53% como se muestra en la Fig. 4.12.

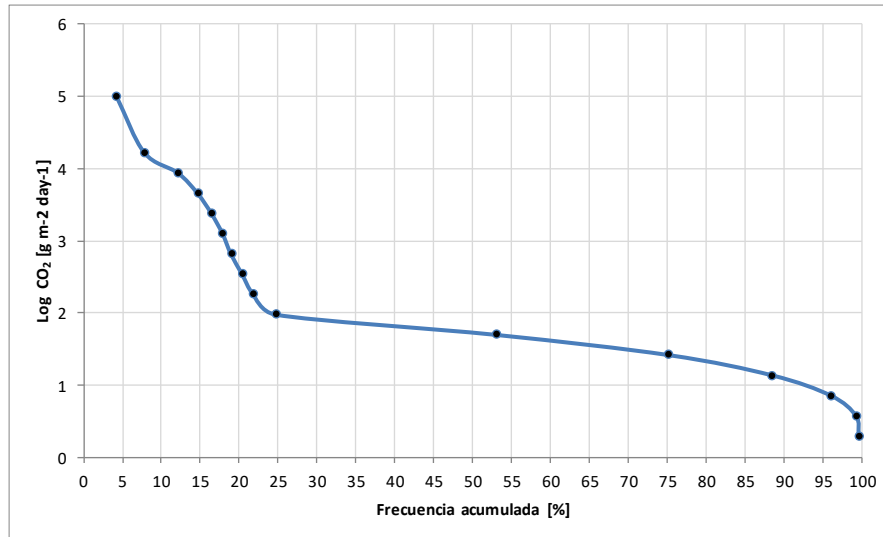


Fig. 4.12 Determinación de tres puntos de inflexión con base al cambio del valor de la pendiente.

De acuerdo con el método de Sinclair, la determinación de tres puntos de inflexión en la curva de la frecuencia acumulada indica la presencia de cuatro grupos de muestras del flujo de CO<sub>2</sub>, es decir, cuatro posibles orígenes del mismo: Grupo A, B, C y D.

En la Fig. 4.12 el primer punto de inflexión divide la curva de la frecuencia acumulada asignando el 12.2% de las concentraciones del flujo de CO<sub>2</sub> a un conjunto de muestras que llamaremos Grupo A. La diferencia entre 17.9% y 12.2% nos da proporción de 5.7% para un Grupo B. De forma similar, la diferencia entre el tercer y segundo punto de inflexión (53% y 17.9%) indica la proporción de un Grupo C igual a 35.1%. Con las proporciones asignadas de los Grupos A, B, y C, el último Grupo D tiene el 47% de los datos del flujo de CO<sub>2</sub>. En la Fig. 4.13 se muestran los cuatro grupos de muestras identificados con base al método de Sinclair, que dado el objetivo de la presente investigación reflejan cuatro posibles orígenes del flujo de CO<sub>2</sub> medido en la zona geotérmica de Los Azufres (Aocolco, Puebla).

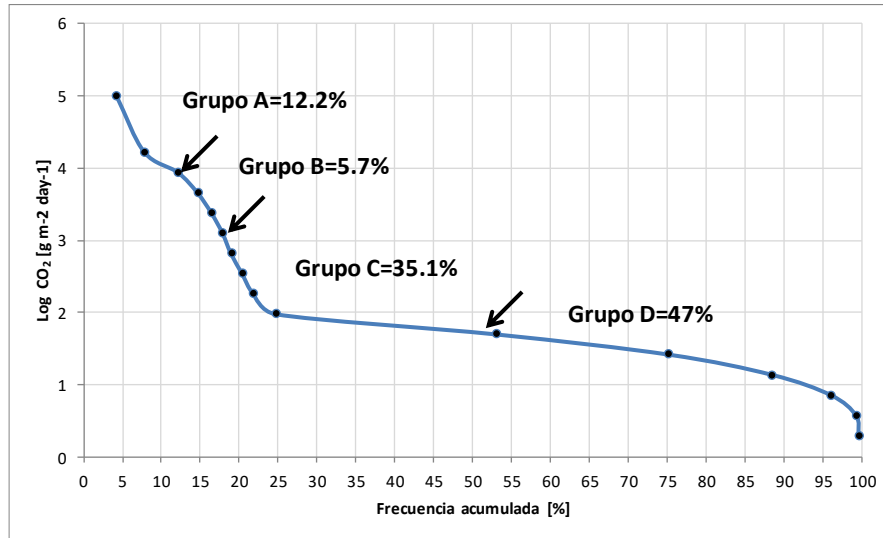


Fig. 4.13 Cuatro grupos identificados en las mediciones del flujo de CO<sub>2</sub> muestreado en Los Azufres.

Continuando con el procedimiento de Sinclair, cada grupo identificado se debe representar por una línea recta para determinar la probabilidad de la mezcla entre los grupos, es decir, delimitar los umbrales límite entre cada grupo.

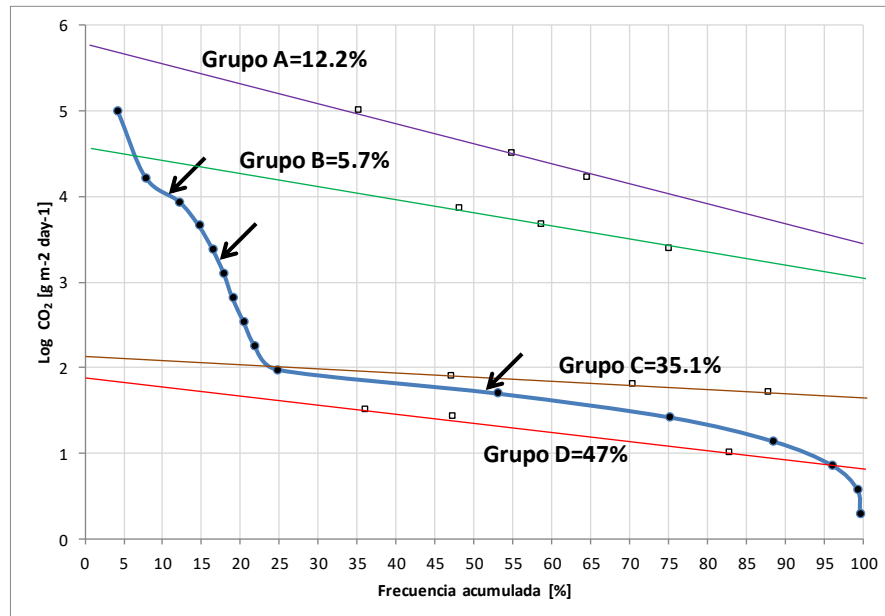


Fig. 4.14 Representación de los Grupos A, B, C, y D por medio de una línea recta.

En la Fig. 4.14 se calcularon los recuadros en blanco para representar cada uno de los grupos identificados en las mediciones del flujo de CO<sub>2</sub> a través de una línea recta. Por ejemplo, el primer punto en color negro que aparece en la

curva de la frecuencia acumulada graficado al nivel de las ordenadas de 5 ( $\text{Log CO}_2 \text{ g m}^{-2} \text{ d}^{-1}$ ) representa el 4.3% del total de los datos. Así mismo, también representa  $(4.3/12.2 \times 100) = 35.2\%$  de las muestras de  $\text{CO}_2$  pertenecientes al Grupo A. De acuerdo con el método de Sinclair, este procedimiento se repite hasta obtener un mínimo de dos puntos y poder definir el Grupo A por una línea recta (línea de color morado). El resto de los Grupos de muestras B, C, y D están representados por líneas de color verde, café y roja, respectivamente.

- (7) Los resultados finales e interpretación de estos conforme a la discriminación del flujo de  $\text{CO}_2$  utilizando el método de Sinclair, se presentan en el siguiente Capítulo.

### 4.3 Pseudocódigo de los Módulos 1, 2 y 3

Para fortalecer la descripción de las metodologías desarrolladas en los Módulos 1, 2 y 3, a continuación, se presentan los pseudocódigos correspondientes a cada una de ellas.

Pseudocódigo: Metodología del Módulo 1

**Inicio**

**Compilación** en literatura de datos geoquímicos (gases y temperatura) almacenados en BDGM=591

**Pruebas** estadísticas univariadas y multivariadas a BDGM

**Creación** de 3 sub-bases de datos ( $\text{SubDB}_1=527$ ,  $\text{SubBD}_2=498$  y  $\text{SubBD}_3=97$ )

**Diseño** de 13 combinaciones diferentes usando los gases geotérmicos

**Repetir**

**Subrutina A)**

Para cada sub-base de datos ( $\text{SubDB}_1=527$ ,  $\text{SubBD}_2=498$  y  $\text{SubBD}_3=97$ ) **HACER**

A1. Script para el entrenamiento de la RNA

A2. Entrenamiento de la RNA usando las 13 combinaciones

A3. Almacenar arquitecturas derivadas del entrenamiento de la RNA

**Fin Subrutina A)**

**Selección** de arquitecturas de las RNA con base al método MCDA

**Nuevos** geotermómetros de gas ( $\text{GasG}_1 - \text{GasG}_8$ ) con base a las arquitecturas seleccionadas

**Compilación** de 25 geotermómetros de gas reportados en la literatura

**Creación** de base de datos para validación (NGWDB=265)

**Separación** de muestras geoquímicas provenientes de NWGDB en yacimientos de líquido—dominante (LIQDR=178) y vapor—dominante (VAPDR=87)

**Conversión** de unidades de concentración de NWGDB para los geotermómetros de gas

**Desarrollo** de nuevo programa para aplicación de los 8 nuevos geotermómetros: Gas\_GeoT

**Análisis** geoquimiométrico para comparar resultados de predicción de temperaturas usando los geotermómetros de gas (8 nuevos y 25 existentes) a través de parámetros estadísticos: DIFF%, RMSE, MAE, MAPE y Theil's U

**Resultados** finales: para yacimientos LIQDR los 8 geotermómetros (GasG<sub>1</sub> – GasG<sub>8</sub>) presentan el mejor desempeño; y para yacimientos de VAPDR son los geotermómetros GasG<sub>1</sub> y Gas<sub>3</sub>; el geotermómetro más confiable en ambos tipos de yacimientos es el GasG<sub>3</sub>

### Pseudocódigo: Metodología del Módulo 2

**Inicio**

**Selección** de zona geotérmica en la Caldera de Acoculco, Puebla

**Instalación** de equipo micrometeorológico Eddy-Covariance (proyecto P09-CeMIE-Geo)

**Creación** de base de datos BDEddy=1685 con mediciones de equipo Eddy-Covariance

**Repetir**

**Subrutina A)**

Para cada nueva selección de variables (stepwise selection-forward selection) **HACER**

A1. Script para el entrenamiento de la RNA

A2. Entrenamiento de la RNA usando las variables seleccionadas

A3. Almacenar arquitecturas derivadas del entrenamiento de la RNA

**Fin** Subrutina A)

**Selección** de la arquitectura ANN-7 con base al método MCDA

**Validación** estadística de predicciones de la ANN-7 usando las pruebas F y t-studen

**Identificación** de gaps para ser rellenados con la predicción de flujo de CO<sub>2</sub> de la ANN-7

**Predicción** de datos ausentes utilizando la ANN-7

**Resultados** finales: la ANN-7 es capaz de reproducir los perfiles del flujo de CO<sub>2</sub> en los periodos de ausencia de datos del equipo Eddy-Covariance a causa de la aparición de neblina y bajas temperaturas

### Pseudocódigo: Metodología del Módulo 3

**Inicio**

**Selección** de zona geotérmica en la Caldera de Acoculco, Puebla

**Medición** de flujo de CO<sub>2</sub> en la superficie utilizando el método de Cámara de Acumulación (CA)

colaboración con grupo de investigación del proyecto P09-CeMIE-Geo

**Creación** de base de datos BDCA=279 con mediciones de CA

**Desarrollo** del método de Sinclair mejorado para identificar el origen del flujo de CO<sub>2</sub> de BDCA

**Subproceso A)**

A1. Calcular parámetro de la distribución de los datos (máximo, mínimo, # clases e incremento en cada clase)

A2. Calcular la frecuencia acumulada y graficarla

A3. Identificación del punto de inflexión en la frecuencia acumulada utilizando el principio de la prueba de rachas (cambio en el signo de la pendiente de una función  $f(x)$  y la interpolación de Lagrange

**Subproceso B)**

B1. Definición de arreglos secuenciales en subconjuntos de 4 datos  $n = 1 \dots, n - 3$

B2. Repetir

**Subrutina A)**

Por cada arreglo secuencial  $i = 1$  mientras  $i \leq n - 3$  HACER

Calcular la pendiente para cada arreglo secuencial  $i += 1$

Fin Subrutina A)

B3. Evaluar cambios de las pendientes en todos los arreglos secuenciales

B4. Repetir

**Subrutina B)**

Por cada cambio en el valor de la pendiente  $p_{inflexión} \neq 0$  HACER

Crear arreglos secuenciales usando interpolación de Lagrange ( $n_{Lagrange} += 1$ )

Repetir

**Subrutina C)**

Por cada arreglo secuencial  $j = 1$  mientras  $j < n_{Lagrange}$  HACER

Calcular el valor de la pendiente para cada arreglo secuencial  $j += 1$

Fin Subrutina C)

Evaluar cambios de las pendientes de arreglos  $n_{Lagrange}$

Obtención de coordenadas  $(x, y)$  de punto de inflexión  $p_{inflexión} += 1$

Fin Subrutina B)

**Graficar** los 3 puntos de inflexión determinados por el algoritmo descrito en sección previa (3.1)

**Con base** a Sinclair, 3 puntos de inflexión representan de 4 orígenes del flujo de CO<sub>2</sub>

**Interpretación** de resultados: de las 279 mediciones del flujo de CO<sub>2</sub> el 12.2% son Magmáticas, 5.7% Atmosféricas, 35.1% Antropogénicas y 47% Biogénicas.

## CAPÍTULO 5

### Resultados del Proyecto de Investigación

**Resumen del Capítulo.** En este Capítulo se presentan los resultados derivados de los tres módulos planteados como objetivos específicos. Durante el desarrollo del Módulo 1, se obtuvieron 8 nuevos geotermómetros de gases para ser utilizados en la etapa de exploración y/o explotación de sistemas geotérmicos a través de un nuevo programa llamado GaS\_GeoT (ver ANEXO 1 y ANEXO 2). Para el Modulo 2, se obtuvieron diversos modelos para predecir el flujo de CO<sub>2</sub>, entre los cuales fue seleccionada la RNA más eficiente a través de un análisis multi-criterio sustentado en la metodología de *MCDA*. En el desarrollo del Módulo 3, se discriminó exitosamente el origen del flujo de CO<sub>2</sub> a través de una nueva metodología geoquimiométrica computacional utilizando de base el método de Sinclair.

#### 5.1 Resultados del Módulo 1

En esta sección se presentan los resultados del desarrollo de nuevos geotermómetros de gases para la predicción de temperaturas de fondo.

En las Tablas 10-12 del ANEXO 1 se presentan los principales resultados del entrenamiento realizado con cada una de las tres sub-bases de datos (ANN-1 a ANN-39: **Artificial Neural Network**). Para la primera sub-base de datos WG\_SubDB<sub>1</sub> (q<sub>1</sub>=527 muestras) el resultado de las 13 combinaciones de los gases aparece en la Tabla 10 del mismo ANEXO.

Los coeficientes de correlación más altos que fueron calculados entre los valores medidos de *BHT* y la temperatura estimada corresponden a la arquitectura ANN-12 y ANN-13 con las siguientes características: ANN-12: 4 variables de entrada  $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$  y  $\ln(\text{H}_2)$ ; 13 neuronas en la capa oculta,  $r_{\text{global}}=0.8379$ ,  $r_{\text{entrenamiento}}=0.8568$ ,  $r_{\text{validación}}=0.7824$  y  $r_{\text{test}}=0.7499$ . ANN-13: 5 variables de entrada  $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ ,  $\ln(\text{H}_2/\text{CO}_2)$  y  $\ln(\text{H}_2)$ ; 9 neuronas en la capa oculta,  $r_{\text{global}}=0.8273$ ,  $r_{\text{entrenamiento}}=0.8457$ ,  $r_{\text{validación}}=0.7937$  y  $r_{\text{test}}=0.7165$ . Gráficamente se muestran en la Fig. 7 del ANEXO 1.



Para la segunda sub-base de datos WG\_SubDB<sub>2</sub> (q<sub>2</sub>=498 muestras) los resultados aparecen en la Tabla 11 del ANEXO 1. Las dos mejores arquitecturas son ANN-22 y ANN-25. La ANN-22 tiene 3 variables de entrada  $\ln(\text{CO}_2/\text{CH}_4)$ ,  $\ln(\text{H}_2\text{S}/\text{CH}_4)$ ,  $\ln(\text{H}_2/\text{CH}_4)$ ; 34 neuronas en la capa oculta,  $r$  global=0.8546,  $r$  entrenamiento=0.8838,  $r$  validación=0.7992 y  $r$  test=0.7872. La ANN-25: 4 variables de entrada  $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$  y  $\ln(\text{H}_2)$ ; 10 neuronas en la capa oculta,  $r$  global=0.8505,  $r$  entrenamiento=0.8635,  $r$  validación=0.8024 y  $r$  test=0.7906.

Para la última sub-base de datos WG\_SubDB<sub>3</sub> (q<sub>3</sub>=97 muestras) los resultados aparecen en la Tabla 12 del ANEXO 1. Las dos mejores arquitecturas son ANN-33 y ANN-38. La ANN-33 tiene 3 variables de entrada  $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ ,  $\ln(\text{H}_2/\text{CO}_2)$ ; 8 neuronas en la capa oculta,  $r$  global=0.9403,  $r$  entrenamiento=0.9530,  $r$  validación=0.9242 y  $r$  test=0.9109. La ANN-38: 4 variables de entrada  $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$  y  $\ln(\text{H}_2)$ ; 8 neuronas en la capa oculta,  $r$  global=0.9211,  $r$  entrenamiento=0.9422,  $r$  validación=0.9207 y  $r$  test=0.9197.

Los coeficientes de correlación más altos y los valores más pequeños del error medido por el parámetro RMSE y MAE, fueron obtenidos con los resultados del entrenamiento con la sub-base de datos WG\_SubDB<sub>3</sub>. La razón fundamental es que esta sub-base de datos fue altamente filtrada con los criterios de control de calidad descritos en el Capítulo 4. Si bien es cierto que durante el entrenamiento los resultados son más que prometedores, al ser una sub-base de datos tan pequeña difícilmente representará la complejidad de la totalidad de las muestras gaseosas de zonas geotérmicas.

Para evaluar más objetivamente y de forma integral las 39 ANNs, en la presente investigación se utilizó por primera vez en la literatura el método *MCDA* para seleccionar los mejores modelos predictivos con base a métricas estadísticas divididas en tres secciones: (i) los parámetros obtenidos durante la etapa de entrenamiento; (ii) las métricas RMSE, MAE y MAPE calculados con todos los datos de la sub-base de datos más representativa (WG\_SubDB<sub>1</sub>); y, (iii) las mismas

métricas pero calculadas solo con las muestras que cumplían las condiciones de aplicabilidad con base a la etapa de entrenamiento (ver Tabla 4, ANEXO 2).

Con los datos calculados, se crearon 4 escenarios con ponderaciones diferentes para cada sección de condiciones de aplicabilidad (ver Fig. 1B, ANEXO 2). Los detalles del procedimiento *MCDA* se muestran en las Tablas S1-S4 del material suplementario y en la Tabla 5 del mismo ANEXO. Los modelos seleccionados a través el método *MCDA* son: ANN-22, ANN-12, ANN-25, ANN-13, ANN-23, ANN-7, ANN-9 y ANN-20. Conforme al análisis de sensibilidad calculado por el algoritmo de Garson (1991), la mayoría de las variables que utilizan el CO<sub>2</sub> contribuyen con más aportación porcentual para estimar la *temperatura*. Dichos porcentajes van desde un 24.6% de la ANN-23 hasta el 46% de la ANN-7, con la variable ln(CO<sub>2</sub>/H<sub>2</sub>) y ln(H<sub>2</sub>S/CO<sub>2</sub>), respectivamente.

**Ecuación general geotermométrica.** Todos los coeficientes obtenidos por las 8 ANNs seleccionadas se muestran en la Tabla 7 y las condiciones de aplicabilidad en la Tabla 8 del ANEXO 2. La ecuación general para todas las ANNs es la siguiente:

$$BHT_{GasG_i} (^{\circ}C) = \left\{ \left[ \frac{\left( \sum_{j=1}^m LW_{(m,n)} \cdot \left( \frac{2}{1 + \exp(-2 \cdot (\sum_k IW_{(m,k)} \cdot X_{(k)} + b_{1(m)}) - 1)} \right) + b_{2(n)} \right) + 1 \right]}{2} \right] \cdot \alpha \right\} + 170 \quad \text{Ec. (5.1)}$$

donde *IW* corresponden a los pesos sinápticos entre la capa de entrada y la capa oculta; *LW* los pesos entre la capa oculta y la capa de salida; *b*<sub>1</sub> los valores umbral para las neuronas ocultas; *b*<sub>2</sub> los valores umbral para las neuronas de salida y los subíndices *k, m, n* se refieren a las neuronas de entrada, oculta y salida, respectivamente.

**Programa GaS\_GeoT con nuevos y mejorados geotermómetros de gases.** Los 8 nuevos geotermómetros de gases (derivados de las 8 modelos de ANNs seleccionados; a partir de ahora denotados como GasG<sub>1</sub> – GasG<sub>8</sub>) fueron programados en GaS\_GeoT (Java) para su fácil y practica utilización (ver Fig. 5.1).

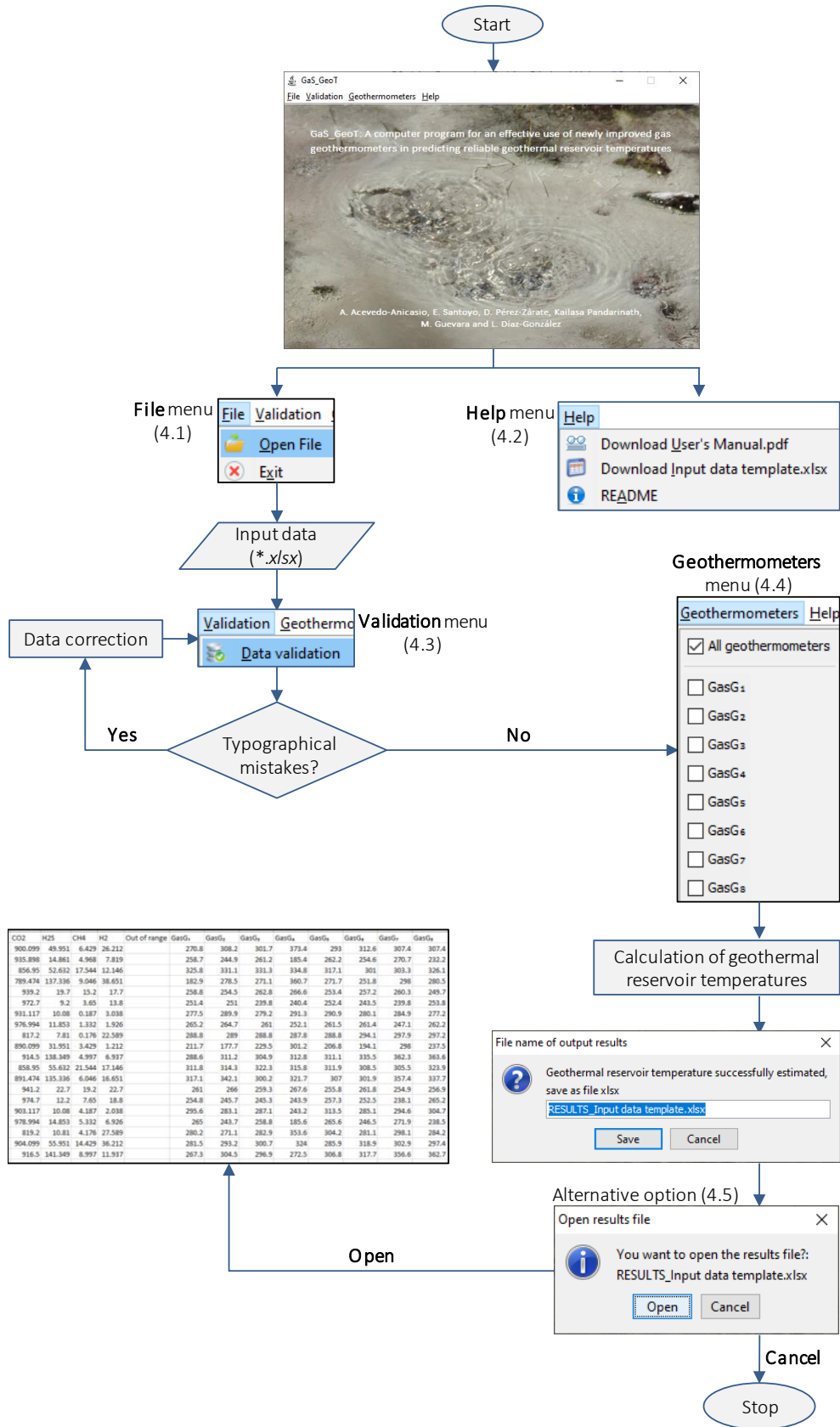


Fig. 5.1 Diagrama general de operación del programa GaS\_GeoT.

Para evaluar el desempeño de los nuevos geotermómetros y compáralos con 25 geotermómetros de la literatura, se compilo una base de datos de validación con 265 muestras geoquímicas de gases. Estos fueron separados en muestras provenientes de sistemas de líquido— y vapor—dominante, con 178 y 85 muestras, respectivamente.

Los resultados finales del análisis comparativo mostraron que los nuevos geotermómetros GasG<sub>1</sub> y GasG<sub>8</sub> se posicionaron en los primeros 8 lugares respecto a un ranking en predicción para sistemas de líquido—dominante. Con valores de RMSE que van desde los 31.6 hasta 34.7, mientras que los resultados de geotermómetros de la literatura son de 36.9 en adelante (p. ej. AG85b: geotermómetro propuesto por Arnórsson & Gunnlaugsson, 1985). Por otro lado, para los sistemas de vapor—dominante los geotermómetros GasG<sub>3</sub> y GasG<sub>1</sub> estimaron los valores más bajos, con 35.6 y 38, respectivamente. Seguidos del geotermómetro ND84c y A98c con valores de 38.9 y 39.8, respectivamente (ND84c: propuesto por Nehring & D'Amore, 1984; A98c: propuesto por Arnórsson et al., 1998). Favor de consultar el ANEXO 2 para más detalles de este extenso análisis comparativo.

## **5.2 Resultados del Módulo 2**

En esta sección se presentan los resultados del desarrollo de un nuevo modelo de predicción para modelar anomalías o emisiones naturales de CO<sub>2</sub> en sistemas geotérmicos.

En el ANEXO 3 se presentan los resultados del entrenamiento de las ANNs para modelar el flujo de CO<sub>2</sub> medido en Acoculco, Puebla (BDEddy). El periodo de medición fue de los meses marzo, abril y mayo del año 2016, lo que consideramos el periodo de secas. Se seleccionó este periodo para evitar la humedad y la afectación de las mediciones en condiciones climáticas de lluvia.

Las arquitecturas que exhibieron los mejores resultados se enlistan en la Tabla 1 del ANEXO 3, siendo la mejor de ellas la ANN-4 que utiliza 6 neuronas en la capa

de entrada (RN: radiación neta, LE: calor latente, H: calor sensible, TA: temperatura del aire, TS: temperatura del suelo y Tdew: temperatura del punto de rocío). En la capa oculta 14 neuronas y una neurona en la capa de salida para modelar el flujo de CO<sub>2</sub>. Esta misma arquitectura, presentó los siguientes resultados en términos de coeficiente de correlación entre el valor real y el simulado  $r_{\text{global}}=0.8903$ ,  $r_{\text{entrenamiento}}=0.91$ ,  $r_{\text{validación}}=0.89$  y  $r_{\text{test}}=0.86$ . Así mismo, los valores más bajos de RMSE y MAE, con 5.05 y 2.81, respectivamente.

En la Fig. 5 del ANEXO 3 se muestran los valores medidos del flujo de CO<sub>2</sub> y simulados por la ANN-4. Y en la Fig. 7 del mismo ANEXO la gráfica de regresión lineal entre ambos valores. Posteriormente al análisis de sensibilidad por el algoritmo de Garson (1991), se observó que la variable más importante con respecto a su contribución para simular el flujo de CO<sub>2</sub> es el calor latente (LE); con valores que van desde el 12.2% hasta 34.6% (ver Tabla 3, ANEXO 3). Con estos resultados observamos que no solo es posible sino viable la implementación de las ANNs en el modelado de anomalías geotérmicas. Posteriormente a la publicación del ANEXO 3, se procedió al análisis de *gaps* para poder modelarlos mediante simulación de las ANNs. Sin embargo, observamos que en ausencia del flujo de CO<sub>2</sub>, también faltaban las mediciones del calor latente. Esto nos condujo a un nuevo planteamiento para desarrollar una nueva arquitectura que pudiera simular el flujo de CO<sub>2</sub> aun en ausencia de la variable LE.

Para ello, se procedió a realizar la selección de variables adecuadas al modelo utilizando la técnica *Stepwise Selection (forward selection)*, la cual sugiere agregar una a una las variables y evaluar el desempeño del modelo neuronal. Cuando el modelo mejora, la variable recién agregada se conserva, de lo contrario se descarta y se continua con el resto de las variables. El resultado de esta selección de variables se muestra en la Tabla 5.1 con 16 de las mejores arquitecturas entrenadas para estimar el flujo de CO<sub>2</sub> censado con el equipo micrometeorológico *Eddy-Covariance*.

Tabla 5.1 Principales resultados del entrenamiento neuronal para simular el flujo de CO<sub>2</sub>.

ANN	# Variables de entrada	# Neuronas ocultas	Arquitectura	Métricas estadísticas de evaluación					
				(r)-global	(r)-entrenamiento	(r)-test	RMSE	MAE	MAPE
1	6	21	[6-21-1]	0.8722	0.8803	0.8594	7.6004	3.9412	1.4819
2	7	23	[7-23-1]	0.8262	0.8611	0.7637	8.7756	4.4442	1.6710
3	8	23	[8-23-1]	0.8919	0.9361	0.8104	7.0682	3.4853	1.3105
4	9	22	[9-22-1]	0.8635	0.8758	0.8405	7.8306	4.0838	1.5355
5	10	15	[10-15-1]	0.8382	0.8435	0.8342	8.4718	4.2135	1.5843
6	11	9	[11-9-1]	0.8559	0.8597	0.8489	8.0292	4.0048	1.5058
7	12	14	[12-14-1]	0.9028	0.9095	0.8915	6.7120	3.8106	1.4328
8	13	6	[13-6-1]	0.8631	0.8688	0.8578	7.8645	4.2184	1.5861
9	14	5	[14-5-1]	0.8440	0.8609	0.8570	8.4891	4.3076	1.6197
10	15	16	[15-16-1]	0.8808	0.8971	0.8620	7.4045	3.9560	1.4874
11	16	5	[16-5-1]	0.8727	0.8787	0.8635	7.5797	3.9505	1.4854
12	17	20	[17-20-1]	0.8257	0.8359	0.8128	8.7901	4.3970	1.6533
13	18	21	[18-21-1]	0.8922	0.9107	0.8561	7.0107	3.7910	1.4254
14	19	10	[19-10-1]	0.8776	0.8976	0.8385	7.4365	3.9081	1.4695
15	20	5	[20-5-1]	0.8753	0.8944	0.8427	7.5185	3.9338	1.4791
16	21	8	[21-8-1]	0.8579	0.8768	0.8205	7.9783	4.0974	1.5406

**MCDA para evaluar modelos neuronales.** Para evaluar las arquitecturas de la Tabla 5.1 y seleccionar la que exhiba el mejor desempeño, se implementó el método *MCDA* descrito ampliamente en el ANEXO 2. Como primer paso se normalizaron las variables de acuerdo con las Ecs. (2) y (3) del ANEXO 2 en las cuales se sugiere normalizar las métricas estadísticas de la Tabla 5.1 de acuerdo con su representación. Por ejemplo, si nos referimos al coeficiente de correlación lineal ( $r$ ) obtenido entre los valores medidos del flujo de CO<sub>2</sub> y los valores estimados por un modelo de red neuronal artificial, sabemos a priori que la correlación perfecta estaría representada por la unidad. Es decir, todos los valores cercanos a la unidad se deben normalizar en términos de maximización (Ec. 3, ANEXO 2). Por otro lado, las métricas estadísticas (RMSE, MAE y MAPE) utilizadas para medir el error obtenido entre los valores medidos en campo y simulados del flujo de CO<sub>2</sub>, se deben normalizar en términos de minimización (Ec. 2, ANEXO 2), dado que buscamos el modelo neuronal que exhiba los errores más bajos de predicción.

Tabla 5.2 Resultados del entrenamiento neuronal para simular el flujo de CO<sub>2</sub> normalizados (0-10).

ANN	# Variables de entrada	# Neuronas ocultas	Arquitectura	Métricas estadísticas de evaluación normalizadas					
				(r)-global	(r)-entrenamiento	(r)-test	RMSE	MAE	MAPE
1	6	21	[6-21-1]	6.03	4.43	7.49	5.72	5.25	5.25
2	7	23	[7-23-1]	0.06	2.52	0.00	0.07	0.00	0.00
3	8	23	[8-23-1]	8.58	10.00	3.65	8.29	10.00	10.00
4	9	22	[9-22-1]	4.90	3.98	6.01	4.62	3.76	3.76
5	10	15	[10-15-1]	1.62	0.75	5.51	1.53	2.41	2.41
6	11	9	[11-9-1]	3.91	2.37	6.66	3.66	4.58	4.58
7	12	14	[12-14-1]	10.00	7.34	10.00	10.00	6.61	6.61
8	13	6	[13-6-1]	4.85	3.28	7.37	4.45	2.35	2.35
9	14	5	[14-5-1]	2.37	2.49	7.30	1.45	1.42	1.42
10	15	16	[15-16-1]	7.14	6.10	7.70	6.67	5.09	5.09
11	16	5	[16-5-1]	6.09	4.27	7.81	5.82	5.15	5.15
12	17	20	[17-20-1]	0.00	0.00	3.84	0.00	0.49	0.49
13	18	21	[18-21-1]	8.63	7.46	7.23	8.56	6.81	6.81
14	19	10	[19-10-1]	6.73	6.15	5.85	6.51	5.59	5.59
15	20	5	[20-5-1]	6.43	5.84	6.18	6.12	5.32	5.32
16	21	8	[21-8-1]	4.17	4.08	4.45	3.91	3.62	3.62

En la Tabla 5.2 se observan las métricas estadísticas de evaluación normalizadas de acuerdo con el método *MCDA* para evaluar integralmente las 16 arquitecturas. Como se puede observar, algunos valores de la Tabla 5.2 aparecen como 0.00 (el más bajo) y otros como 10 (más alto). Esto se debe a que al momento de la normalización ya sea en términos de minimización o maximización (Ecs. 2 y 3, ANEXO 2) esos valores son los datos extremos. Por ejemplo, la ANN-3 en la columna 6 de izquierda a derecha, corresponde al coeficiente de correlación obtenido para el set de entrenamiento entre los datos medidos de flujo de CO<sub>2</sub> y datos estimados por la red neuronal. Aparece con el valor igual a 10 dado que en la Tabla 5.1 es el coeficiente más alto en toda la columna 6, con un valor de 0.9361. De forma similar, el valor igual a cero de la ANN-2 en la columna 7 (r-test en Tabla 5.2) representa el valor más bajo obtenido para el set de prueba durante el entrenamiento de la ANN-2, con un valor igual a 0.7637.

Continuando con el método *MCDA* para evaluar integralmente las 16 arquitecturas neuronales, se procedió a la creación de distintos escenarios con ponderaciones variadas en cada una de las métricas estadísticas normalizadas.

Tabla 5.3 Cuatro escenarios creados con ponderaciones diferentes en los coeficientes de correlación.

Métricas de evaluación	Escenario 1		Escenario 2		Escenario 3		Escenario 4	
	Pesos ponderados	Pesos (%)	Pesos ponderados	Pesos (%)	Pesos ponderados	Pesos (%)	Pesos ponderados	Pesos (%)
(r) - global	10	16.67	<b>20</b>	28.57	10	14.29	10	14.29
(r) - entrenamiento	10	16.67	10	14.29	<b>20</b>	28.57	10	14.29
(r) - test	10	16.67	10	14.29	10	14.29	<b>20</b>	28.57
RMSE	10	16.67	10	14.29	10	14.29	10	14.29
MAE	10	16.67	10	14.29	10	14.29	10	14.29
MAPE	10	16.67	10	14.29	10	14.29	10	14.29
Total	60	100%	70	100%	70	100%	70	100%

En la Tabla 5.3 están representados los cuatro escenarios creados para evaluar el desempeño integralmente de las 16 arquitecturas seleccionadas para modelar el flujo de CO<sub>2</sub>. En el primer escenario los pesos ponderados tienen un valor igual a 10 puntos con la finalidad de dar un equilibrio a cada métrica estadística utilizada. Con el objetivo de evaluar integralmente todas las arquitecturas y seleccionar la que exhiba el mejor desempeño para modelar el flujo de CO<sub>2</sub> sin caer en el sobreajuste de la red neuronal; en el segundo, tercer y cuarto escenario se le da una ponderación mayor al coeficiente de correlación global, de entrenamiento y de test, respectivamente (resaltado en negrita). Los resultados de la ponderación asignada en los cuatro escenarios se muestran en las Figs. 5.2 - 5.5

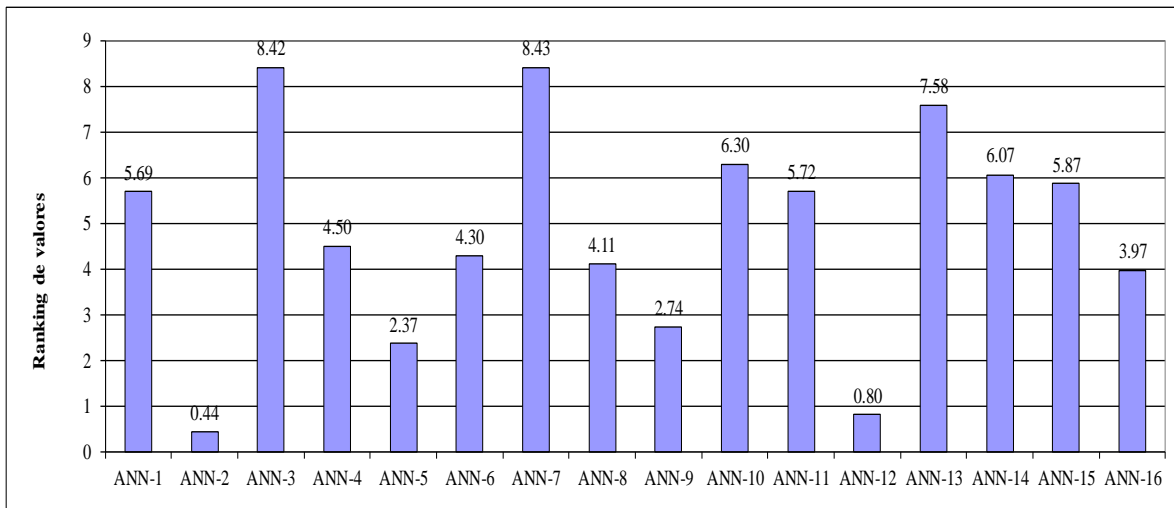


Fig. 5.2 Resultados de la ponderación equilibrada a 10 puntos en el escenario 1.



En la Fig. 5.2 sobresalen dos arquitecturas con el puntaje más alto en términos de un ranking general para evaluar las 16 arquitecturas seleccionadas para modelar el flujo de CO<sub>2</sub>. La ANN-3 y la ANN-7 con valores 8.42 y 8.43, respectivamente.

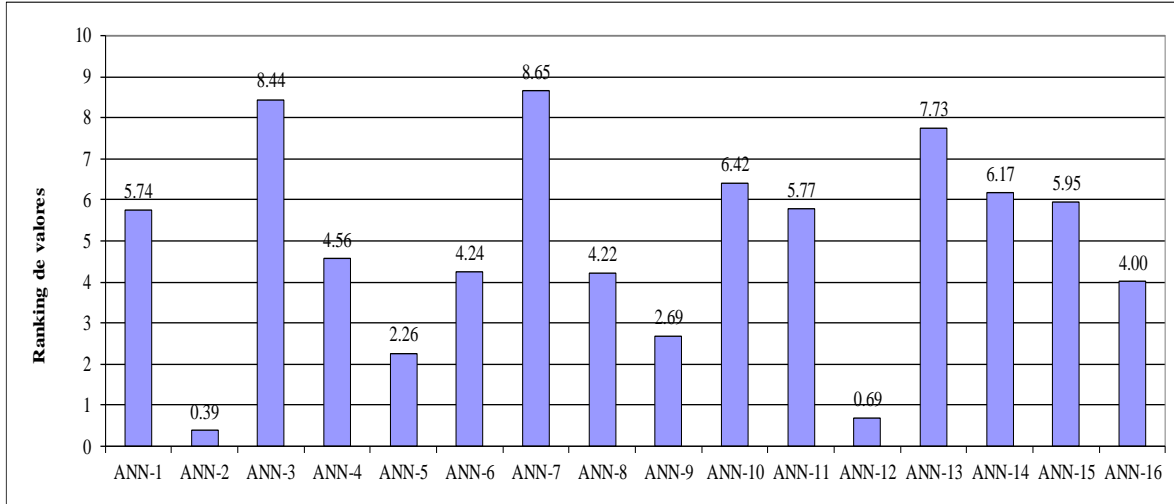


Fig. 5.3 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación global en el escenario 2.

En la Fig. 5.3 nuevamente aparece las arquitecturas ANN-3 y ANN-7 posicionadas en el ranking superior con valores de 8.44 y 8.65, respectivamente.

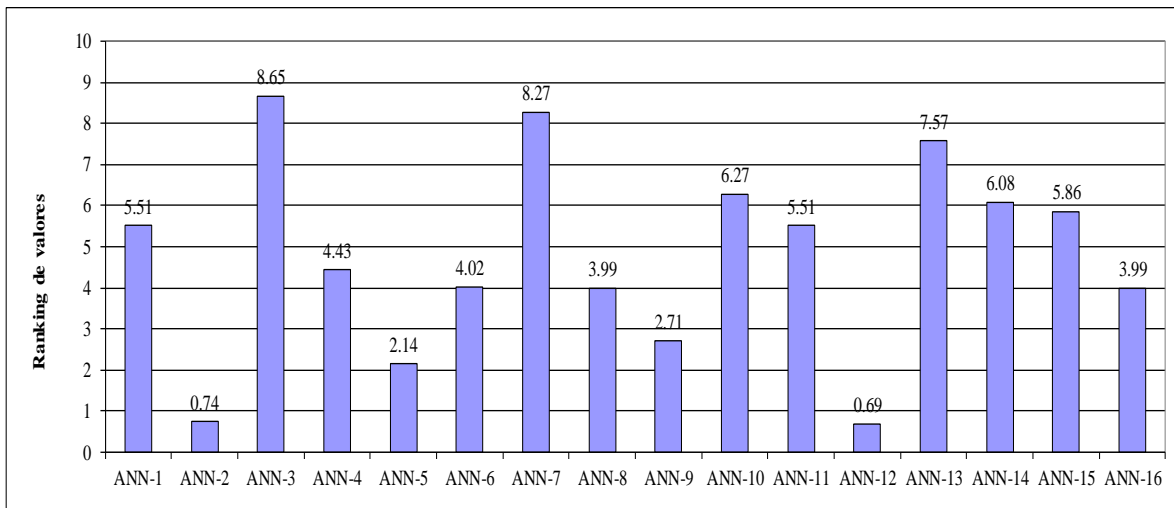


Fig. 5.4 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación de entrenamiento en el escenario 3.

En la Fig. 5.4 la ANN-3 se posiciona en mejor ranking con respecto a la ANN-7, con valores de 8.65 y 8.27, respectivamente. Esto tiene sentido si revisamos los valores

originales del coeficiente de correlación obtenido en el set de entrenamiento de ambas arquitecturas, donde se tienen coeficientes ( $r$ ) de 0.9361 y 0.9095 de la Tabla 5.1.

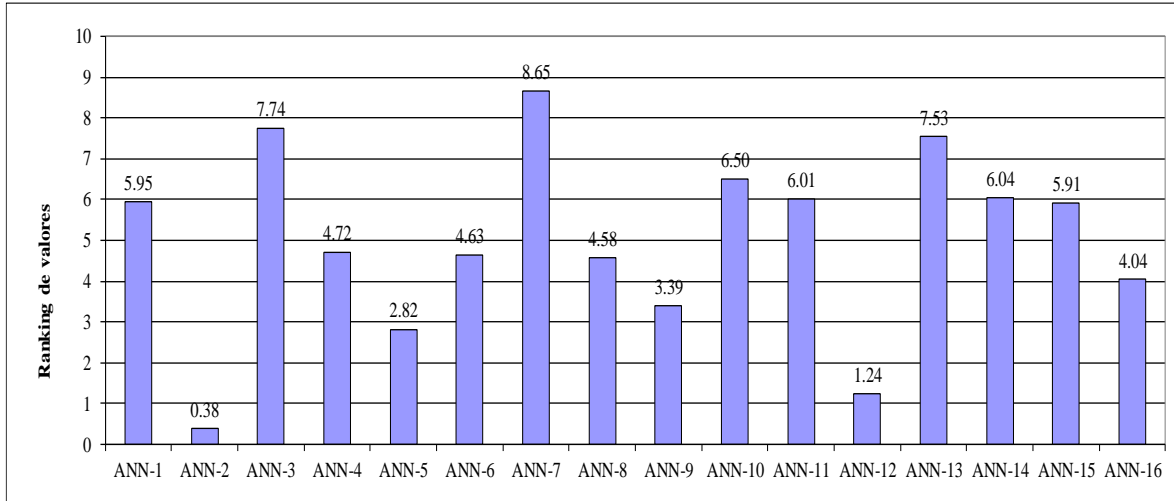


Fig. 5.5 Resultados de la ponderación de 20 puntos asignados al coeficiente de correlación del conjunto de prueba en el escenario 4.

En la Fig. 5.5 aparecen tres arquitecturas con los valores más altos del ranking, la ANN-3, ANN-7 y ANN-13 con valores de 7.74, 8.65 y 7.53, respectivamente. Dado que la creación de los cuatro escenarios fue evaluar las 16 arquitecturas asignando pesos ponderados diferentes en cada escenario, la selección de la arquitectura con base al método *MCDA* está basada a un promedio del desempeño en los cuatro escenarios, ver Tabla 5.4.

Tabla 5.4 Resultado final del método *MCDA* para seleccionar la mejor arquitectura neuronal.

No.	ANN	Valor promedio
1	<b>ANN-7</b>	<b>8.4998</b>
2	ANN-3	8.3122
3	ANN-13	7.6044
4	ANN-10	6.3711
5	ANN-14	6.0903
6	ANN-15	5.8984
7	ANN-11	5.7518
8	ANN-1	5.7242
9	ANN-4	4.5526
10	ANN-6	4.2971

No.	ANN	Valor promedio
11	ANN-8	4.2233
12	ANN-16	4.0008
13	ANN-9	2.8825
14	ANN-5	2.3995
15	ANN-12	0.8550
16	ANN-2	0.4855

**Resultados del MCDA – Ecuación neuronal.** Con base MCDA, se seleccionó la ANN-7 como la mejor arquitectura para modelar el flujo de CO<sub>2</sub> medido con la estación *Eddy-Covariance*. Este resultado está respaldado por una metodología en la cual se evaluaron 16 arquitecturas neuronales con los valores obtenidos durante el entrenamiento, siendo la ecuación general de la ANN-7 la siguiente:

$$ANN - 7 = \left\{ \left[ \frac{\left( \sum_{j=1}^m LW_{(m,n)} \cdot \left( \frac{2}{1 + \exp(-2 \cdot (\sum_k IW_{(m,k)} \cdot X_{(k)} + b_{1(m)}))} - 1 \right) \right) + b_{2(n)} \right] + 1 \right]}{2} \cdot 285.34 \right\} + 0.011 \quad \text{Ec. (5.2)}$$

donde *IW* corresponden a los pesos sinápticos entre la capa de entrada y la capa oculta; *LW* los pesos entre la capa oculta y la capa de salida; *b*<sub>1</sub> los valores umbral para las neuronas ocultas; *b*<sub>2</sub> para las neuronas de salida y los subíndices *k, m, n* se refieren a las neuronas de entrada, oculta y salida, respectivamente.

Los pesos sinápticos (coeficientes) relacionados a la Ec. 5.2 se presentan en las Tablas 5.5 y 5.6.

Tabla 5.5 Coeficientes de la ANN-7 asociados a la capa de entrada y la capa oculta de Ec. 5.2

Neurona	<i>IW</i> ( <i>m,k</i> )											
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12
1	0.4211	-2.1936	0.9194	-1.0572	0.5003	-0.0956	0.6508	1.0783	-2.2869	2.3360	1.4374	-0.0131
2	-0.7503	1.8381	-0.6254	0.0700	-0.2060	0.0456	-1.3657	-1.2267	-0.2514	-2.2177	-2.9124	1.3949
3	0.5930	-0.0909	-1.0887	-1.0802	-0.2257	0.3807	-1.8703	-0.7563	4.9948	-0.5253	-1.0179	-0.6228
4	1.0226	0.4837	2.3319	0.7987	-1.7274	-0.0978	-0.2823	1.8874	0.5583	-0.0490	0.2896	-0.1836
5	-0.2106	-0.4979	0.4173	-0.4116	0.0544	0.0077	-0.2414	0.7151	-1.6957	-0.3937	-1.1312	0.0380
6	-0.4436	2.0076	-0.7901	0.7446	-0.5980	0.0782	-0.0227	-0.7268	1.7636	-0.6004	-1.4344	-0.2322
7	-0.1886	-1.2242	-0.0225	-0.1488	-0.2213	0.1709	-0.2216	0.9842	-1.6530	-0.3793	-1.4542	0.4179
8	-2.0956	-0.5002	-5.0462	2.0089	5.7706	-0.6412	-4.6180	3.5481	2.4116	2.0858	4.2653	-1.0002

Neurona	$IW(m,k)$											
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12
9	-0.2078	-1.4839	0.0501	-0.3476	-0.1698	0.0575	0.0232	0.5636	-0.7225	0.1577	-1.2079	0.2866
10	1.9982	0.9637	4.8947	-1.9816	-5.7732	0.6764	4.7344	-4.2834	-2.1666	-2.0466	-4.0588	1.0445
11	-1.4980	3.8954	0.3621	2.9022	7.2693	0.2640	1.9033	-4.8820	0.4938	-0.9972	-2.2605	1.8593
12	0.3657	-3.8410	2.7849	-0.3528	-1.1611	-2.6584	2.9443	1.3703	-8.0885	-0.7125	-1.0347	1.4109
13	0.2341	0.1273	-0.2827	0.3443	0.0110	0.0481	-0.0770	-0.4753	0.8914	-0.0050	0.9987	0.0467
14	-0.1486	0.8889	-0.3961	0.5969	-0.1458	0.1768	-0.5300	1.2934	-1.2543	-2.1138	-1.4666	0.5806

ANN-7 ( $k = 12, m = 14, n = 1$ ); Var1-Var12 representan las 12 variables de entrada utilizadas por la ANN-7.

Tabla 5.6 Coeficientes correspondientes a la ANN-7 asociados a la capa oculta y la capa de salida de Ec. 5.2

Neurona	$b1(m,1)$	$LW(m,n)$	$b2(n)$
	Bias	Pesos	Bias
1	-1.8534	0.5140	2.5273
2	-3.5582	-0.2139	
3	-0.3980	-0.1514	
4	-1.0520	-0.3713	
5	-0.7809	1.5561	
6	0.0704	0.6434	
7	-0.9344	-1.6367	
8	-0.3950	-2.9405	
9	-1.4933	1.7115	
10	-0.0953	-2.8834	
11	3.3766	-2.8007	
12	2.0182	-1.1518	
13	1.1141	1.9736	
14	0.8999	0.5107	

ANN-7 ( $k = 12, m = 14, n = 1$ )

Bajo estos resultados, la ANN-7 utilizó las siguientes variables: RN (radiación neta), H (calor sensible), SHF (flujo de calor del suelo), TA (temperatura del aire), TS (temperatura del suelo), SWC (contenido de agua en el suelo), %\_signal (relación entre la señal de CO<sub>2</sub> y la señal de CO<sub>2</sub> esperada), TKE (energía cinética turbulenta), Pitch (ángulo de rotación), Air\_heat\_capacity (calor específico a presión constante del aire en el ambiente), W\_var (varianza del sensor para medir el viento), RH (humedad relativa). La contribución relativa de cada variable utilizada (análisis de sensibilidad) durante el entrenamiento de dicha arquitectura, se obtuvieron los valores de la Tabla 5.7.

Tabla 5.7 Variables utilizadas para modelar el flujo de CO<sub>2</sub>.

Id	Variable	Contribución relativa (%)
1	RN	4.7
2	H	11.2
3	SHF	8.1
4	TA	6.3
5	TS	7.4
6	SWC	2.0
7	%_signal	6.7
8	TKE	11.4
9	Pitch	16.5
10	Air_heat_capacity	7.2
11	W_var	14.4
12	RH	4.1

**Mediciones del flujo de CO<sub>2</sub>.** Estas mediciones censadas durante el periodo de medición (marzo, abril y mayo), se pudo identificar un comportamiento periódico diario con respecto al día y a la noche. Esto tiene que ver con el ecosistema y la activación de la fotosíntesis por la radiación solar (emisión de oxígeno y secuestro de CO<sub>2</sub>) como aparece en la Fig. 5.6.

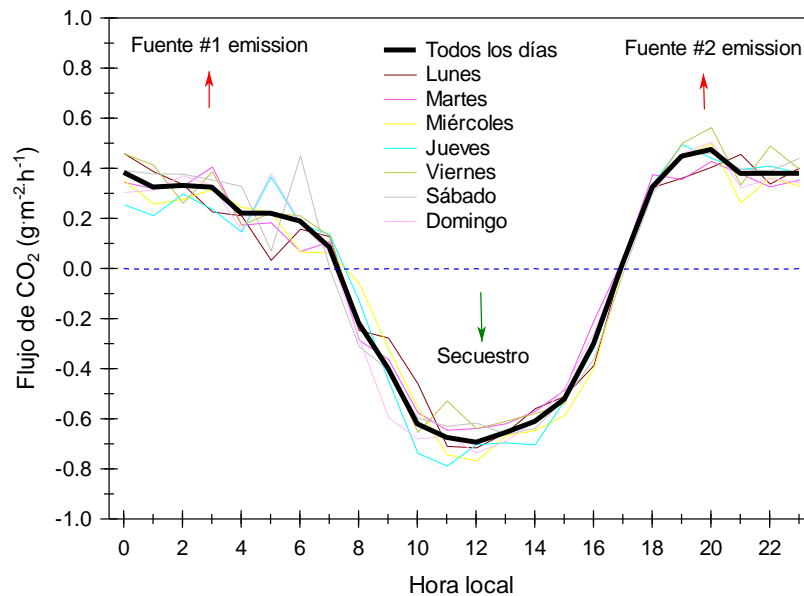


Fig. 5.6 Perfil del flujo de CO<sub>2</sub> por hora.

En la Fig. 5.6 se puede observar que aproximadamente de las 7:00 hrs. a las 18:00 hrs. existen mediciones negativas (secuestro de CO<sub>2</sub>) y este perfil es constante

independientemente del día. Así mismo, en la Fig. 5.7 se muestra el perfil por día durante el mismo periodo de medición.

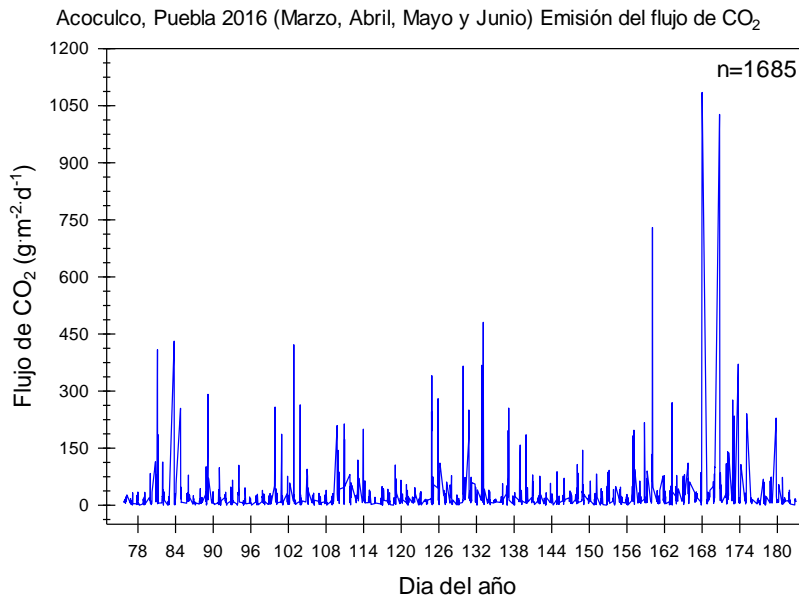


Fig. 5.7 Perfil del flujo de CO<sub>2</sub> por día.

En la Fig. 5.7 aparecen mediciones fuera del perfil (picos en color azul), las cuales pueden originarse por varias causas: (i) mediciones reales en un periodo corto ocasionado por la respiración natural del ecosistema; (ii) datos anómalos ocasionados por medición de flujo de CO<sub>2</sub> no necesariamente del ecosistema, p. ej. antropogénico, mezcla de emisiones de lejanas que por la dirección del viento, fueron censadas por el equipo Eddy siendo que provenían de otro lado; y, (iii) error de medición en el equipo, dado que como todo sistema electrónico puede presentar fallas siendo consideradas dentro de esta técnica de medición hasta un 5% de valores a tomar con reserva (Burba, 2013) .

El resultado de utilizar las variables referidas en la Tabla 5.5, fue la obtención de valores de coeficiente de correlación ( $r$ ) global=0.9028, entrenamiento=0.9095, y test=0.8915. Dados estos valores principalmente en el set de test, podemos asegurar que el modelo neuronal no tuvo un sobreajuste.

En la Fig. 5.8 se graficó una regresión lineal entre los valores medidos y simulados del flujo de CO<sub>2</sub>.

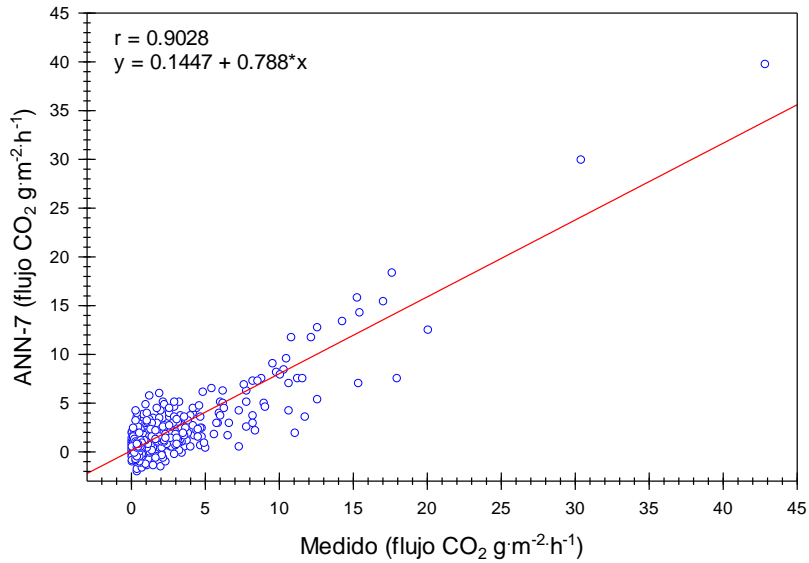


Fig. 5.8 Regresión lineal entre valores medidos y simulados del flujo de CO<sub>2</sub>.

**Validación estadística – Prueba F y t-student.** Evaluando estadísticamente los valores predichos por la ANN-7 para las mediciones del flujo de CO<sub>2</sub>, se aplicó la prueba de F y t-student con el objetivo de determinar si los valores calculados por el modelo neuronal son estadísticamente similares y que provienen de una misma población.

El planteamiento fue el siguiente: al 99% de confianza, para la prueba F, la hipótesis H<sub>0</sub>=las dos muestras tienen la misma varianza y H<sub>1</sub>=Existen diferencias significativas en las varianzas de las muestras. Los resultados son F=1.1537 siendo el valor crítico de referencia 1.2815, por lo cual la H<sub>0</sub> es aceptada al ser menor el valor de F comparado con el valor crítico.

Para la prueba t-student la hipótesis H<sub>0</sub>=las medias de las muestras son iguales y H<sub>1</sub>=existen diferencias significativas en las medias de las muestras. Los resultados de t=0.66099 y el valor crítico correspondiente es 2.33316, por lo tanto, la H<sub>0</sub> es aceptada por lo que las medias de las muestras son iguales.

En la Fig. 5.9 se observan los valores medidos (color rojo) y los valores simulados (línea punteada en color azul) del flujo de CO<sub>2</sub>.

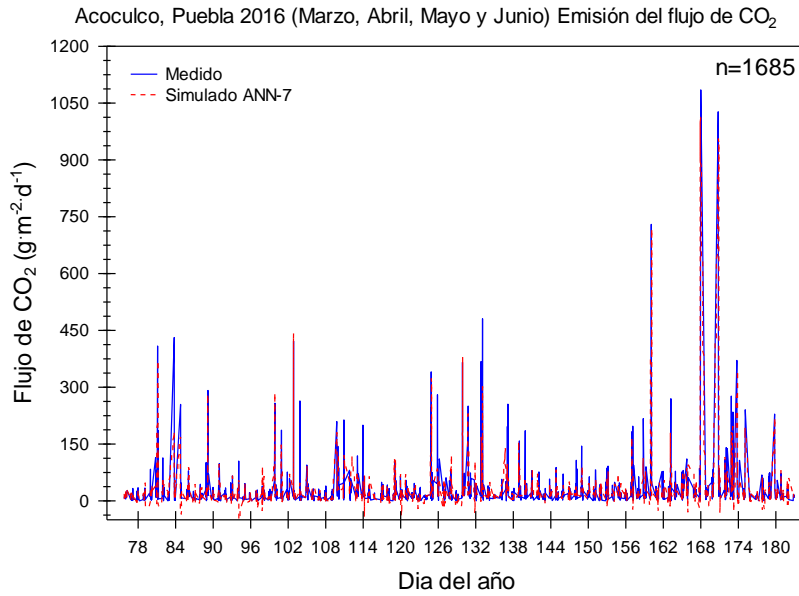


Fig. 5.9 Perfil por día del flujo de CO<sub>2</sub> medido versus simulado.

**Identificación y relleno de *gaps*.** Continuando con los objetivos del Módulo 2, se procedió a identificar los *gaps* donde hay ausencia de mediciones. En el periodo de medición marzo, abril, mayo y junio se encontraron un total de 208 *gaps*. Estos están distribuidos de la siguiente manera:

- 35 *gaps* que representan un 16.8% del total de las mediciones a 30 minútales durante el mes de marzo;
- 70 *gaps* durante el mes de abril equivalente a un 33.6% de las mediciones durante el mes;
- 24 *gaps* en el mes de mayo con un porcentaje del 11.5% del total de las mediciones; y
- 79 *gaps* observados en el mes de junio que representa el 37.9% de las mediciones durante el mes.

En la Fig. 5.10 se calcularon los valores ausentes del flujo de CO<sub>2</sub> utilizando el modelo neuronal ANN-7. Como se observa los valores exhiben un perfil similar a las mediciones de la Fig. 5.7 donde aparecen solo las mediciones reales.



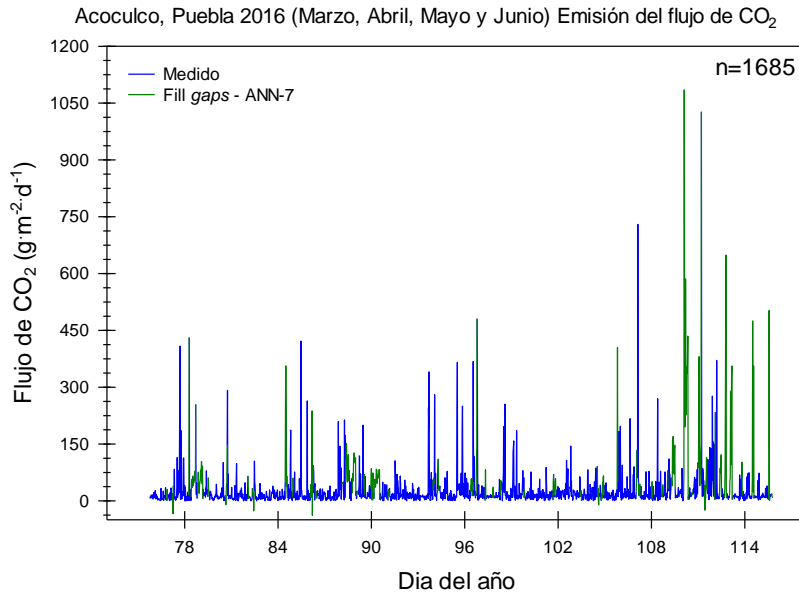


Fig. 5.10 Perfil por día del flujo de CO<sub>2</sub> con relleno de *gaps* utilizando la ANN-7.

Dado que se tienen 1685 datos medidos en el periodo, poder visualizar a detalle en la Fig. 5.10 es poco práctico, dicha figura nos exhibe un perfil completo. Un acercamiento de estos datos simulados se observa en la Fig. 5.11 (A) y (B).

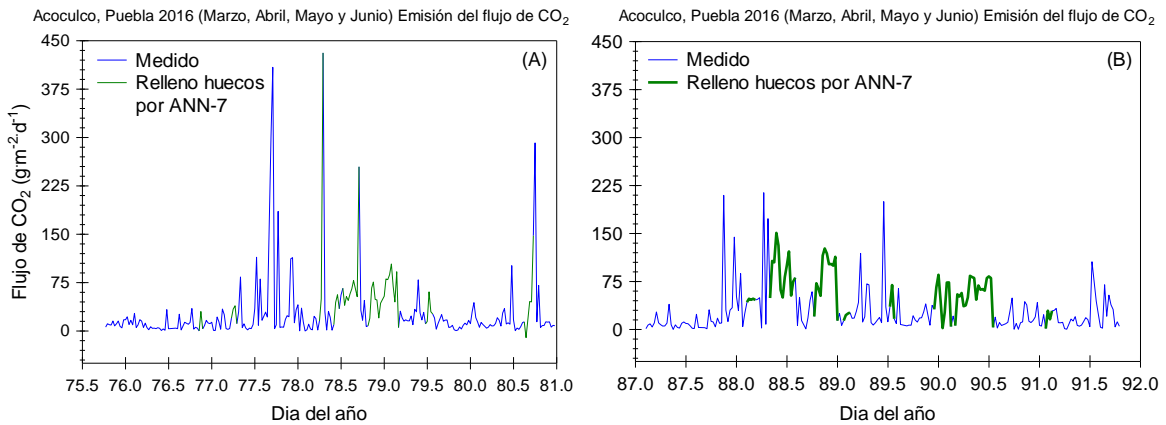


Fig. 5.11 Relleno de huecos en el mes de marzo (A) y abril (B).

Como se puede observar en el acercamiento de la Fig. 5.11, los huecos rellenos por la ANN-7 es una aproximación muy cercana al perfil con las mediciones reales. Por lo que la selección de variables, el entrenamiento de las ANNs, la selección del mejor modelo (ANN-7) utilizando la metodología de *MCDA* y la predicción de los huecos se desarrolló con éxito.

### 5.3 Resultados del Módulo 3

En esta sección se presentan los resultados del desarrollo de una nueva metodología geoquímica computacional apoyada en el método de Sinclair para la discriminación del transporte de gas en sistemas geotérmicos.

Durante las campañas de muestreo, se visitó la zona de Los Azufres ubicada dentro de la caldera de Acoculco, Puebla. Esta zona es conocida por su modesta actividad hidrotermal, dado el planteamiento en el Capítulo 2 como un sistema geotérmico de roca seca caliente “oculto”. El muestreo del flujo de CO<sub>2</sub> en la superficie de la Tierra se realizó con una Cámara de Acumulación como se muestra en la Fig. 5.12.



Fig. 5.12 Mediciones del flujo de CO<sub>2</sub> con el equipo de medición montado (A) e instalado en el subsuelo (B).

**Resultados con el método de Sinclair mejorado.** Aplicando el método de Sinclair a los datos medidos en Los Azufres, con base a la metodología descrita la sección 4.3 del Capítulo 4 resaltando la solución propuesta para la selección del punto de inflexión no arbitrariamente, tenemos la gráfica final representada en la Fig. 5.13. De acuerdo con el método de Sinclair, existe un grupo definido con los valores más altos del flujo de CO<sub>2</sub> pertenecientes al Grupo A, desde de los valores en escala logarítmica de 4.6 (g m<sup>-2</sup> d<sup>-1</sup>) hacia arriba, representada por la línea discontinua "a". Así mismo, existe una mezcla en los Grupos A y B comprendido entre los valores de 3.4 y 4.6, líneas discontinuas "b" y "a", respectivamente.

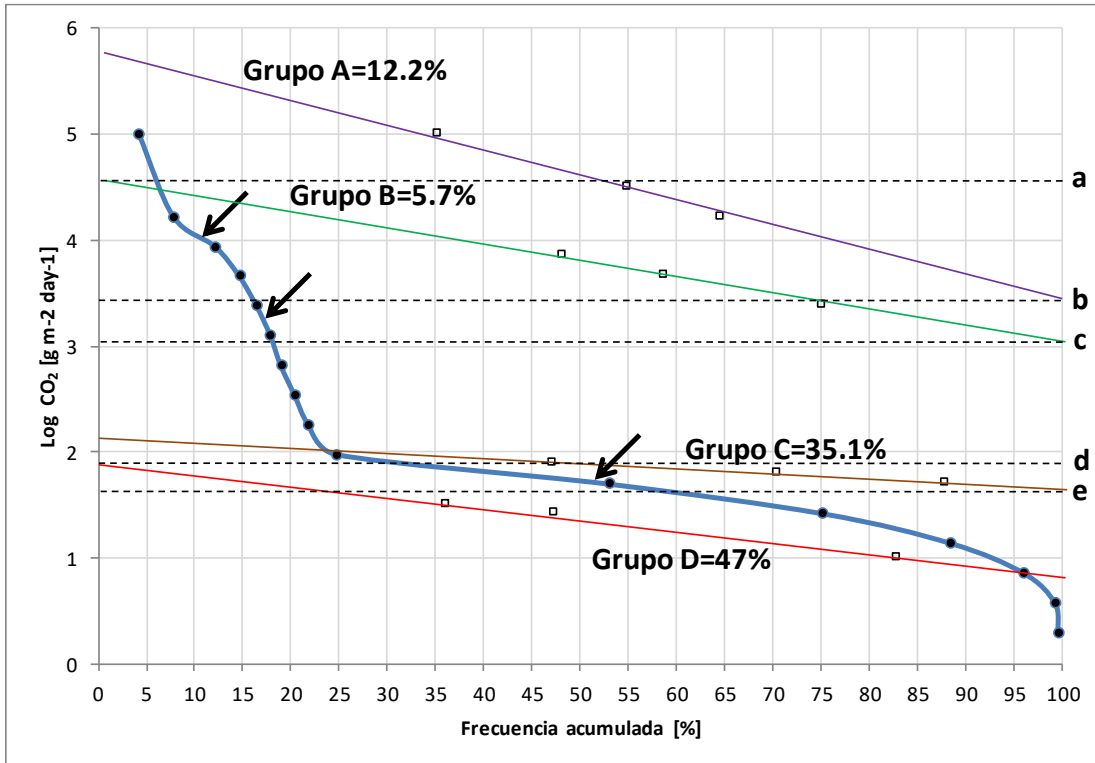


Fig. 5.13 Grafica de los resultados finales al aplicar el método de Sinclair mejorado.

. El Grupo B está definido entre los valores de 3 y 4.6, líneas "c" y "a", respectivamente. Y solo está presente el Grupo B entre los valores de 3 y 3.4 líneas "c" y "b", respectivamente.

El Grupo C está definido entre los valores de 1.9 y 3, líneas "d" y "c", respectivamente. Así mismo, existe una mezcla entre los Grupos C y D entre los valores de 1.6 y 1.9, líneas "e" y "d", respectivamente.

El grupo D con los valores más bajos del flujo de CO<sub>2</sub> está presente a partir del valor 1.6 hacia abajo, línea "e".

**Interpretación de resultados.** De los 279 datos graficados a través de la frecuencia acumulada (mediciones de flujo de CO<sub>2</sub> compilados en la BDCA), se derivan cuatro posibles orígenes del flujo de CO<sub>2</sub> medido en Los Azufres:

- De acuerdo con la literatura (Chiodini et al., 1998), las mediciones de CO<sub>2</sub> con los valores más altos corresponden al origen Magmático, es decir, el 12.2% (34 muestras del flujo de CO<sub>2</sub>) pertenecientes al Grupo A son de origen Magmático.
- Las mediciones de origen Biogénico de CO<sub>2</sub> generalmente son las más bajas de acuerdo con el metabolismo de las raíces de las plantas y la mineralización de materia orgánica (Martínez et al., 2008), por lo tanto, el 47% (131 muestras del flujo de CO<sub>2</sub>) pertenecientes al Grupo D corresponden al origen Biogénico.
- De acuerdo con Burba (2013), el CO<sub>2</sub> de origen Atmosférico solo es medible en altas concentraciones por lo que el 5.7% (16 muestras del flujo de CO<sub>2</sub>) identificadas en el Grupo B son de origen Atmosférico.
- Dado que se han identificado los orígenes de los Grupos A, B y D, el último de ellos de acuerdo con la literatura (Thomas, 2018) correspondería al origen Antropogénico (mezcla de varios orígenes) con el 35.1% (98 muestras del flujo de CO<sub>2</sub>) pertenecientes al Grupo C son de origen Antropogénico.

En la Fig. 5.14 se muestra la distribución porcentual con base a los puntos de inflexión identificados en la curva de la frecuencia acumulada del flujo de CO<sub>2</sub> medido en la zona de Los Azufres.

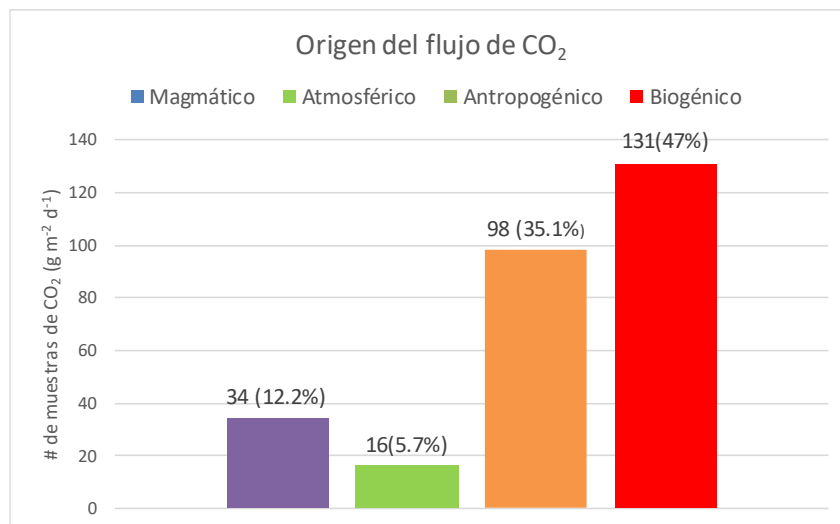


Fig. 5.14 Distribución porcentual del origen del flujo de CO<sub>2</sub> medido en Los Azufres, Acoculco, Puebla.

Los resultados finales de la implementación del método de Sinclair a las mediciones del flujo de CO<sub>2</sub> se presentan en la Tabla 5.8.

Tabla 5.8 Identificación de las muestras de CO<sub>2</sub> con sus respectivos orígenes y mezclas encontradas.

Grupo	Origen flujo CO <sub>2</sub>	# Muestras de un solo origen	Intervalo del flujo CO <sub>2</sub> (g m <sup>-2</sup> d <sup>-1</sup> )	Orígenes de la mezcla	# Muestras de mezclas	Intervalo del flujo CO <sub>2</sub> (g m <sup>-2</sup> d <sup>-1</sup> )
A	Magmático	8	40300 - 79200	Magmático y Atmosférico	33	2920 - 23800
B	Atmosférico	7	1220 - 2210			
C	Antropogénico	15	89 - 801	Antropogénico y Biogénico	30	40.8 - 66
D	Biogénico	186	0.559 - 38.5			
	Subtotal	216		Subtotal	63	Total = 279

De acuerdo a la Fig. 5.13 en la cual se identificaron las muestras pertenecientes a un solo origen y las mezclas entre los mismos; se tienen 8 muestras de origen exclusivamente Magmáticas y 7 Atmosféricas, con 33 muestras mezcladas entre estos dos orígenes. Las concentraciones más altas corresponden al origen Magmático con valores superiores a los 40300 (g m<sup>-2</sup> d<sup>-1</sup>). Así mismo, se tienen 15 muestras Antropogénicas definidas y 186 Biogénicas, con una mezcla entre ambas igual a 30.

Como era de esperarse, la gran mayoría de las muestras son de origen Biogénico con las concentraciones más bajas (0.559 - 38.5 g m<sup>-2</sup> d<sup>-1</sup>) dado que las mediciones se realizaron en el subsuelo a ~40 cm de profundidad y esto nos otorga un conocimiento a priori del metabolismo de las raíces de las plantas y la mineralización de materia orgánica como se reporta en la literatura (Martínez et al., 2008). Así mismo, se pudo desarrollar un procedimiento mejorado para la implementación del método de Sinclair el cual es muy utilizado en estudios de exploración en ciencias de la Tierra, como la geotermia.

## CAPÍTULO 6

### Conclusiones y Trabajos Futuros

#### Conclusiones

El éxito de la presente investigación obedece a varios factores importantes, como son: (i) contar con una amplia base de datos de muestras geoquímicas de gases medidos en pozos geotérmicos productores. Esto nos permitió desarrollar nuevos geotermómetros de gases con una calibración mucho más amplia que los geotermómetros reportados en la literatura; (ii) la evaluación integral de múltiples arquitecturas de RNA con base a una nueva metodología existente en otros campos de la ciencia pero aplicada por primera vez en geotermometría de gases usando RNA, la cual es *Multi-Criteria Decision Analysis: MCDA*; (iii) evaluar las nuevas ecuaciones geotermométricas con base a la predicción de temperaturas en una nueva base de datos de validación, con ello se demostró la efectividad de los nuevos geotermómetros de gases al ser comparados con los reportados en la literatura; (iv) se simuló exitosamente el flujo de CO<sub>2</sub> (censado con el equipo micrometeorológico *Eddy-Covariance*) utilizando RNA como una herramienta confiable para modelar anomalías geotérmicas; (v) se evaluaron los resultados bajo la metodología *MCDA* seleccionando la arquitectura con el mejor desempeño; (vi) se procedió a resolver el problema de ausencia de datos a causa de la aparición de neblina (*gaps*); (vii) se propuso una alternativa viable para discriminar el flujo de CO<sub>2</sub> mediante la implementación del método de Sinclair mejorado; y, (viii) se identificaron las mediciones del flujo de CO<sub>2</sub> (medido con la Cámara de Acumulación) con base a su origen: Magmático, Atmosférico, Antropogénico, Biogénico y las mezclas entre algunos de ellos.

Finalmente, se pudo demostrar que el estudio del transporte de gases de origen geotérmico es una importante fuente de información para ser utilizada durante la exploración y explotación de un recurso geotérmico permitiendo la estimación de temperaturas de fondo con un alto grado de confiabilidad (temperaturas razonables) determinadas por los nuevos geotermómetros de gases y más realistas en

comparación con otras herramientas geotermométricas reportadas en la literatura. Paralelamente, el modelado y discriminación del origen del gas mayoritario en ambientes geotérmicos se logró exitosamente a través de una metodología integral que comprende el uso de las RNA-MCDA y el método de Sinclair mejorado.

### Trabajo a futuro

Desarrollar un nuevo geotermómetro de gases complejo que estiman la *BHT* en función de variables más independientes, como el exceso de vapor y los coeficientes de distribución de los gases con las fases líquido/vapor, entre otros.

La idea general para dicho desarrollo se presenta a continuación:

$$\begin{array}{l} \text{Geotermómetro simple} \\ BHT = f(x_i) \end{array} \quad \text{Ec. (6.1)}$$

$$\begin{array}{l} \text{Geotermómetro complejo} \\ BHT = f(x_i, k_i, \beta_i, y) \end{array} \quad \text{Ec. (6.2)}$$

donde  $x_i$  es composición geoquímica de los gases medidos (p. ej. CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, H<sub>2</sub>);  $k_i$  es la constante de equilibrio de la reacción;  $\beta_i$  son los coeficientes de distribución;  $y$  es la fracción de vapor.

**La constante de equilibrio** ( $k$ ) de una reacción química es el valor de su cociente de reacción en el equilibrio químico. Cuando tiene lugar una reacción química reversible se observa que llegado un determinado momento las cantidades netas de reactivos y productos se mantienen invariables. Este hecho no significa que la reacción se detenga, sino que la velocidad a la que se forman los productos se iguala a la velocidad a la que se regeneran los reactivos. Es decir, nuestra reacción está en equilibrio.

De manera general, podemos representar una reacción química en equilibrio así:



En la reacción de izquierda a derecha los reactivos A y B reaccionan para formar los productos C y D. Pero como es una reacción reversible podríamos interpretar que los productos también reaccionan para formar los reactivos (sentido inverso). Las velocidades a las que tienen lugar ambos procesos se pueden representar mediante las siguientes ecuaciones de velocidad:

$$v_1 = k_1[A]^a[B]^b \quad \text{Ec. (6.4)}$$

$$v_2 = k_2[C]^c[D]^d \quad \text{Ec. (6.5)}$$

donde los superíndices  $a, b, c, d$  son el número de moléculas (coeficientes estequiométricos) de las especies químicas representadas por las letras mayúsculas.

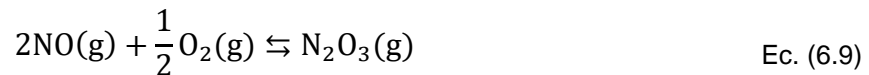
En el equilibrio las velocidades de las reacciones directa e inversa son iguales, por lo tanto:

$$v_1 = v_2 \rightarrow k_1[A]^a[B]^b = k_2[C]^c[D]^d \quad \text{Ec. (6.6)}$$

$$\frac{k_1}{k_2} = \frac{[C]^c[D]^d}{[A]^a[B]^b} \quad \text{Ec. (6.7)}$$

$$k = \frac{[C]^c[D]^d}{[A]^a[B]^b} \quad \text{Ec. (6.8)}$$

Por ejemplo, aplicando la Ec. (6.8) a la siguiente reacción tenemos:



$$k = \frac{[\text{N}_2\text{O}_3]}{[\text{NO}]^2[\text{O}_2]^{1/2}} \quad \text{Ec. (10)}$$

**Los coeficientes de distribución** ( $\beta$ ) indican la cantidad de gas disuelto en un líquido. La Ley de Henry propuesta en 1803 por William Henry enuncia que, “*a una temperatura constante, su concentración en el disolvente es directamente proporcional a su presión*”. De acuerdo a Mazor & Truesdell (1984), los coeficientes



de distribución pueden ser calculados por la Ec. (6.11):

$$\beta = \frac{RT}{k_h V} \quad \text{Ec. (6.11)}$$

$$k_h = \frac{R}{P} \quad \text{Ec. (6.12)}$$

donde  $R$  es la concentración del gas;  $T$  es la temperatura absoluta;  $k_h$  es la constante de Henry;  $V$  es el volumen específico del vapor y  $P$  es la presión del gas. La fracción de vapor ( $y$ ) puede ser calculada por si se conoce la temperatura o presión de separación y la entalpia de la mezcla mediante la Ec. (6.13):

$$y = \frac{H_O - H_L}{H_V - H_L} \quad \text{Ec. (6.13)}$$

donde  $H_O$  es la entalpia de la mezcla;  $H_L$  y  $H_V$  son la entalpia del líquido y del vapor, respectivamente. Ambas entalpias se determinan usando las tablas de vapor.

## Referencias

- Almasi, A., Jafarirad, A., Afzal, P., Rahimi, M. (2015). Prospecting of gold mineralization in Saqez area (NW Iran) using geochemical, geophysical and geological studies based on multifractal modelling and principal component analysis. *Arabian Journal of Geosciences*, vol. 8, no. 8, 5935-5947.
- Arnórsson, S., & Gunnlaugsson, E. (1985). New gas geothermometers for geothermal exploration—calibration and application. *Geochimica et Cosmochimica Acta*, vol. 49, no. 6, 1307-1325.
- Arnórsson, S. (1987). Gas chemistry of the Krísuvík geothermal field, Iceland, with special reference to evaluation of steam condensation in upflow zones. *Jökull*, vol. 37, 31-48.
- Arnórsson, S., Fridriksson, T., & Gunnarsson, I. (1998). Gas chemistry of the Krafla geothermal field, Iceland. In *Intl Symp Water-Rock Interaction, Auckland, New Zealand*, vol. 613, p. 616.
- Bayram, A. F. (2001). Application of an artificial neural network model to a Na–K geothermometer. *Journal of Volcanology and Geothermal Research*, vol. 112, no. 1, 75-81.
- Bertani, R. (2016). Geothermal power generation in the world 2010–2014 update report. *Geothermics*, vol. 60, 31-43.
- Blaise, T., Barbarand, J., Kars, M., Ploquin, F., Aubourg, C., Brigaud, B., ... & Janots, D. (2014). Reconstruction of low temperature (< 100 C) burial in sedimentary basins: a comparison of geothermometer in the intracontinental Paris Basin. *Marine and Petroleum Geology*, vol. 53, 71-87.
- Blamey, N. J. (2006). H<sub>2</sub>S concentrations in geothermal and hydrothermal fluids—a new gas geothermometer. In *Proceedings, Thirty-first Workshop of Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA*, p. 403-407.
- Burba, G. (2013). Eddy covariance method for scientific, industrial, agricultural and regulatory applications: A field book on measuring ecosystem gas exchange and areal emission rates. *LI-Cor Biosciences*, p. 349.
- Can, I. (2002). A new improved Na/K geothermometer by artificial neural networks. *Geothermics*, vol. 31, no. 6, 751-760.
- Chamorro-Camazón, C. (2009). Energía eléctrica a partir de recursos geotérmicos: Estado actual y perspectivas a nivel mundial. *DYNA-Ingeniería e Industria*, vol. 84, no. 1, 44-51.
- Chiodini, G., Cioni, R., Guidi, M., Raco, B., & Marini, L. (1998). Soil CO<sub>2</sub> flux measurements in volcanic and geothermal areas. *Applied Geochemistry*, vol. 13, no. 5, 543-552.
- Chiodini, G., Frondini, F., Cardellini, C., Granieri, D., Marini, L., & Ventura, G. (2001). CO<sub>2</sub> degassing and energy release at Solfatara volcano, Campi Flegrei, Italy. *Journal of Geophysical Research: Solid Earth*, vol. 106, no. B8, 16213-16221.
- Chiodini, G., Caliro, S., Cardellini, C., Avino, R., Granieri, D., & Schmidt, A. (2008). Carbon isotopic composition of soil CO<sub>2</sub> efflux, a powerful method to discriminate different sources feeding soil

- CO<sub>2</sub> degassing in volcanic-hydrothermal areas. *Earth and Planetary Science Letters*, vol. 274, no. 3, 372-379.
- D'Amore, F., & Panichi, C. (1980). Evaluation of deep temperatures of hydrothermal systems by a new gas geothermometer. *Geochimica et Cosmochimica Acta*, vol. 44, no. 3, 549-556.
- Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & dos Reis Alves, S. F. (2017). Introduction. In: *Artificial Neural Networks*. Springer, Cham. [https://doi.org/10.1007/978-3-319-43162-8\\_1](https://doi.org/10.1007/978-3-319-43162-8_1)
- De Lima, P.C. (2005). Paralelização Do Algoritmo De Backpropagation Para Clusters Beowulf. Tese para obtenção do grau Bacharel. Instituto Educacional San Agustín, Montes Claros, Brasil, p. 92.
- Díaz-González, L., Hidalgo-Dávila, C. A., Santoyo, E., Hermosillo-Valadez, J. (2013). Evaluación de técnicas de entrenamiento de redes neuronales para estudios geotermométricos de sistemas geotérmicos. *Revista mexicana de ingeniería química*, vol. 12, no. 1, 105-120.
- Dickson M.H., Fanelli M., (2004). What is geothermal energy?, Consulta en línea, disponible en: <http://www.geothermal-energy.org/geo/geoenergy.php>.
- Dickson, Mary H., & Mario Fanelli. (2013). *Geothermal energy: utilization and technology*. Routledge.
- Dobson, P. F. (2016). A review of exploration methods for discovering hidden geothermal systems. *GRC Trans.*, 695-706.
- Ebrahimi, D., Nouraliee, J., Dashti, A. (2019). Assessing the geothermal potential of the Shahin Dezh Region, based on the geological, geochemical and geophysical evidence. *Journal of African Earth Sciences*, vol. 152, 84-94.
- Espinoza-Ojeda, O. M., Santoyo, E., & Andaverde, J. (2011). A new look at the statistical assessment of approximate and rigorous methods for the estimation of stabilized formation temperatures in geothermal and petroleum wells. *Journal of Geophysics and Engineering*, vol. 8, no. 2, 233-258.
- Espinoza-Ojeda, O. M., & Santoyo, E. (2016). A new empirical method based on log-transformation regressions for the estimation of static formation temperatures of geothermal, petroleum and permafrost boreholes. *Journal of Geophysics and Engineering*, vol. 13, no. 4, 559-596.
- Fridriksson, T., Padrón, E., Óskarsson, F., & Pérez, N. M. (2016). Application of diffuse gas flux measurements and soil gas analysis to geothermal exploration and environmental monitoring: Example from the Reykjanes geothermal field, SW Iceland. *Renewable energy*, vol. 86, 1295-1307.
- García-Gil, A., Vázquez-Suñe, E., Alcaraz, M. M., Juan, A. S., Sánchez-Navarro, J. Á., Montlleó, M., Rodríguez G., Lao, J. (2015). GIS-supported mapping of low-temperature geothermal potential taking groundwater flow into account. *Renewable energy*, vol. 77, 268-278.
- García-López, C. G., Pandarinath, K., & Santoyo, E. (2014). Solute and gas geothermometry of geothermal wells: a geochemometrics study for evaluating the effectiveness of geothermometers to predict deep reservoir temperatures. *International Geology Review*, vol. 56, no. 16, 2015-2049.
- García-López, C. G. (2016). Evaluación y desarrollo de geotermómetros de gases para la predicción de temperaturas de sistemas geotérmicos. Tesis para obtener el grado de Doctora en Ingeniería.

- Instituto de Energías Renovables, UNAM, Temixco, Morelos, México, p. 160.
- Garson, G.D. (1991). Interpreting neural-network connection weights. *Artificial Intelligence Expert*, vol. 6, no. 7, 47–51.
- Giammanco, S., Sims, K. W. W., & Neri, M. (2007). Measurements of  $^{220}\text{Rn}$  and  $^{222}\text{Rn}$  and  $\text{CO}_2$  emissions in soil and fumarole gases on Mt. Etna volcano (Italy): Implications for gas transport and shallow ground fracture. *Geochemistry, Geophysics, Geosystems*, vol. 8, no. 10, 1-14.
- Giggenbach, W. F. (1991). Chemical techniques in geothermal exploration. Application of geochemistry in geothermal reservoir development. *Series of Technical Guides on the Use of Geothermal Energy*, UNITAR/UNDP Centre Press, Rome, 119-144.
- 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*, vol. 21, no. 1-2, 121-140.
- Gomez, M. L. (2016). Estimación de parámetros de transporte y mecanismos que gobiernan su respuesta frente a la contaminación: caso de estudio de un acuífero poroso clástico. *Boletín de la sociedad geológica Mexicana*, vol. 68, no. 2, 339-356.
- Guerrero-Martínez, F. J., Prol-Ledesma, R. M., Carrillo-De La Cruz, J. L., Rodríguez-Díaz, A. A., & González-Romo, I. A. (2020). A three-dimensional temperature model of the Aocolco caldera complex, Puebla, Mexico, from the Curie isotherm as a boundary condition. *Geothermics*, vol. 86, 1-11.
- Gutiérrez-Negrín, L. C., Canchola Félix, I., Romo-Jones, J. M., & Quijano-León, J. L. (2020). Geothermal energy in Mexico: update and perspectives. In *Proceedings, Proceedings World Geothermal Congress*, Reykjavik, Iceland, p. 13.
- Harvey, M. C., Rowland, J. V., Chiodini, G., Rissmann, C. F., Bloomberg, S., Fridriksson, T., & Óladóttir, A. A. (2017).  $\text{CO}_2$  flux geothermometer for geothermal exploration. *Geochimica et Cosmochimica Acta*, vol. 213, 1-16.
- Hiriart, L. B. G., Gutiérrez-Negrín, L. C., Quijano-León, H. L., Ornelas-Celis, A., Espíndola, S., & Hernández-Carrillo, I. (2011). Evaluación de la energía geotérmica en México. Informe para el banco Interamericano de Desarrollo y la Comisión Reguladora de Energía, p. 167.
- Honglin, H., Guirui, Y., Zhang, Leiming., S., X., & Wen, S. (2006). Simulating  $\text{CO}_2$  flux of three different ecosystems in ChinaFLUX based on artificial neural networks. *Science in China Series D: Earth Sciences*, vol. 49, 252-261.
- Huttrer, G. W. (2020). Geothermal power generation in the world 2015-2020 update report. In *World Geothermal Congress; International Geothermal Association: Reykjavik, Iceland*. 17 p.
- Idroes, R., Yusuf, M., Saiful, S., Alatas, M., Subhan, S., Lala, A., ... & Mahlia, T. M. I. (2019). Geochemistry Exploration and Geothermometry Application in the North Zone of Seulawah Agam, Aceh Besar District, Indonesia. *Energies*, vol. 12, no. 23, 1-17.
- Iglesias, E., Arellano, V., & Torres, R.J. (2005). Estimación del recurso y prospectiva tecnológica de la geotermia en México. Gerencia de Geotermia, IIE. Proyecto IIE No. 3753, INFORME

- Incropera, F. P., & DeWitt, D. P. (1999). *Introducción. Fundamentos de transferencia de calor* 4th. Edición. Pearson Educación. pp. 2-29.
- Iorio, M., Carotenuto, A., Corniello, A., Di Fraia, S., Massarotti, N., Mauro, A., ... & Vanoli, L. (2020). Low enthalpy geothermal systems in structural controlled areas: A sustainability analysis of geothermal resource for heating plant (the Mondragone case in Southern Appennines, Italy). *Energies*, vol. 13, no. 5, 1237.
- Isaka, B. A., Ranjith, P. G., & Rathnaweera, T. D. (2019). The use of super-critical carbon dioxide as the working fluid in enhanced geothermal systems (EGSs): a review study. *Sustainable Energy Technologies and Assessments*, vol. 36, no. 100547, 1-18.
- Jackson, P. C. (2019). *Introduction to artificial intelligence*. Third Edition, Courier Dover Publications, Mineola, New York. pp. 454
- Kaya, E., Zarrouk, S. J., & O'Sullivan, M. J. (2011). Reinjection in geothermal fields: a review of worldwide experience. *Renewable and sustainable energy reviews*, vol. 15, no. 1, 47-68.
- Klusman, R. W., Moore, J. N., & LeRoy, M. P. (2000). Potential for surface gas flux measurements in exploration and surface evaluation of geothermal resources. *Geothermics*, vol. 29, no. 6, 637-670.
- Koga, A., Kita, I., Hikino, E., Nitta, K., & Taguchi, S. (1995). New gas geothermometers using CO<sub>2</sub>/H<sub>2</sub> and CH<sub>4</sub>/H<sub>2</sub> ratios. *Journal of the Geothermal Research Society of Japan*, vol. 17, no. 3, 201-211.
- Kumari, W. G. P., & Ranjith, P. G. (2019). Sustainable development of enhanced geothermal systems based on geotechnical research—A review. *Earth-Science Reviews*, vol. 199, 1-25.
- Ledesma, R. M. P. (1993). *El calor de la Tierra*. SEP, Subsecretaría de Educación Superior e Investigación Científica. pp. 1-41.
- Lepeltier, C. (1969). A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.*, vol. 64, no. 5, 538-550.
- Lewicki, J. L., & Oldenburg, C. M. (2004). *Strategies for detecting hidden geothermal systems by near-surface gas monitoring*. Earth Sciences Division Ernest Orlando Lawrence Berkeley National Laboratory Berkeley, CA, USA, p. 77.
- Llopis, G., & Rodrigo, V. (2008). *Guía de la energía geotérmica*. Dirección general de la industria energía, y minas, consejería de economía y consumo, pp. 13-184.
- López, R. F., & Fernández, J. M. F. (2008). *Redes Neuronales Artificiales. Las redes neuronales artificiales fundamentos teóricos y aplicaciones prácticas*. Netbiblo, pp. 16-41.
- Lowenstern, J. B., Bergfeld, D., Evans, W. C., & Hunt, A. G. (2015). Origins of geothermal gases at Yellowstone. *Journal of Volcanology and Geothermal Research*, vol. 302, 87-101.
- Lund, J. W., & Toth, A. N. (2020). Direct Utilization of geothermal energy 2020 Worldwide review. *Geothermics*, 101915.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, vol. 5, no. 4, 115-133.
- Martínez, E., Fuentes, J. P., & Acevedo, E. (2008). *Carbono orgánico y propiedades del suelo*.

- Revista de la ciencia del suelo y nutrición vegetal, vol. 8, no.1, 68-96.
- Marzolf, N. C. (2014). Emprendimiento de la energía geotérmica en Colombia. Banco Interamericano de Desarrollo, pp. 1-86.
- Mazor, E., & Truesdell, A. H. (1984). Dynamics of a geothermal field traced by noble gases: Cerro Prieto, Mexico. *Geothermics*, vol. 13, no. 1-2, 91-102.
- Mazur, M. A Step by Step Backpropagation Example. <https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/> (Consulta mayo 2018).
- Melesse, A. M., & Hanley, R. S. (2005). Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecological Modelling*, vol. 189, no. 3-4, 305-314.
- Méndez, J. T. P., & Morales, R. M. (2008). *Inteligencia artificial. Métodos, técnicas y aplicaciones*. Primera Edición. McGraw-Hill, Murcia, España. pp. 1022.
- Nehring, N. L., & D'Amore, F. (1984). Gas chemistry and thermometry of the Cerro Prieto, Mexico, geothermal field. *Geothermics*, vol. 13, no. 1-2, 75-89.
- Nicholson, K. (2012). *Geothermal fluids: chemistry and exploration techniques*. Springer Science & Business Media, p. 263.
- Noorollahi, Y., Itoi, R., Fujii, H., Tanaka, T. (2008). GIS integration model for geothermal exploration and well siting. *Geothermics*, vol. 37, no. 2, 107-131.
- Olasolo, P., Juárez, M. C., Morales, M. P., Liarte, I. A. (2016). Enhanced geothermal systems (EGS): A review. *Renewable and Sustainable Energy Reviews*, vol. 56, 133-144.
- Oporto, S. (2006). *Redes Neuronales Artificiales – Introducción*. Fuente original: <http://www.wiphala.net/oporto>; actualmente disponible: <https://slideplayer.es/slide/92349/>
- Pang, Z. (2001). Isotope and chemical geothermometry and its applications. *Science in China Series E: Technological Sciences*, vol. 44, no. 1, 16-20.
- Parks, M. M., Caliro, S., Chiodini, G., Pyle, D. M., Mather, T. A., Berlo, K., ... & Raptakis, C. (2013). Distinguishing contributions to diffuse CO<sub>2</sub> emissions in volcanic areas from magmatic degassing and thermal decarbonation using soil gas 222 Rn–d 13 C systematics: Application to Santorini volcano, Greece. *Earth and Planetary Science Letters*, vol. 377, 180-190.
- Peiffer, L., Bernard-Romero, R., Mazot, A., Taran, Y. A., Guevara, M., & Santoyo, E. (2014). Fluid geochemistry and soil gas fluxes (CO<sub>2</sub>–CH<sub>4</sub>–H<sub>2</sub>S) at a promissory Hot Dry Rock Geothermal System: The Acoculco caldera, Mexico. *Journal of Volcanology and Geothermal Research*, vol. 284, 122-137.
- Ping, Z., & Ármannsson, H. (1996). Gas geothermometry in selected Icelandic geothermal fields with comparative examples from Kenya. *Geothermics*, vol. 25, no. 3, 307-347.
- Ponce, J., Soto, A., & Quezada, F. (2014). *Inteligencia artificial* (Doctoral dissertation, Open Textbook Library).
- Portugal, E., P Verma, M., M Barragán, R., & Mañón, A. (1994). Geoquímica isotópica de 13 C, D y 18 O de fluidos del sistema geotérmico Los Hornos, Puebla (México). *Geofísica Internacional*, vol. 33, no. 4, 607-618.

- Powell, T. (2000). A review of exploration gas geothermometry. In Proceedings, 25th Workshop on Geothermal Reservoir Engineering, p. 9.
- Powell, T., & Cumming, W. (2010). Spreadsheets for geothermal water and gas geochemistry. In Proceedings, 35th Workshop on Geothermal Reservoir Engineering, p. 10.
- Pruess, K., & Spycher, N. (2009). Enhanced geothermal systems (EGS) with CO<sub>2</sub> as heat transmission fluid--A scheme for combining recovery of renewable energy with geologic storage of CO<sub>2</sub> (No. LBNL-1949E). Lawrence Berkeley National, CA, USA, p. 6.
- Pulido, C. L., Armenta, M. F., & Silva, G. R. (2011). Caracterización de un yacimiento de roca seca caliente en la zona geotérmica de Aocolco, Pue. *Geotermia*, vol. 24, no. 1, 59-69.
- Pulido, J. R. G., Michel, E. M. R., & Díaz, R. A. (2018). Inteligencia artificial y aprendizaje máquina: Aplicaciones y tendencias. *Tecnologías disruptivas de información*. Dirección General de Publicaciones de la Universidad de Colima, pp. 177.
- Raine, R., & Reay, D. M. (2019). A review of geothermal reservoir properties of Triassic, Permian and Carboniferous sandstones in Northern Ireland, Geological Survey of Northern Ireland, Internal Report 19/01, 62 p.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, vol. 323, no. 6088, 533-536.
- Russell, S., & Norvig, P. (2002). *Artificial intelligence: a modern approach*. Tercera Edición. Prentice Hall, Saddle River, New Jersey, pp. 1132.
- Russell, M. J., Hall, A. J., Boyce, A. J., & Fallick, A. E. (2005). 100th Anniversary Special Paper: On Hydrothermal Convection Systems and the Emergence of Life. *Economic Geology*, vol. 100, no. 3, 419-438.
- Santoyo, E., & Barragán-Reyes, R. M. (2010). Energía geotérmica. *Revista de la Academia Mexicana de Ciencias*, vol. 61, 40-51.
- Santoyo, E., & Torres-Alvarado, I. S. (2010). Escenario futuro de explotación de la energía geotérmica: hacia un desarrollo sustentable. *Revista Digital Universitaria*, vol. 11, no. 10, 1-26.
- Sánchez-Robles, A. (2010). Comportamiento de la permeabilidad en núcleos naturalmente fracturados. Tesis para obtener el título de ingeniero petrolero. Facultad de ingeniería, UNAM, 100 p.
- Sarmiento, Z. F., Sagala, B., & Siagian, H. (2019). The Sokoria Geothermal System, Flores Island, Indonesia. *Geothermics*, vol. 82, 282-295.
- Sinclair, A. J. (1974). Selection of threshold values in geochemical data using probability graphs. *Journal of Geochemical Exploration*, vol. 3, no. 2, 129-149.
- Speckman, H. N., Frank, J. M., Bradford, J. B., Miles, B. L., Massman, W. J., Parton, W. J., & Ryan, M. G. (2015). Forest ecosystem respiration estimated from eddy covariance and chamber measurements under high turbulence and substantial tree mortality from bark beetles. *Global change biology*, vol. 21, no. 2, 708-721.
- Stelling, P., Shevenell, L., Hinz, N., Coolbaugh, M., Melosh, G., & Cumming, W. (2016). Geothermal

- systems in volcanic arcs: Volcanic characteristics and surface manifestations as indicators of geothermal potential and favorability worldwide. *Journal of Volcanology and Geothermal Research*, vol. 324, 57-72.
- Supranto, S., Budianto, T., Djoko, W., & Idrus, A. (1996). Proposed empirical gas geothermometer using multidimensional approach. In *Proceedings, Twenty First Workshop on Geothermal Reservoir Engineering: Stanford, California, Stanford University, USA*, p. 195-199.
- Tabaco, F., Verma, M. P., Nieva, D., & Portugal, E. (1991). Características geoquímicas e isotópicas del carbono en el sistema geotérmico de Los Azufres, Michoacán. *Geofísica internacional*, Vol. 30, No. 3, pp. 173-181.
- Thomas, Dana Louise. (2018). *Geochemistry of Carbon Dioxide Sequestration in Basalts: Implications from Natural Analogues and Experimental Observations*. Doctoral dissertation, Stanford University, pp. 246.
- Tomaszewska, B., Pająk, L., Bundschuh, J., Bujakowski, W. (2018). Low-enthalpy geothermal energy as a source of energy and integrated freshwater production in inland areas: Technological and economic feasibility. *Desalination*, vol. 435, 35-44.
- Verma, S. P. (2002). *Optimisation of the exploration and evaluation of geothermal resources. Geothermal Energy Resources for Developing Countries: Rotherdam, Swets & Zeitlinger, BV, AA Balkena Publishers*, 195-224.
- Viggiano-Guerra, J. C., Armenta, M. F., & Silva, G. R. (2011). Evolución del sistema geotérmico de Acoculco, Pue., México: un estudio con base en estudios petrográficos del pozo EAC-2 y en otras consideraciones. *Geotermia*, vol. 24, no. 1, 14-24.
- Wallace, P. J. (2005). Volatiles in subduction zone magmas: concentrations and fluxes based on melt inclusion and volcanic gas data. *Journal of volcanology and Geothermal Research*, vol. 140, no. 1-3, 217-240.
- Wang, G. L., Zhang, W., Ma, F., Lin, W. J., Liang, J. Y., Zhu, X. (2018). Overview on hydrothermal and hot dry rock researches in China. *China Geology*, vol. 1, no. 2, 273-285.
- Williams, C., Jones, M. A., Durden, D., Luo, H., & Metzger, S. (2017). Eddy covariance: Measuring an ecosystem's breath [Video file]. Courtesy: Battelle - National Ecological Observatory Network. doi:10.13140/RG.2.2.32889.06244.
- Wetzel, T. G., Balan, C., Byrd, C. M., & Rajendran, V. P. (2006). U.S. Patent No. 7,124,584. Washington, DC: U.S. Patent and Trademark Office. pp. 6-10.
- Welles, J. M., Demetriades-Shah, T. H., & McDermitt, D. K. (2001). Considerations for measuring ground CO<sub>2</sub> effluxes with chambers. *Chemical Geology*, vol. 177, no. 1, 3-13.
- Xu, T., Pruess, K., & Apps, J. (2008). Numerical studies of fluid-rock interactions in enhanced geothermal systems (EGS) with CO<sub>2</sub> as working fluid. In *Proceedings, Thirty-Third Workshop on Geothermal Reservoir Engineering: Stanford, California, Stanford University, USA*, p. 9.
- Zhang, Q., Yu, H., Barbiero, M., Wang, B., & Gu, M. (2019). Artificial neural networks enabled by nanophotonics. *Light: Science & Applications*, vol. 8, no. 1, 1-14.



## ANEXOS

Todos los anexos están disponibles en la literatura internacional o pueden ser solicitados al correo electrónico del autor de la tesis:

[agustin.acevedo.anicasio@gmail.com](mailto:agustin.acevedo.anicasio@gmail.com)

### ANEXO 1.

Pérez-Zárate, D., Santoyo, E., Acevedo-Anicasio, A., Díaz-González, L., & García-López, C. (2019). Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. *Computers & Geosciences*, vol. 129, 49-68.

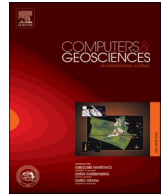
### ANEXO 2.

Acevedo-Anicasio, A., Santoyo, E., Pérez-Zárate, D., Pandarinath, K., Guevara, M., Díaz-González, L. (2021). GaS\_GeoT: A computer program for an effective use of newly improved gas geothermometers in predicting reliable geothermal reservoir temperatures. *Geotherm Energy*, vol. 9, no. 1, 1-41

### ANEXO 3.

Santoyo, E., Acevedo-Anicasio, A., Pérez-Zarate, D., & Guevara, M. (2018). Evaluation of artificial neural networks and eddy covariance measurements for modelling the CO<sub>2</sub> flux dynamics in the Acoculco geothermal caldera (Mexico). *International Journal of Environmental Science and Development*, vol. 9, no. 10.

# ANEXO 1



# Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids

D. Pérez-Zárate<sup>a</sup>, E. Santoyo<sup>b,1,\*</sup>, A. Acevedo-Anicasio<sup>c</sup>, L. Díaz-González<sup>d</sup>, C. García-López<sup>e</sup>

<sup>a</sup> CONACYT-Instituto de Geofísica, Universidad Nacional Autónoma de México, Circuito interior s/n, Coyoacán, 04510, Ciudad de México, Mexico

<sup>b</sup> Depto. de Física y Matemáticas, Universidad Iberoamericana, Ciudad de México, 01219, México

<sup>c</sup> Doctorado en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos; Av. Universidad 1001, Col. Chamilpa, 62209, Morelos, Mexico

<sup>d</sup> Centro de Investigación en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos; Av. Universidad 1001, Chamilpa, 62209, Morelos, Mexico

<sup>e</sup> Posgrado en Ingeniería (Energía), Instituto de Energías Renovables, Universidad Nacional Autónoma de México, Priv. Xochicalco s/n, Temixco, 62580, Morelos, Mexico

## ARTICLE INFO

### Keywords:

Geothermal energy  
Geothermal prospection  
Gas geothermometers  
Fluid geochemistry  
Artificial intelligence  
Machine learning

## ABSTRACT

Three-layer artificial neural networks were used for the multivariate analysis of the gas-phase composition of fluids, and the prediction of geothermal reservoir temperatures. The major gas-phase composition of geothermal fluids (CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub>) was defined as input variables whereas the measured bottomhole temperatures were used as output. Multivariate statistical analysis and log-ratio transformations were used for the normalization of input variables. These data sets were randomly divided into training (80%), validation (10%) and testing (10%). Matlab script algorithms based on a learning process of artificial neural networks (ANN) were programmed. The ANN architectures were determined by using up to five input neurons with thirteen different combinations of the input normalized variables for the input layer, a variable number of neurons for the hidden layer ranging from 1 to 35, and one neuron for the output layer. The Levenberg-Marquardt algorithm, the hyperbolic tangent sigmoid, and linear transfer functions were used for the training of the neural networks with tuned learning rates after using some initialization values (ranging from 0.01 to 0.00001). A sensitivity analysis to evaluate the relative importance of input variables on the output was also performed. Six ANN architectures were selected as the best from a comprehensive geochemometric analysis based on a statistical comparison and linear regressions between simulated and measured data. From an additional external validation (using a different geochemical database containing 13 samples from the Olkaria, Kenya geothermal field), the most efficient ANN made possible to predict geothermal reservoir temperatures with an acceptable accuracy (with mean error differences between simulated and measured data ranging from 2 to 11%). The best neural network was given by the ANN-33 architecture [3-8-1] which was characterized by three variables in the input layer [ $\ln(H_2S/CO_2)$ ,  $\ln(CH_4/CO_2)$ , and  $\ln(H_2/CO_2)$ ], 8 neurons for the hidden layer and one neuron for the output layer. These results support, for the first time, the use of ANN as a gas geothermometry tool to predict geothermal reservoir temperatures.

## 1. Introduction

The knowledge of reservoir temperatures in geothermal systems is still recognized as a crucial geochemical task for the reliable evaluation of energy reserves (Neupane et al., 2016; Guo et al., 2017). Solute, gas and isotopic geothermometers are low-cost geochemical tools used for

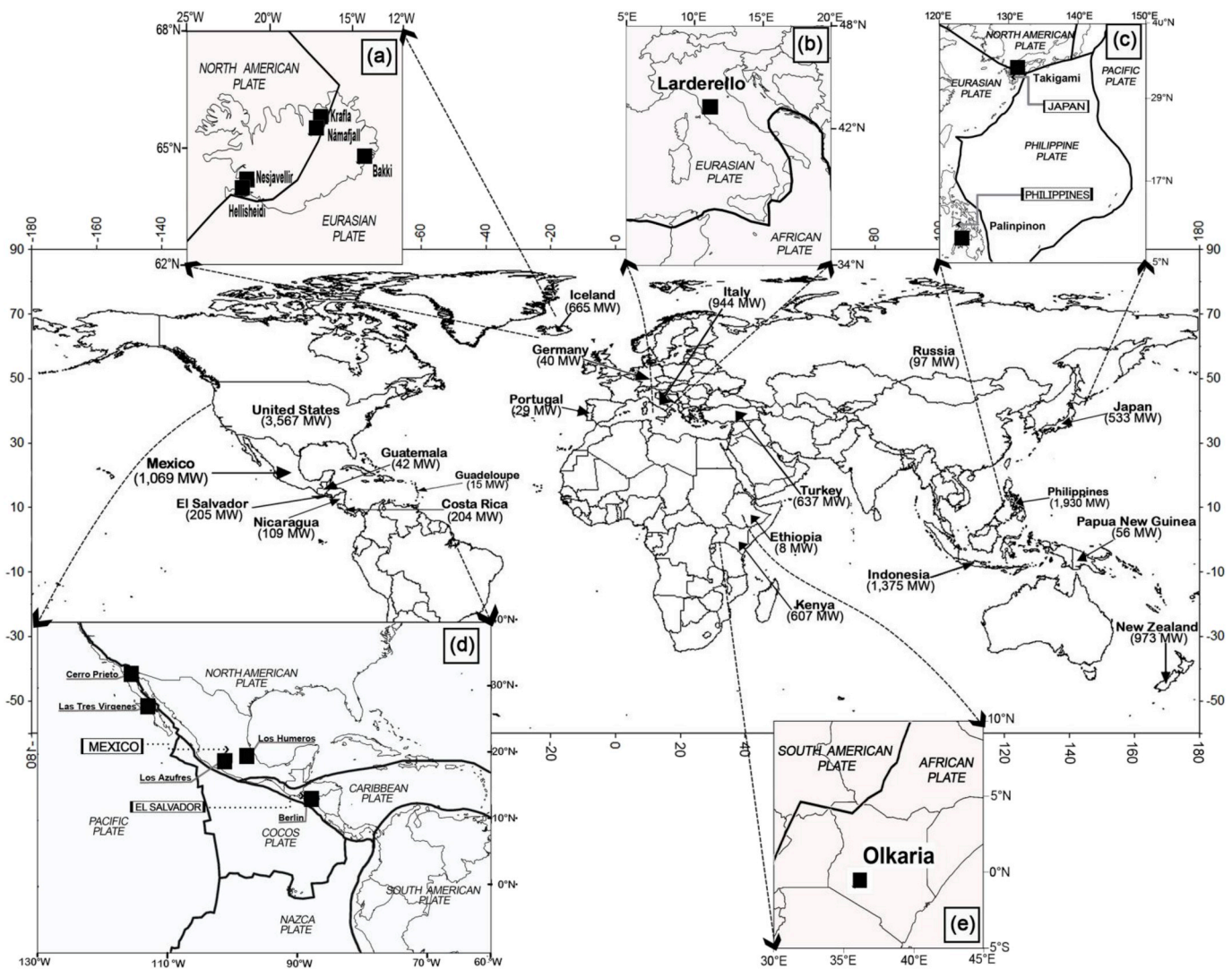
the prediction of reservoir temperatures in geothermal systems (Verma and Santoyo, 1997; Arnórsson, 2000; Díaz-González et al., 2008; Yan-guang et al., 2017).

In geothermal exploration, the chemistry of geothermal gases has been broadly studied for estimating reservoir temperatures (e.g., Stefánsson et al., 2017). A geothermal system above cooling magma may

\* Corresponding author.

E-mail addresses: [depez@igeofisica.unam.mx](mailto:depez@igeofisica.unam.mx) (D. Pérez-Zárate), [esg@ier.unam.mx](mailto:esg@ier.unam.mx) (E. Santoyo), [agustin.acevedoa@uaem.edu.mx](mailto:agustin.acevedoa@uaem.edu.mx) (A. Acevedo-Anicasio), [ldg@uaem.mx](mailto:ldg@uaem.mx) (L. Díaz-González), [cggal@ier.unam.mx](mailto:cggal@ier.unam.mx) (C. García-López).

<sup>1</sup> Instituto de Energías Renovables, Universidad Nacional Autónoma de México, Priv. Xochicalco s/n, Temixco, 62580, Morelos, México.



**Fig. 1.** Schematic map showing the current worldwide installed capacity of geothermal electricity, and the location of the geothermal fields considered for the creation of the World Geochemical Database (WGDB).

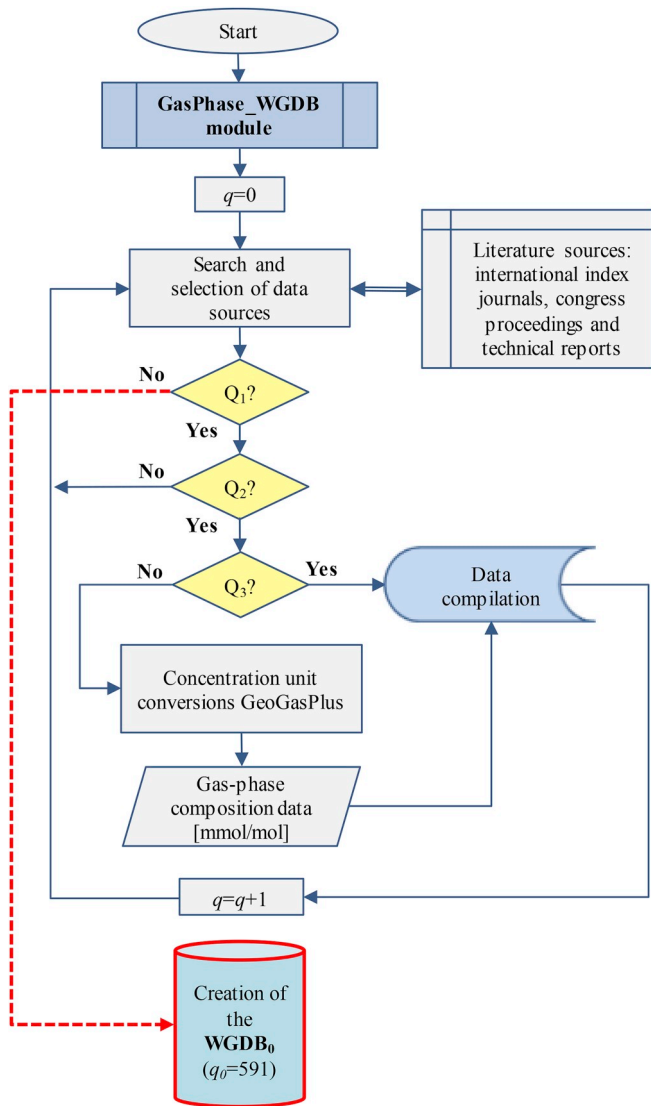
behave either as an open thermodynamic system due to the escape of magma gases or as a physicochemical trap owing to gas-gas and gas-mineral reactions that may consume or produce these gases (Arnórrsson et al., 2010). The equilibria attainment among gas species constitutes the basis for gas geothermometry, which has been preferentially used for the prediction of reservoir temperatures in vapor-dominated geothermal systems where fumaroles, steam or gas bubbles are the main surface manifestations (Arnórrsson and Gunnlaugsson, 1985).

Gas geothermometers have been developed based upon the analysis of gas-phase compositions, and the assumption of a thermodynamic equilibrium among gas-gas and gas-mineral reactions (Giggenbach, 1980; Arnórrsson and D'Amore, 2000). These reactions are thermodynamically expressed by equations based upon the free energy change of gas species, which may be correlated with temperature-dependent equilibrium constants (D'Amore and Panichi, 1980; D'Amore et al., 1983; Giggenbach, 1991; Blamey, 2006). Other complex gas geothermometers consider independent reservoir parameters such as the steam excess, the distribution coefficients of gases between liquid and steam, the gas/steam ratio, and the vapor fraction (*VF*) which may also affect the equilibria of gas reactions (Nehring and D'Amore, 1984; Bertrami et al., 1985; Giggenbach, 1988; Barragán et al., 2016). Gas geothermometers currently available in the literature include: (1)

Simple empirical equations that use geochemical databases containing gas-phase compositions of fluids from hot-water and wet-steam geothermal wells, and bottomhole temperature measurements (Arnórrsson, 1985); (2) Complex numerical methods for drawing temperature and *VF* grids using the thermodynamic analysis of chemical reactions among gas-gas and gas-mineral assemblages (D'Amore and Nuti, 1977; Bertrami et al., 1985; Stefánsson and Arnórrsson, 2002; Barragán et al., 2016); and (3) Geothermometric equations from water-rock-gas interaction experiments carried out at laboratory conditions (Reed and Spycher, 1984; Pang, 2001).

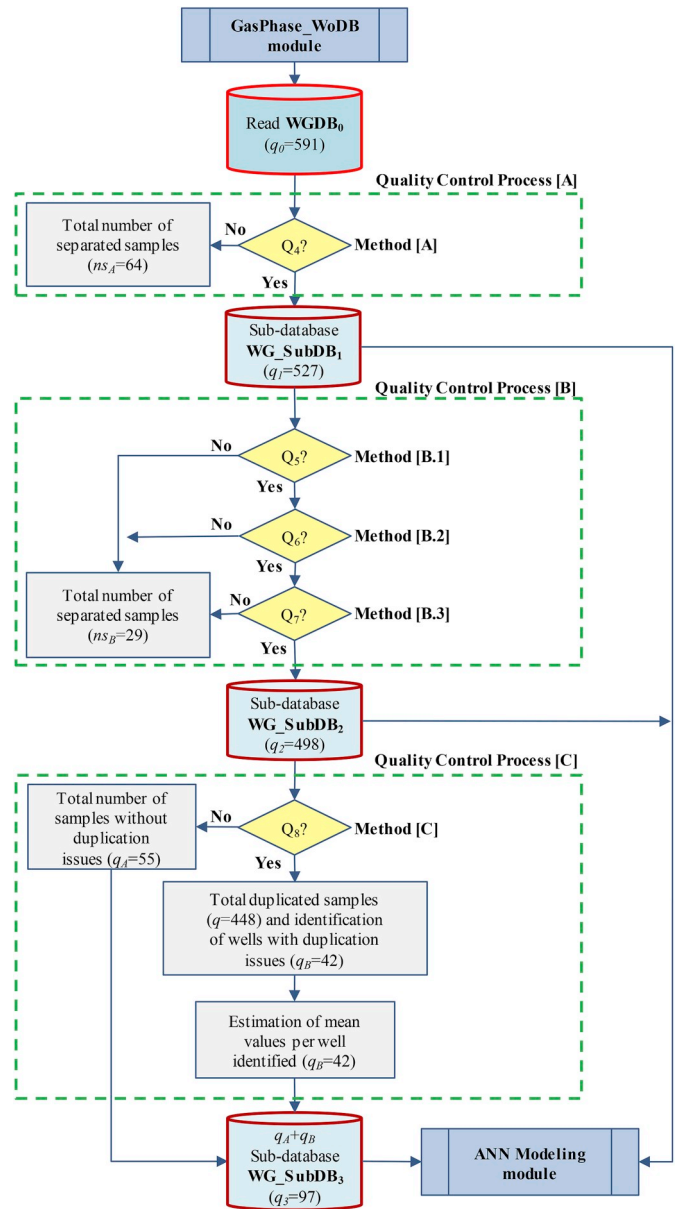
Although the advances achieved in this knowledge area, it is recognized that most of the reservoir temperatures predicted by gas geothermometers still exhibit significant discrepancies when these estimates are compared with bottomhole temperature (*BHT<sub>m</sub>*) measurements (Arnórrsson and Gunnlaugsson, 1985; Giggenbach, 1988; Powell and Cumming, 2010; García-López et al., 2014). Such discrepancies are mainly caused by uncertainty sources such as:

- (i) the use of a limited number of actual temperature measurements for the calibration of gas geothermometers (García-López et al., 2014);



**Fig. 2.** Schematic flow diagram showing the GasPhase\_WGDB module used for the creation of the world geochemical database (WGDB<sub>0</sub>). Q<sub>i</sub> represent the following conditional statements: Q<sub>1</sub>: compilation of a new record; Q<sub>2</sub>: availability of gas-phase concentration and BHT(°C) data; and Q<sub>3</sub>: standard concentration unit [mmol/mol].

- (ii) the use of small data sets of gas-phase fluid compositions and  $BHT_m$  measurements for the derivation of gas geothermometers (Koga et al., 1995; Supranto et al., 1996);
- (iii) the lack of equilibrium into the reservoir which affect the gas-phase composition of deep fluids, as well as other geochemical processes, such as the fluid-mineral equilibria, the gas reactions, and the distribution coefficients (Giggenbach, 1980; Pang and Reed, 1998);
- (iv) the effect of secondary geochemical processes such as steam loss, the feeding of secondary fluids, and mixing with shallow colder fluids, which may affect the original composition of deep fluids (Giggenbach, 1987);
- (v) the gas sampling and analytical errors which may affect the representativeness of the gas samples (Kacandes and Grandstaff, 1989; Arnórrsson et al., 2006); and
- (vi) the erroneous use of non-appropriated regressions for the derivation of geothermometer equations (Verma and Santoyo, 1997; Díaz-González et al., 2008).



**Fig. 3.** Schematic flow diagram of the GasPhase\_WorkDB module used for the creation of the working sub-databases (WG\_SubDB1, WG\_SubDB2, and WG\_SubDB3). Q<sub>i</sub> represent the following conditional statements: Q<sub>4</sub>: full availability of the gas compositions of CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub> per sample; Q<sub>5</sub>: total sum of gas concentrations = 1000 ± 50 mmol/mol; Q<sub>6</sub>: temperature interval: 170 °C ≤  $BHT_m$  ≤ 365 °C; Q<sub>7</sub>: the CO<sub>2</sub> concentration ≥ 70%wt; and Q<sub>8</sub>: exist duplicate  $BHT_m$  data?.

To face out these issues, innovative geochemical methods in geothermal exploration are still needed for the development of new improved gas geothermometers. Bearing in mind these needs, artificial neural networks are explored, up to our knowledge for the first time, for the prediction of reservoir temperatures using the gas-phase compositions of geothermal fluids. Discrepancy types (i, ii, v, and vi) may be either minimized or controlled by an efficient multivariate analysis through the use of artificial neural network techniques. Artificial neural networks (ANN) have been used as computational tools in Earth Sciences for modeling and solving a wide variety of multivariate problems in geophysics, geochemistry and geology (Van der Baan and Jutten, 2000; Dawson and Wilby, 2001; Poulton, 2001, 2002; Hagan et al., 2014; Trépanier et al., 2016). ANN are parallel-distributed information processing techniques that may found relationships among input and

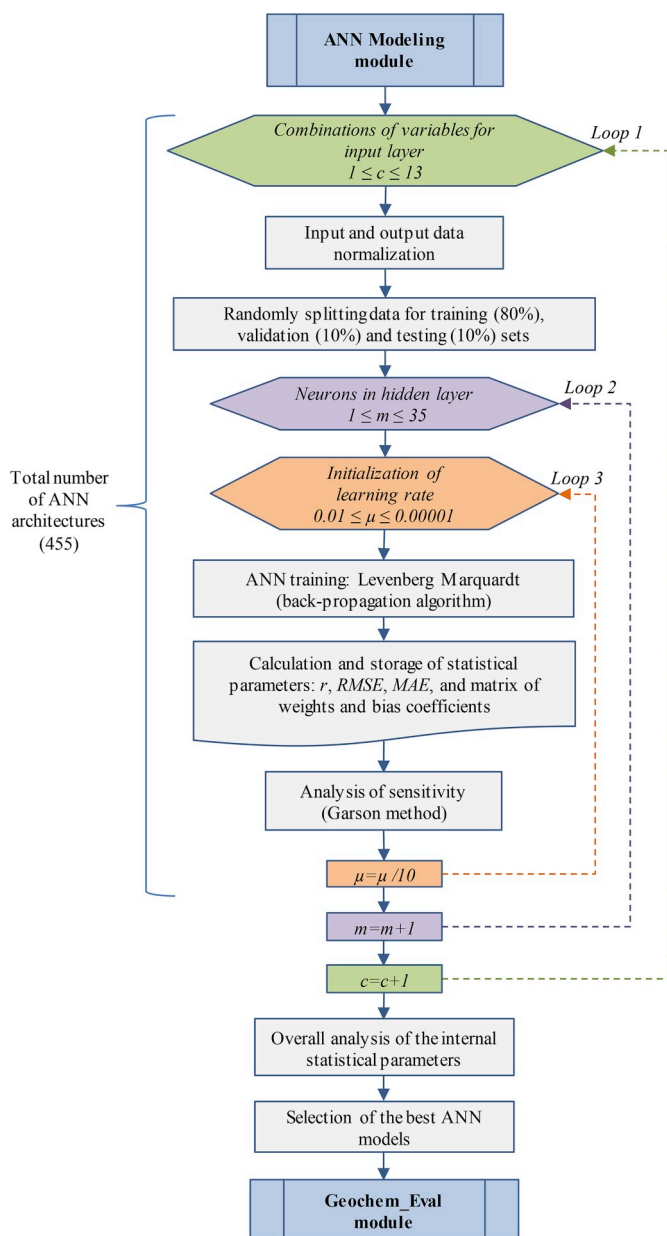


Fig. 4. Schematic flow diagram showing the ANN\_Modeling module developed to design and apply the ANN architectures for the prediction of deep reservoir temperatures.

output variables of a complex system through a network with highly interconnected neurons (Hornik et al., 1989).

In some earlier geothermal applications of the ANN, new versions of solute geothermometers have been derived from the multivariate analysis of geochemical databases of the liquid-phase composition of fluids and bottomhole temperatures (e.g., Bayram, 2001; Can, 2002; Serpen et al., 2010; Díaz-González et al., 2013). Some other ANN applications have been reported for modelling dynamic fluid flow and heat transport in geothermal wells (e.g., Bassam et al., 2010, 2015; Álvarez del Castillo et al., 2012; Porkhial et al., 2015). Although these novel applications, complex multivariate issues still exist in other geothermal areas that may be addressed by using ANN techniques.

In this work, a new application of the ANN techniques has been successfully carried out to correlate the gas-phase composition of geothermal fluids with measured bottomhole temperatures for the determination of deep reservoir temperatures. A geochemical database containing gas-phase fluid compositions and temperature measurements

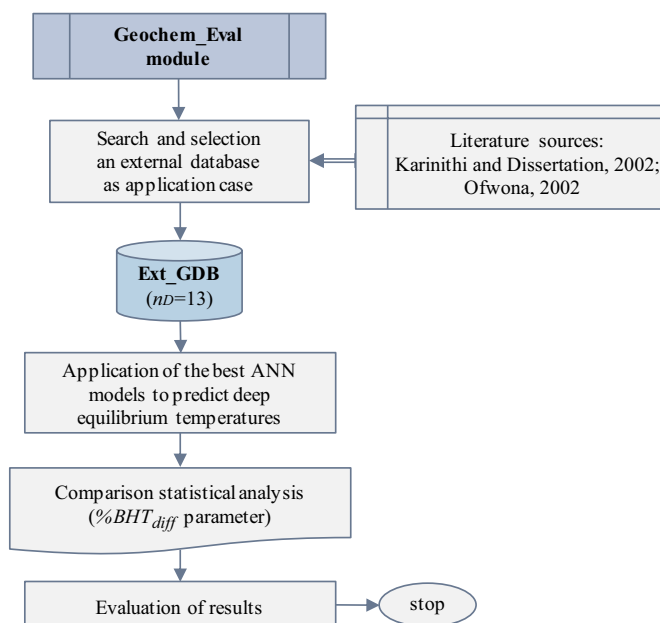


Fig. 5. Schematic flow diagram showing the Geochem\_Eval module developed to evaluate the prediction performance of the ANN architectures using an external geochemical database (Ext\_GDB).

has been created from production wells located in several world geothermal fields (Fig. 1). A detailed description of the computational methodology in terms of the design and learning processes of ANN, and a comprehensive geochemometric analysis for the evaluation of the best ANN architectures are outlined.

## 2. Work methodology

To carry out the multivariate analysis among the gas-phase compositions of geothermal fluids and bottomhole temperature measurements ( $BHT_m$ ), a general computational methodology was developed (see Figs. 2–5). Four computational modules were specifically created with some conditional statements (Q1-Q8):

- (i) the GasPhase\_WGDB module to compile a primary world geochemical database containing raw data of the gas-phase composition of geothermal fluids and  $BHT_m$  (Fig. 2),
- (ii) the GasPhase\_WoDB module to create representative working sub-databases with refined data of composition and temperature for controlling bias issues detected in the primary world geochemical database (Fig. 3),
- (iii) the ANN\_Modeling module to design ANN architectures to be evaluated for the prediction of geothermal reservoir temperatures (Fig. 4), and
- (iv) the Geochem\_Eval module to evaluate the prediction performance of the ANN architectures for predicting geothermal reservoir temperatures by using an external geochemical database as a validation or application case (Fig. 5).

### 2.1. GasPhase\_WGDB module

This module was used for the creation of a primary World Geochemical Database (hereafter referred as WGDB<sub>0</sub>) containing raw data of gas-phase compositions of geothermal fluids and  $BHT_m$  as a strict criterion of compilation, which was defined as the conditional statement Q<sub>2</sub> (Fig. 2). A total number of  $q_0 = 591$  samples were compiled in the WGDB<sub>0</sub> (14 columns and 591 rows). Steam fractions ( $S_j$ ) and gas-steam ratios ( $X_g$ ) were also compiled when these parameters were reported.

**Table 1**  
World geothermal fields and literature sources used for the creation of the primary WGDB<sub>0</sub> database ( $q_0 = 591$ ).

Geothermal field	Country	Number of gas-phase composition samples	Literature sources
Los Humeros (LHGF)	Mexico	268	López-Mendiola and Munguía, 1989; Barragán et al. (1991); Tello, 1992; Robles and Munguía (1993); Quijano and Torres (1995); Tello et al., 2000; Arellano et al. (2003); Arellano et al. (2006); Barragán et al. (2008)
Los Azufres (LAGF)	Mexico	197	Nieva et al. (1983), 1987; Portugal et al. (1984); Quijano-León (1985); Santoyo et al. (1991); Arellano et al. (2005); Barragán et al. (2005); González-Partida et al. (2005); Barragán et al. (2008)
Cerro Prieto (CPGF)	Mexico	35	Nieva et al. (1982); Nehring and D'Amore (1984)
Larderello (LGF)	Italy	30	D'Amore et al. (1983)
Krafla (KGF)	Iceland	26	Arnórsson and Gunnlaugsson (1985); Ping and Ármannsson (1996); Guðmundsson and Arnórsson (2002); Arnórsson et al. (2010)
Hellisheidi (HGF)	Iceland	10	Stefánsson et al. (2011)
Námafjall (NaGF)	Iceland	8	Arnórsson and Gunnlaugsson (1985); Guðmundsson and Arnórsson (2002)
Takigami (TGF)	Japan	6	Furuya et al. (2000)
Berlin (BeGF)	El Salvador	6	D'Amore and Mejía (1999)
Olkaria (OGF)	Kenya	3	Opondo, 2008
Nesjavellir (NGF)	Iceland	2	Ping and Ármannsson (1996)

The primary source of WGDB<sub>0</sub> was created from a wide selection of geothermal literature sources (i.e., papers published in international journals, congress proceedings, and technical reports). The gas-phase compositions of fluids were compiled from chemical analyses carried out in 149 production wells of eleven geothermal fields located in six countries (El Salvador, Iceland, Italy, Japan, Kenya, and Mexico: see Fig. 1 for their location). A summarized description of each geothermal field and the original data sources are also reported in Table 1.

The compilation of the gas-phase compositions included the concentration of major ( $CO_2$ ,  $H_2S$ , and  $NH_3$ ) and trace ( $H_2$ ,  $CH_4$  and  $N_2$ ) gas species; whereas the  $BHT_m$  data (in °C) were logged from transient production well tests. As the concentration of gas species are reported in a wide variety of concentration units [e.g., mol/kg steam as wet-basis, % wt, mmol/mol as dry-basis, %mol, mg/kg, mol fraction, among others], a single concentration unit was adopted for this study [mmol/mol]. The molar concentration unit was selected as working standard unit because it is used by most of the existing gas geothermometers for the determination of geothermal reservoir temperatures. This conditional statement was defined as  $Q_3$  (Fig. 2). To calculate all the necessary unit conversions in WGDB<sub>0</sub>, a Fortran computer code (GeoGasPlus) was programmed. This computer code uses the steam fraction ( $S_f$ ) to estimate the quantities from either dry-to-wet basis or vice versa. A summary of the descriptive statistics applied to the WGDB<sub>0</sub> is reported in Table 2.

## 2.2. GasPhase\_WorkDB module

This module was used to create working sequential sub-databases to

be used for the design and implementation of ANN architectures. As the original database (WGDB<sub>0</sub>) showed some bias issues due to incomplete chemical analyses of gases and duplicate data of  $BHT_m$  (i.e., samples with different gas-phase compositions for a same geothermal well but related with a single value of  $BHT_m$ ), quality control methods (A, B and C) were used for creating three working sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>), and with these, to eliminate such issues (Fig. 3).

### 2.2.1. Quality control method [A]

The original WGDB<sub>0</sub> ( $q_0 = 591$  samples) was analyzed by applying a criterion to verify the full availability of the concentrations of  $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$  per sample, as conditional statement  $Q_4$  (Fig. 3). All the samples that did not fulfill this condition were separated from WGDB<sub>0</sub>. After applying this criterion, 64 samples were separated ( $ns_A$ ). The remaining data set with  $q_1 = 527$  samples was hereafter referred as the first working sub-database (WG\_SubDB<sub>1</sub>). A summary of the descriptive statistics of the WG\_SubDB<sub>1</sub> is presented in Table 3.

### 2.2.2. Quality control method [B]

The first working sub-database WG\_SubDB<sub>1</sub> ( $q_1 = 527$  samples) was afterwards analyzed by applying a second quality control method under the following sequential conditions (Fig. 3)

- (B.1) the total sum of gas concentrations must be equal to 1000 mmol/mol (as conditional statement  $Q_5$ ). According to Nicholson (1993) this parameter may have an analytical uncertainty of  $\pm 5\%$  ( $\pm 50$  mmol/mol). Gas-phase compositions of fluids falling outside this concentration interval were separated,
- (B.2) the interval of geothermal temperatures used for this study was limited from 170 °C to 365 °C (as conditional statement  $Q_6$ ), which was assumed to consider middle-to high-enthalpy geothermal systems.  $BHT_m$  data falling outside this interval were also separated, and
- (B.3) the concentration of  $CO_2$  reported for the gas-phase composition of each sample must be greater than 70%wt (as conditional statement  $Q_7$ ). Gas-phase compositions of fluids with  $CO_2$  concentrations lower than this percentage were separated.

After applying these sequential conditions, a total number of 29 samples were identified and separated from the WG\_SubDB<sub>1</sub>. As a result of this, a remaining data set with  $q_2 = 498$  samples were obtained, which was referred as the second working sub-database (WG\_SubDB<sub>2</sub>): Fig. 3. A summary of the descriptive statistics of the gas-phase compositional data of the WG\_SubDB<sub>2</sub> is presented in Table 4.

### 2.2.3. Quality control method [C]

The second working sub-database WG\_SubDB<sub>2</sub> ( $q_2 = 498$  samples) was subsequently analyzed to identify bias problems caused by duplicate samples containing gas-phase compositions (coming from different sampling and analysis dates) for the same well but associated to a single value of  $BHT_m$ . This problem was defined by means of the conditional statement  $Q_8$  (Fig. 3).

The problem of duplicate  $BHT_m$  data is commonly due to the high-cost that represent the temperature logging in geothermal well completion operations ( $\sim 10\%$  of the overall cost), which justifies the limited number of  $BHT_m$  reported per well (García et al., 1998).

After applying the C method, 448 samples were identified with these problems, from which 445 samples corresponded to 41 geothermal wells of Mexico (9 wells from LHGF with 269 samples; 19 wells from LAGF with 141 samples; and 13 wells from CPGF with 35 samples), and 3 samples from Iceland (1 well from KGF). To solve these bias problems, the single values of  $BHT_m$  (corresponding to 42 geothermal wells) required to be statistically associated to mean concentration values of the gases ( $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$ ) per well. The mean concentration values for each gas were calculated assuming that the group of samples per well (with a single  $BHT_m$ ) comes from a statistical normal

**Table 2**  
Summary of the descriptive statistics of the WGDB<sub>0</sub> database ( $q_0 = 591$ ).

	<i>n</i>	<i>Min</i>	<i>Max</i>	Central tendency statistical parameters				Dispersion statistical parameters				
				<i>Mean</i> $\bar{x}$	Robust estimators			<i>Standard deviation s</i>	Robust estimators <sup>a</sup>			<i>S<sub>n</sub></i>
			<i>Median</i> $\tilde{x}$	<i>Trimean</i> $(E_L)_T$	<i>Gastwirth mean</i> $(E_L)_G$	<i>Median absolute deviation</i> <i>MAD</i>						
<b>CO<sub>2</sub> (mmol/mol)</b>												
Berlin	6	398	3206	1453	1130	1280	1130	1022	950	930	1100	1070
Cerro Prieto	35	713	911	842	844	840	840	35	32	31	31	30
Hellisheidi	10	396	846	597	608	590	590	163	256	224	191	182
Krafla	26	606	938	792	795	800	800	89	81	94	93	90
Larderello	30	928	981	952	951	950	950	16	20	19	16	17
Los Azufres	197	336	996	935	969	960	970	98	25	25	25	29
Los	268	699	994	920	921	920	920	50	41	42	42	43
Humeros												
Námafjall	8	311	938	681	699	700	700	222	311	272	287	254
Nesjavellir	2	328	801	564	564			335				
Olkaria	3	864	944	904	904		900	40				
Takigami	6	676	851	780	794	790	790	74	97	93	101	87
<b>H<sub>2</sub>S (mmol/mol)</b>												
Berlin	6	10	197	82	61	71	61	70	70	69	79	78
Cerro Prieto	35	26	83	54	54	54	54	11	6	8	8	8
Hellisheidi	10	65	357	228	247	244	243	95	102	110	110	107
Krafla	26	4	188	77	80	77	79	42	44	43	42	42
Larderello	30	5	26	16	16	16	16	6	7	6	6	6
Los Azufres	197	3	108	24	16	17	17	22	11	11	11	12
Los	268	0.1	193	55	56	54	55	34	30	32	30	32
Humeros												
Námafjall	8	14	211	108	90	102	94	83	112	108	84	95
Nesjavellir	2	144	230	187	187			61				
Olkaria	3	2	90	33	7		8	50				
Takigami	6	41	110	71	65	69	65	25	29	28	27	30
<b>CH<sub>4</sub> (mmol/mol)</b>												
Berlin	6	0.6	4.9	2.5	2.7	2.3	2.6	1.5	1.8	1.9	2.3	1.8
Cerro Prieto	35	20.7	56.8	38.7	40	39.1	39.6	7.9	6.5	7.9	7.7	7.7
Hellisheidi	10	0.3	5.9	2.5	2.3	2.4	2.3	1.6	1.7	1.7	1.7	1.7
Krafla	26	0.1	17.1	3.2	0.99	1.5	1.3	4.7	1	0.95	1.3	1.4
Larderello	30	1.9	19.3	9.2	9.4	9	9.3	4.9	6.7	5.9	5.1	5.2
Los Azufres	163	0.0001	3.2	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.4	0.5
Los	241	0.00006	49.5	2.9	0.95	1.1	1	6.6	0.99	0.99	0.97	1.1
Humeros												
Námafjall	8	0.5	26	4.5	1.3	1.5	1.4	8.7	1	1.1	1.2	1.2
Nesjavellir	2	0.96	2.9	1.9	1.9			1.4				
Olkaria	3	2	7.1	4.3	3.9		3.9	2.6				
Takigami	6	5.1	12	8.1	7.9	7.9	7.9	2.3	1.8	2.1	2.6	2.2
<b>H<sub>2</sub> (mmol/mol)</b>												
Berlin	6	3.1	18	12.8	15.5	14	15.5	6.3	4.4	4.3	5	5.8
Cerro Prieto	35	17.7	65.4	35.7	36.9	35.6	35.4	12.5	13.5	12.6	12.7	12.9
Hellisheidi	10	5.3	194.2	121	134.9	129	134	67.1	86	78	70	76
Krafla	25	2.6	174.5	65.1	63.1	65	62	42.8	52	46	45	44.7
Larderello	30	9.8	27.8	18.1	17.4	17.8	17.5	4.9	5.2	5.3	5.3	5.2
Los Azufres	167	0.6	120.8	7.2	3.9	4.5	4.4	11.5	2.9	3	3	3.3
Los	263	0.1	126.8	10.1	8.2	8.1	7.9	11.3	6.4	6.8	5.9	6.6
Humeros												
Námafjall	8	0.5	477	131.3	17.7	76	41.6	180.1	27.9	23.3	25.4	38
Nesjavellir	2	15.5	420.2	217.9	217.9			286.2				
Olkaria	2	1.7	24.5	13.1	13.1			16.2				
Takigami	6	0.4	2	0.97	0.9	0.9	0.9	0.6	0.4	0.4	0.4	0.4
<b>BHT<sub>m</sub> (°C)</b>												
Berlin	6	239	305	279	292	283	292	27	15	17	15	20
Cerro Prieto	35	279	325	297	296	296	296	13	14	13	14	14
Hellisheidi	10	270	320	288	283	284	284	16	12	15	16	15
Krafla	26	170	340	289	300	304	305	48	38	35	29	40
Larderello	30	230	270	249	250	250	247	10	15	12	20	10
Los Azufres	197	248	352	285	278	281	281	20	16	16	13	17
Los	268	204	384	300	340	306	330	60				
Humeros												
Námafjall	8	182	320	249	240	246	243	45	54	48	59	51
Nesjavellir	2	278	306	292	292			20				
Olkaria	3	240	343	307	338		338	58				
Takigami	6	205	246	222	216	221	216	16	13	14	14	18

<sup>a</sup> Dispersion robust parameters adjusted with correction factors to convert the sample estimates to population parameters (Verma et al., 2016b, 2017).



**Table 3**  
Summary of the descriptive statistics of the WG\_SubDB<sub>1</sub> sub-database ( $q_1 = 527$ ).

	<i>n</i>	<i>Min</i>	<i>Max</i>	Central tendency statistical parameters				Dispersion statistical parameters				
				<i>Mean</i> $\bar{x}$	Robust estimators			<i>Standard deviation s</i>	Robust estimators <sup>a</sup>			<i>S<sub>n</sub></i>
			<i>Median</i> $\tilde{x}$	<i>Trimean</i> $(E_L)_T$	<i>Gastwirth mean</i> $(E_L)_G$	<i>Median absolute deviation</i> <i>MAD</i>						
<b>CO<sub>2</sub> (mmol/mol)</b>												
Berlin	6	398	3206	1453	1130	1280	1130	1022	950	930	1100	1070
Cerro Prieto	35	713	911	842	844	840	840	35	32	31	31	30
Hellisheidi	10	396	846	597	608	590	590	163	256	224	191	182
Krafla	25	606	938	788	786	790	790	89	82	95	93	92
Larderello	30	928	981	952	951	950	950	16	20	19	16	17
Los Azufres	163	410	990	934	964	960	960	81	26	26	26	31
Los	240	699	994	920	921	930	920	52	43	43	41	45
Humeros												
Námafjall	8	311	938	681	699	700	700	222	311	272	287	254
Nesjavellir	2	328	801	564	564			335				
Olkaria	2	864	904	884	884			29				
Takigami	6	676	851	780	794	790	790	74	97	93	101	87
<b>H<sub>2</sub>S (mmol/mol)</b>												
Berlin	6	10	197	82	61	71	61	70	70	69	79	78
Cerro Prieto	35	26	83	54	54	54	54	11	6	8	8	8
Hellisheidi	10	65	357	228	247	244	243	95	102	110	110	107
Krafla	25	4	188	76	77	75	76	42	49	46	46	43
Larderello	30	5	26	16	16	16	16	6	7	6	6	6
Los Azufres	163	3	108	26	18	18	18	24	13	13	12	13
Los	240	0.1	193	54	56	52	54	35	31	32	30	33
Humeros												
Námafjall	8	14	211	108	90	102	94	83	112	108	84	95
Nesjavellir	2	144	230	187	187			61				
Olkaria	2	7	90	49	49			59				
Takigami	6	41	110	71	65	69	65	25	29	28	27	30
<b>CH<sub>4</sub> (mmol/mol)</b>												
Berlin	6	0.6	4.9	2.5	2.7	2.3	2.6	1.5	1.8	1.9	2.3	1.8
Cerro Prieto	35	20.7	56.8	38.7	40	39.1	39.6	7.9	6.5	7.9	7.7	7.7
Hellisheidi	10	0.3	5.9	2.5	2.3	2.4	2.3	1.6	1.7	1.7	1.7	1.7
Krafla	25	0.1	17.1	3.2	0.98	1.4	1.1	4.8	0.99	0.9	1.2	1.3
Larderello	30	1.9	19.3	9.2	9.4	9	9.3	4.9	6.7	5.9	5.1	5.2
Los Azufres	163	0.0001	3.2	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.4	0.5
Los	240	0.00006	49.5	2.9	0.95	1.1	1	6.6	1	0.99	0.97	1.1
Humeros												
Námafjall	8	0.5	26	4.5	1.3	1.5	1.4	8.7	1	1.1	1.2	1.2
Nesjavellir	2	0.96	2.9	1.9	1.9			1.4				
Olkaria	2	2	7.1	4.5	4.5			3.6				
Takigami	6	5.1	12	8.1	7.9	7.9	7.9	2.3	1.8	2.1	2.6	2.2
<b>H<sub>2</sub> (mmol/mol)</b>												
Berlin	6	3.1	18	12.8	15.5	14	15.5	6.3	4.4	4.3	5	5.8
Cerro Prieto	35	17.7	65.4	35.7	36.9	35.6	35.4	12.5	13.5	12.6	12.7	12.9
Hellisheidi	10	5.3	194.2	121	134.9	129	134	67.1	86	78	70	76
Krafla	25	2.6	174.5	65.1	63.1	65	62	42.8	52	46	45	44.7
Larderello	30	9.8	27.8	18.1	17.4	17.8	17.5	4.9	5.2	5.3	5.3	5.2
Los Azufres	163	0.6	32.2	6.2	3.8	4.4	4.3	6	2.7	2.9	2.9	3.2
Los	240	0.1	126.8	9.7	8.1	7.7	7.8	11	6.6	7	5.9	6.6
Humeros												
Námafjall	8	0.5	477	131.3	17.7	76	41.6	180.1	27.9	23.3	25.4	38
Nesjavellir	2	15.5	420.2	217.9	217.9			286.2				
Olkaria	2	1.7	24.5	13.1	13.1			16.2				
Takigami	6	0.4	2	0.97	0.9	0.9	0.9	0.6	0.4	0.4	0.4	0.4
<b>BHT<sub>m</sub> (°C)</b>												
Berlin	6	239	305	279	292	283	292	27	15	17	15	20
Cerro Prieto	35	279	325	297	296	296	296	13	14	13	14	14
Hellisheidi	10	270	320	288	283	284	284	16	12	15	16	15
Krafla	25	170	340	289	300	304	306	49	38	33	39	41
Larderello	30	230	270	249	250	250	247	10	15	12	20	10
Los Azufres	163	248	352	287	287	285	285	20	22	24	13	21
Los	240	204	384	298	340	306	325	61				
Humeros												
Námafjall	8	182	320	249	240	246	243	45	54	48	59	51
Nesjavellir	2	278	306	292	292			20				
Olkaria	2	338	343	341	341			4				
Takigami	6	205	246	222	216	221	216	16	13	14	14	18

<sup>a</sup> Dispersion robust parameters adjusted with correction factors to convert the sample estimates to population parameters (Verma et al., 2016b, 2017).

**Table 4**  
Summary of the descriptive statistics of the WG\_SubDB<sub>2</sub> sub-database ( $q_2 = 498$ ).

	<i>n</i>	<i>Min</i>	<i>Max</i>	Central tendency statistical parameters				Dispersion statistical parameters					
				<i>Mean</i> $\bar{x}$	Robust estimators			<i>Standard deviation s</i>	Robust estimators <sup>a</sup>			<i>S<sub>n</sub></i>	<i>Q<sub>n</sub></i>
			<i>Median</i> $\tilde{x}$	<i>Trimean</i> $(E_L)_T$	<i>Gastwirth mean</i> $(E_L)_G$	<i>Median absolute deviation</i> <i>MAD</i>							
<b>CO<sub>2</sub> (mmol/mol)</b>													
Berlin	1	790	790	790	790								
Cerro Prieto	35	713	911	842	844	840	840	35	32		31	31	30
Hellisheidi	2	828	846	837	837			12					
Krafla	22	707	938	811	806	810	800	68	71		76	79	73
Larderello	30	928	981	952	951	950	950	16	20		19	16	17
Los Azufres	160	761	990	943	964	960	960	52	25		25	25	29
Los	236	699	994	920	921	930	920	52	43		43	42	45
Humeros													
Námafjall	4	718	938	857	886	870	890	98					
Nesjavellir	1	801	801	801	801								
Olkaria	2	864	904	884	884			29					
Takigami	5	718	851	801	831	820	830	60	36		32	53	53
<b>H<sub>2</sub>S (mmol/mol)</b>													
Berlin	1	197	197	197	197								
Cerro Prieto	35	26	83	54	54	54	54	11	6		8	8	8
Hellisheidi	2	65	102	84	84			27					
Krafla	22	4	134	75	80	78	81	35	43		36	40	38
Larderello	30	5	26	16	16	16	16	6	7		6	6	6
Los Azufres	160	3	108	26	18	18	18	24	13		13	12	13
Los	236	0.1	193	54	56	52	55	36	31		33	30	33
Humeros													
Námafjall	4	14	119	57	47	52	47	46					
Nesjavellir	1	144	144	144	144								
Olkaria	2	7	90	49	49			59					
Takigami	5	55	110	77	70	72	71	23	27		24	27	29
<b>CH<sub>4</sub> (mmol/mol)</b>													
Berlin	1	2.7	2.7	2.7	2.7								
Cerro Prieto	35	20.7	56.8	38.7	40	39.1	39.6	7.9	6.5		7.9	7.7	7.7
Hellisheidi	2	0.96	5.9	3.4	3.4			3.5					
Krafla	22	0.1	17.1	3.4	0.99	1.4	1.2	5	1		0.95	1.3	1.4
Larderello	30	1.9	19.3	9.2	9.4	9	9.3	4.9	6.7		5.9	5.1	5.2
Los Azufres	160	0.0001	3.2	0.7	0.6	0.6	0.6	0.6	0.5		0.5	0.4	0.5
Los	236	0.00006	49.5	3	0.95	1.1	1	6.7	1		0.99	0.97	1.1
Humeros													
Námafjall	4	0.5	1.9	1.2	1.1	1.1	1.1	0.7					
Nesjavellir	1	0.96	0.96	0.96	0.96								
Olkaria	2	2	7.1	4.5	4.5			3.6					
Takigami	5	5.1	12	8.4	8	8.2	8.1	2.5	1.8		2.1	3.5	2.6
<b>H<sub>2</sub> (mmol/mol)</b>													
Berlin	1	16.7	16.7	16.7	16.7								
Cerro Prieto	35	17.7	65.4	35.7	36.9	35.6	35.4	12.5	13.5		12.6	12.7	12.9
Hellisheidi	2	5.3	61.3	33.3	33.3			39.6					
Krafla	22	2.6	127.6	61.2	60.3	60	59	36.2	36.4		39.2	39.6	38.9
Larderello	30	9.8	27.8	18.1	17.4	17.8	17.5	4.9	5.2		5.3	5.3	5.2
Los Azufres	160	0.6	30.1	6.1	3.8	4.4	4.3	5.7	2.7		2.8	2.9	3.2
Los	236	0.1	126.8	9.7	8.2	7.8	7.8	11.1	6.7		7.1	6	6.7
Humeros													
Námafjall	4	0.5	18.7	8.5	7.3	7.9	7.3	8.9					
Nesjavellir	1	15.5	15.5	15.5	15.5								
Olkaria	2	1.7	24.5	13.1	13.1			16.2					
Takigami	5	0.4	2	0.98	0.9	0.9	0.9	0.6	0.5		0.6	0.6	0.6
<b>BHT<sub>m</sub> (°C)</b>													
Berlin	1	305	305	305	305								
Cerro Prieto	35	279	325	297	296	296	296	13	14		13	14	14
Hellisheidi	2	270	275	273	273			4					
Krafla	22	170	340	289	300	304	306	51	38		32	38	40
Larderello	30	230	270	249	250	250	247	10	15		12	20	10
Los Azufres	160	248	352	287	287	286	285	21	22		24	13	21
Los	236	204	363	296	340	306	323	61					
Humeros													
Námafjall	4	182	240	222	233	227	233	27					
Nesjavellir	1	278	278	278	278								
Olkaria	2	338	343	341	341			4					
Takigami	5	205	238	217	215	215	215	12	5		8	10	8

<sup>a</sup> Dispersion robust parameters adjusted with correction factors to convert the sample estimates to population parameters (Verma et al., 2016b, 2017).

**Table 5**  
Isometric log-ratio (*ilr*) transformation equations for  $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$  concentrations.

Isometric log-ratio	Equation for transformation
$ilr_1$	$= \sqrt{\frac{1}{2}} \ln \left[ \frac{CO_2}{H_2S} \right]$
$ilr_2$	$= \sqrt{\frac{2}{3}} \ln \left[ \frac{(CO_2 \cdot H_2S)^2}{CH_4} \right]^{\frac{1}{2}}$
$ilr_3$	$= \sqrt{\frac{3}{4}} \ln \left[ \frac{(CO_2 \cdot H_2S \cdot CH_4)^3}{H_2} \right]^{\frac{1}{3}}$

The gas-phase compositional concentrations are in mmol/mol (dry basis).

distribution.

Before to carry out the statistical analysis and to minimize the rejection of a large number of outliers, logarithmic transformations were determined for all the gas compositions ( $CO_2$ ,  $H_2S$ ,  $H_2$ , and  $CH_4$ ) using isometric log-ratios (*ilr*), according to the method proposed by Egozcue et al. (2003). The *ilr* transformation required the use of the geometric mean to describe statistically sample distributions with isolated data. The *ilr* transformations were calculated by means of the Eq. (1):

$$ilr_i = \sqrt{\frac{i}{i+1}} \ln \left[ \frac{(X_1 \cdot X_2 \cdots X_i)^{\frac{1}{i}}}{X_{i+1}} \right]; \quad i = 1, 2, \dots, (k-1) \quad (1)$$

where  $i$  subscript is the *ilr*-transformed variable; and  $k$  subscript is the total number of variables for the gas species ( $k = 4$ :  $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$ );  $X_i$  represent the array of each input variable corresponding to the gas concentration data in the same order. The function  $\ln$  is the natural logarithm. Table 5 presents a listing of the *ilr* log-transformed equations used for these calculations.

When the number of samples per well was less than 5, a univariate analysis based on single-outlier discordancy tests was used for the detection/rejection of discordant outliers, whereas for a group of samples  $\geq 5$ , a multivariate analysis was applied. For the univariate analysis, Dixon, Grubbs, skewness and kurtosis statistical tests at the 99% confidence level were applied using the computer code UDASYS (Verma et al., 2013); whereas for the other analysis, multivariate discordancy tests were applied through the Wilks' test at the 99% confidence level using the computer code DOMuDaF (Verma et al., 2016a, b).

After applying this methodology, mean values of gas concentrations ( $CO_2$ ,  $H_2S$ ,  $H_2$ , and  $CH_4$ ) for each one of the 42 geothermal wells were estimated. In this way, a third working sub-database (WG\_SubDB<sub>3</sub>) was created with a total number of  $q_3 = 97$  samples (i.e., this value given by the sum of 55 original samples that did not report duplicate  $BHT_m$  data, and 42 samples coming from the mean concentration estimations: Fig. 3). A summary of the descriptive statistics of the WG\_SubDB<sub>3</sub> is presented in Table 6.

Although the sample size of WG\_SubDB<sub>3</sub> resulted so small, according to the standard training requirements of ANN, it was necessary to evaluate the effect of duplicate temperature data reported for the geothermal wells on the prediction performance of geothermal reservoir temperatures. A reduced number of training and validation data in costly and complex engineering processes has been actually justified in some previous ANN works reported in the literature (e.g., Johnson et al., 2018).

### 2.3. ANN\_Modeling module

This module was created to design the ANN architectures for the multivariate analysis among the gas compositions and  $BHT_m$  data (Fig. 4). All the ANN architectures used the well-known conceptual feed-forward model proposed by Poulton (2001). A general ANN model of three-layers was specifically adopted for this study, which was

characterized by containing an input layer with up to 5 input variables with 13 different combinations (see Table 7), a hidden layer ranging from 1 to 35 neurons, and an output layer with the predicted output ( $BHT_{ANN}$ ) which was compared against the desired target,  $BHT_m$  (see Fig. 6).

For the input layer, the input neurons  $X_1, X_2, \dots, X_5$  were defined as the independent variables which were computed from the gas concentration data of  $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$ . Up to 5 input variables were used by using the transformations and the different combinations listed in Table 7. The input variables, the concentration intervals, and the corresponding  $BHT_m$  data are described in Table 8.

The neurons in each layer are connected with other neurons in a subsequent layer by means of weighting and bias coefficients, which are adjusted during the training processes (Cortés et al., 2009). The coefficients associated with the hidden layer are grouped into a matrix  $IW$  (given as weights) and the vector  $b_1$  of biases. The output layer computes the weighted sum of inputs provided by the hidden layer, and the calculated coefficients are grouped into a matrix  $LW$  and the vector  $b_2$ , whereas  $Y$  represents the ANN output ( $BHT_{ANN}$ ). According to Fig. 6 and using the conventional matrix notation, the network output is calculated by means of the Eq. (2):

$$Y = BHT_{ANN} = LW \cdot f[IW \cdot X + b_1] + b_2 \quad (2)$$

where  $f$  corresponds to an activation function for the hidden layer, whereas a linear function is used for the output layer.

The activation function produces a non-linear decision boundary via non-linear combinations of the input neurons (Wagh et al., 2018). The activation functions commonly used for hidden layers are the logistic sigmoid and hyperbolic tangent sigmoid, whereas for output layers (if the ANN is used as approximation function, i.e., not for classification), a linear function is suggested (Yadav et al., 2018; Ye and Kim, 2018). For this study, the hyperbolic tangent sigmoid (tansig) function was used for the hidden layer, whereas the linear function (purelin) was adopted for the output layer.

For the learning process, the Levenberg–Marquardt back-propagation (LM) algorithm in combination with the gradient descent and Quasi-Newton methods were used. LM was selected because it is suggested as an efficient algorithm for training of small and medium sized ANN (Marquardt, 1963; Hagan and Menhaj, 1994; Pradeep et al., 2011; Ghritlahre and Prasad, 2018; Johnson et al., 2018; Smith et al., 2018; Ye and Kim, 2018). Like the quasi-Newton methods, the LM algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the cost function has the form of a sum of squares, the Hessian matrix may be approximated as:

$$H = J^T J \quad (3)$$

and the gradient can be computed as:

$$g = J^T e \quad (4)$$

where  $J$  is the Jacobian matrix which contains the first derivatives of the ANN error  $e$  according to the weights and biases. In relation to the present modeling case, the ANN error vector  $e$  given by the difference between the ANN output ( $BHT_{ANN}$ ) and the target ( $BHT_m$ ) for each pattern would be specified by:

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_q \end{bmatrix} = \begin{bmatrix} BHT_{m_1} - BHT_{ANN_1} \\ BHT_{m_2} - BHT_{ANN_2} \\ \vdots \\ BHT_{m_q} - BHT_{ANN_q} \end{bmatrix} \quad (5)$$

where  $q$  is the total number of patterns (or samples).

According to Hagan et al. (2014), the Jacobian matrix may be computed through a standard backpropagation technique which is less complicated than the calculation of the Hessian matrix. The LM algorithm uses this approximation to the Hessian matrix in the following

**Table 6**  
Summary of the descriptive statistics of the WG\_SubDB<sub>3</sub> sub-database ( $q_3 = 97$ ).

	n	Central tendency statistical parameters						Dispersion statistical parameters					
		Min	Max	Mean $\bar{x}$	Robust estimators		Gastwirth mean $(E_L)_G$	Standard deviation $s$	Robust estimators <sup>a</sup>		$S_n$	$Q_n$	$\sigma_n$
					Median $\bar{x}$	Trimean $(E_L)_T$			Median absolute deviation $MAD$				
<b>CO<sub>2</sub> (mmol/mol)</b>													
Cerro Prieto	13	820	875	848	849	850	850	16	15		15	18	17
Hellisheidi	2	828	846	837	837			12					
Krafla	10	805	929	847	838	840	840	38	37		36	35	37
Larderello	30	928	981	952	951	950	950	16	20		19	16	17
Los Azufres	21	894	987	948	959	950	950	30	38		31	28	33
Los	12	832	976	898	892	900	890	42	26		38	46	37
Humeros													
Námafjall	3	860	938	904	912		910	40					
Nesjavellir	1	801	801	801	801								
Olkaria	2	864	904	884	884			29					
Takigami	3	831	851	843	847		850	11					
<b>H<sub>2</sub>S (mmol/mol)</b>													
Cerro Prieto	13	39	68	53	54	55	54	8	7		8	7	9
Hellisheidi	2	65	102	84	84			27					
Krafla	10	40	121	79	80	81	78	26	26		28	30	29
Larderello	30	5	26	16	16	16	16	6	7		6	6	6
Los Azufres	21	6	61	27	22	25	22	18	19		19	16	20
Los	12	13	100	64	68	67	69	27	32		31	32	30
Humeros													
Námafjall	3	14	119	55	32		32	56					
Nesjavellir	1	144	144	144	144								
Olkaria	2	7	90	49	49			59					
Takigami	3	55	110	85	90		90	28					
<b>CH<sub>4</sub> (mmol/mol)</b>													
Cerro Prieto	13	27.4	49.7	38.1	38	38.1	37.7	5.1	4.1		4.0	4.3	4.3
Hellisheidi	2	0.96	5.9	3.4	3.4			3.5					
Krafla	10	0.1	5.7	1.2	0.6	0.6	0.6	1.6	0.6		0.5	0.5	0.6
Larderello	30	1.9	19.3	9.2	9.4	9	9.3	4.9	6.7		5.9	5.1	5.2
Los Azufres	21	0.2	1.8	0.7	0.7	0.7	0.7	0.4	0.4		0.4	0.4	0.4
Los	12	0.4	34.5	11.3	8.1	9.1	8.6	11.7	12.1		12.8	9.1	12.7
Humeros													
Námafjall	3	0.5	1.9	1	0.7		0.7	0.8					
Nesjavellir	1	0.96	0.96	0.96	0.96								
Olkaria	2	2	7.1	4.5	4.5			3.6					
Takigami	3	7.7	12	9.2	8		8	2.4					
<b>H<sub>2</sub>(mmol/mol)</b>													
Cerro Prieto	13	19.8	57.8	36.6	36.1	35.5	35.7	11.6	12		12.6	12.8	12.9
Hellisheidi	2	5.3	61.3	33.3	33.3			39.6					
Krafla	10	3.6	83.3	55.3	60.3	58	61	25.3	28.3		25.8	26.3	27.6
Larderello	30	9.8	27.8	18.1	17.4	17.8	17.5	4.9	5.2		5.3	5.3	5.2
Los Azufres	21	0.96	15.7	5.5	4.9	5	5.1	3.4	2.9		3.3	3.1	3
Los	12	2.2	24.7	13.4	12.9	13.5	13.1	7	9.1		7.8	7.8	7.7
Humeros													
Námafjall	3	0.5	18.7	6.9	1.6		1.6	10.2					
Nesjavellir	1	15.5	15.5	15.5	15.5								
Olkaria	2	1.7	24.5	13.1	13.1			16.2					
Takigami	3	0.4	0.9	0.6	0.6		0.6	0.3					
<b>BHT<sub>m</sub> (°C)</b>													
Cerro Prieto	13	279	325	298	296	296	297	15	16		17	17	16
Hellisheidi	2	270	275	273	273			4					
Krafla	10	215	340	311	325	320	324	36	9		12	16	12
Larderello	30	230	270	249	250	250	247	10	15		12	20	10
Los Azufres	21	248	352	288	278	282	282	26	20		20	21	23
Los	12	204	356	303	313	309	312	44	45		44	39	44
Humeros													
Námafjall	3	182	240	221	240		240	33					
Nesjavellir	1	278	278	278	278								
Olkaria	2	338	343	341	341			4					
Takigami	3	212	238	222	217		217	14					

<sup>a</sup> Dispersion robust parameters adjusted with correction factors to convert the sample estimates to population parameters (Verma et al., 2016b, 2017).

Newton-like update:

$$w_{t+1} = w_t - [J_t^T J_t + \mu I]^{-1} J_t^T e_t \tag{6}$$

where  $w_t$  is the current weight,  $w_{t+1}$  is the next weight,  $I$  is the identity matrix,  $\mu$  is the learning factor.

To minimize the cost function (which is given by the sum of squares), weight and bias values are iteratively adjusted by using the LM

algorithm (Eq. (6)). According to Hagan et al. (2014) and Smith et al. (2018), if the cost function  $F(W)$  is given by the sum of squares function:

$$F(W) = \sum_{i=1}^q e_i^2(W) = e_i^T(W) e_i(W) \tag{7}$$

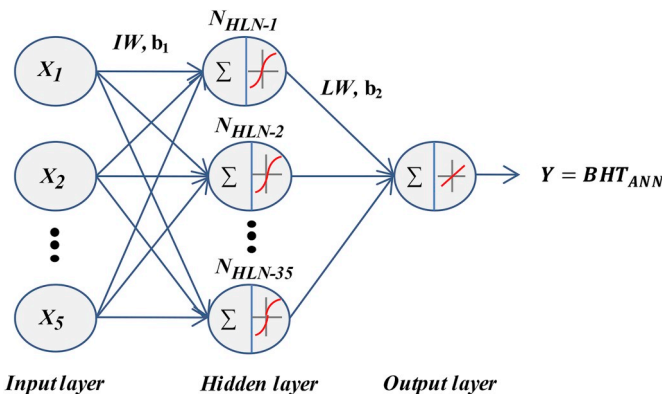
then, the  $j$ th element of the gradient may be expressed as:

**Table 7**

Thirteen arrays or combinations (Comb) of the gas-phase composition data (CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub>; concentration ratios and log-transformations) used as input variables for designing and training of ANN architectures using the three working sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>).

Comb	Input variable (gas concentrations, ratios and log-transformations)	Number Input neurons	Number Hidden neurons	Number Output neurons	First ANN architecture	Last ANN architecture	Number ANN Architectures
1	ln (CO <sub>2</sub> )	1	from 1 to 35	1	[1-1-1]	[1-35-1]	35
2	ln (H <sub>2</sub> S)	1	from 1 to 35	1	[1-1-1]	[1-35-1]	35
3	ln (CO <sub>2</sub> / H <sub>2</sub> )	1	from 1 to 35	1	[1-1-1]	[1-35-1]	35
4	ln (H <sub>2</sub> S/ H <sub>2</sub> )	1	from 1 to 35	1	[1-1-1]	[1-35-1]	35
5	ln (CH <sub>4</sub> / CO <sub>2</sub> ) ln(H <sub>2</sub> S/ H <sub>2</sub> )	2	from 1 to 35	1	[2-1-1]	[2-35-1]	35
6	H <sub>2</sub> S/ CO <sub>2</sub> CH <sub>4</sub> / CO <sub>2</sub> H <sub>2</sub> /CO <sub>2</sub>	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
7	ln (H <sub>2</sub> S/ CO <sub>2</sub> ) ln (CH <sub>4</sub> / CO <sub>2</sub> ) ln (H <sub>2</sub> / CO <sub>2</sub> )	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
8	ln (CO <sub>2</sub> / H <sub>2</sub> S) ln (CH <sub>4</sub> / H <sub>2</sub> S) ln (H <sub>2</sub> / H <sub>2</sub> S)	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
9	ln (CO <sub>2</sub> / CH <sub>4</sub> ) ln (H <sub>2</sub> S/ CH <sub>4</sub> ) ln (H <sub>2</sub> / CH <sub>4</sub> )	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
10	ln (CO <sub>2</sub> / H <sub>2</sub> ) ln (H <sub>2</sub> S/ H <sub>2</sub> ) ln (CH <sub>4</sub> / H <sub>2</sub> )	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
11	itr <sub>1</sub> itr <sub>2</sub> itr <sub>3</sub>	3	from 1 to 35	1	[3-1-1]	[3-35-1]	35
12	ln (CO <sub>2</sub> ) ln (H <sub>2</sub> S) ln (CH <sub>4</sub> ) ln (H <sub>2</sub> )	4	from 1 to 35	1	[4-1-1]	[4-35-1]	35
13	ln (H <sub>2</sub> S/ CO <sub>2</sub> ) ln (CH <sub>4</sub> / CO <sub>2</sub> ) ln (H <sub>2</sub> / CO <sub>2</sub> ) ln (H <sub>2</sub> S) ln (H <sub>2</sub> S/ H <sub>2</sub> )	5	from 1 to 35	1	[5-1-1]	[5-35-1]	35

Total number of evaluated ANN architectures: 455.



**Fig. 6.** Schematic flow diagram showing the conceptual model of the three-layer artificial neural network used in this study.

$$[\nabla F(W)]_j = \frac{\partial F(W)}{\partial w_j} = 2 \sum_{i=1}^q e_i(W) \frac{\partial e_i(W)}{\partial w_j} \quad (8)$$

where  $W$  is representing to the vector of the weight and bias values. The gradient may be then modified in matrix form as:

$$\nabla F(W) = 2J^T(W)e(W) \quad (9)$$

**Table 8**

Concentration interval of the gas-phase composition data compiled in all the WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub> sub-databases, and used for the design and evaluation of the ANN architectures.

Database	Total data (q)	CO <sub>2</sub> mmol/mol, dry basis	H <sub>2</sub> S	CH <sub>4</sub>	H <sub>2</sub>	BHT <sub>m</sub> (°C)
WGDB <sub>0</sub>	591	311–3206	0.1–357	0.00006–56.8	0.1–477	170–384
WG_SubDB <sub>1</sub>	527	311–3206	0.1–357	0.00006–56.8	0.1–477	170–384
WG_SubDB <sub>2</sub>	498	699–994	0.1–197	0.00006–56.8	0.1–127.6	170–363
WG_SubDB <sub>3</sub>	97	801–987	4.6–144	0.1–50	0.4–83	182–356

The Jacobian matrix is calculated as follows:

$$J(W) = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \dots & \frac{\partial e_1}{\partial w_p} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \dots & \frac{\partial e_2}{\partial w_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_q}{\partial w_1} & \frac{\partial e_q}{\partial w_2} & \dots & \frac{\partial e_q}{\partial w_p} \end{bmatrix} \quad (10)$$

where  $p$  is the total number of neurons,  $q$  is the total number of patterns; and  $W$  is the vector of weight and bias values.

During training, and as a result of the combination of the descent gradient and the Gauss–Newton methods, the LM algorithm switched between these two methods. When the combination coefficient or learning rate factor ( $\mu$ ) was very small (nearly zero), the Quasi–Newton method was applied whereas for larger values of  $\mu$ , the descent gradient method was preferentially used (Smith et al., 2018). The learning rate factor acts as a fundamental specification of the ANN models either in identifying over-learning issues or when the training process requires to be stopped (Attoh-Okine, 1999; Sahoo and Jha, 2013). The learning rate factor is one of the most important hyper-parameters to be tuned for training neural networks after using some initialization values which may range from 0.01 to 0.00001. The learning rate factor usually determined the adjustment of the weights and bias during training. A suitable selection of this parameter splits the signal from the noise, and

**Table 9**  
Statistical parameters used for the evaluation of the ANN architectures.

Parameter	Equation	Eq.
Correlation coefficient $r$	$r = \frac{\sum_i^q (BHT_{ANN(i)} - \overline{BHT}_{ANN})(BHT_{m(i)} - \overline{BHT}_m)}{\sqrt{\sum_i^q (BHT_{ANN(i)} - \overline{BHT}_{ANN})^2 \sum_i^q (BHT_{m(i)} - \overline{BHT}_m)^2}}$	(12)
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{q} \sum_{i=1}^q (BHT_{ANN(i)} - BHT_{m(i)})^2}$	(13)
Mean Absolute Error (MAE)	$MAE = \frac{1}{q} \sum_{i=1}^q  BHT_{ANN(i)} - BHT_{m(i)} $	(14)
Sensitivity analysis - Garson method	$I_j = \frac{\sum_{m=1}^{N_h} [ ( IW_{jm}  / \sum_{k=1}^{N_i}  IW_{km} ) \times  LW_{mj}  ]}{\sum_{k=1}^{N_i} \{ \sum_{m=1}^{N_h} [  IW_{km}  / \sum_{k=1}^{N_i}  IW_{km}  ] \times  LW_{mj}  \}}$	(15)

$BHT_m$  and  $\overline{BHT}_m$  is the bottomhole temperature measured for a sample and the corresponding mean values, respectively;  $BHT_{ANN}$  and  $\overline{BHT}_{ANN}$  is the temperature calculated from the ANN models and the corresponding mean values, respectively;  $q$  is de number of samples.  $I_j$  is the relative importance of the  $j$ -th input variable on the output variable;  $N_i$  and  $N_h$  are the number of input and hidden neurons, respectively;  $IW$  and  $LW$  are weights matrix of hidden and output layers, respectively; subscripts 'k', 'm', and 'n' refer to input, hidden, and output neurons, respectively.

avoid an over-fitting of the signal. Small learning rate values cause small weight changes whereas large values produce large changes.

2.3.1. Structure of the ANN architectures

As was previously mentioned, ANN\_Modeling module was specifically used for the design and implementation of all the ANN architectures under evaluation. This module was applied separately to each one of the three working sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>): Fig. 3. A total number of 455 ANN architectures were designed. The total number of ANN architectures was calculated as the sum of ANN architectures designed from the five input variables, the thirteen combinations of these, the number of neurons for the hidden layer ( $m = 1$  to 35), and the neuron of the output layer (see Table 7, and the iterative loops 1 and 2 represented in Fig. 4).

In order to obtain a more efficient ANN training, the input and output variables were systematically normalized between -1 and 1 by using the Eq. (11).

$$y = \frac{2(x - x_{min})}{x_{max} - x_{min}} - 1 \tag{11}$$

where  $y$  is the normalized variable between -1 and 1,  $x$  is the input variable value (i.e., the gas concentrations in mmol/mol units),  $x_{max}$  and

$x_{min}$  are the maximum and minimum values of the input variables, respectively.

According to Demuth and Beale (2003), Wang et al. (2015), Smith et al. (2018), and Ye and Kim (2018), this normalization process leads to a much better prediction performance of the ANN.

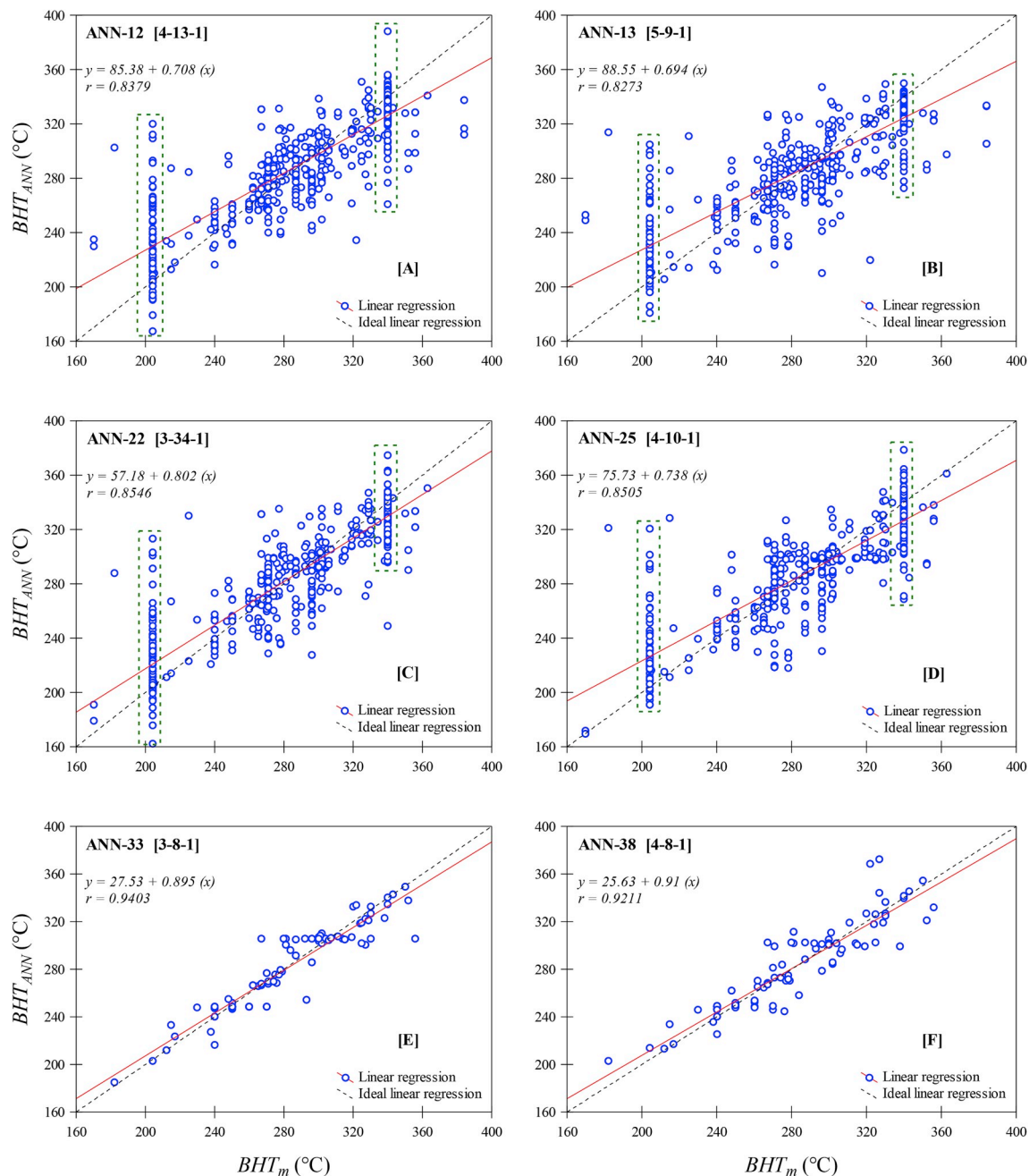
Following the methodology proposed by Serban et al. (2016) and Bhanot et al. (2018), each working sub-database under analysis was randomly split into training (80%), validation (10%), and testing (10%) for all the ANN training simulations. To initialize the evaluation process of the ANNs, each architecture was trained with the LM algorithm by considering four learning rate factors ( $\mu_1 = 0.01$ ,  $\mu_2 = 0.001$ ,  $\mu_3 = 0.0001$ , and  $\mu_4 = 0.00001$ ) which were efficiently tuned. The learning process for all the ANN architectures was simulated by programming computing scripts using the commercial software *Matlab*® for Windows v. R2017b (Math-Works, 2017). As a result of the ANN training simulations, a matrix of output data was generated as a function of some evaluation statistical parameters ( $r$ ,  $RMSE$ , and  $MAE$ ), and the weighting ( $IW$ ,  $LW$ ) and bias ( $b_1$ ,  $b_2$ ) coefficients obtained for each architecture tested.

The prediction performance of each ANN was then evaluated by analyzing the differences between the simulated ( $BHT_{ANN}$ ) and target ( $BHT_m$ ) values through the calculation of the following statistical parameters: (i) the correlation coefficient ( $r$ ) of a linear regression between

**Table 10**  
Summary of the best ANN architectures evaluated for the prediction of reservoir temperature using the sub-database WG\_SubDB<sub>1</sub> ( $q_1 = 527$  samples).

ANN	Comb	Input variables					ANN Architecture	Correlation coefficient (r)				RMSE global	MAE global
		1	2	3	4	5		global	training	validation	testing		
1	1	$\ln(CO_2)$					[1-32-1] $\mu^3$	0.6741	0.6935	0.5831	0.6167	34.8347	27.0785
2	2	$\ln(H_2S)$					[1-3-1] $\mu^4$	0.6596	0.6675	0.5910	0.6716	35.4517	27.5801
3	3	$\ln(CO_2/$ $H_2)$					[1-13-1] $\mu^4$	0.6593	0.6760	0.6361	0.5549	35.4675	27.8452
4	4	$\ln(H_2S/$ $H_2)$					[1-26-1] $\mu^3$	0.4533	0.4814	0.3514	0.3583	42.0594	32.6523
5	5	$\ln(CH_4/$ $CO_2)$	$\ln(H_2S/$ $H_2)$				[2-28-1] $\mu^1$	0.6039	0.6864	0.5650	0.3102	38.2079	27.9366
6	6	$H_2S/CO_2$	$CH_4/CO_2$	$H_2/CO_2$			[3-12-1] $\mu^4$	0.7845	0.7983	0.7409	0.7363	29.2610	20.9794
7	7	$\ln(H_2S/$ $CO_2)$	$\ln(CH_4/$ $CO_2)$	$\ln(H_2/$ $CO_2)$			[3-18-1] $\mu^3$	0.8230	0.8547	0.8008	0.6352	26.8382	18.3898
8	8	$\ln(CO_2/$ $H_2S)$	$\ln(CH_4/$ $H_2S)$	$\ln(H_2/$ $H_2S)$			[3-9-1] $\mu^1$	0.8015	0.8252	0.8047	0.6056	28.2169	19.4521
9	9	$\ln(CO_2/$ $CH_4)$	$\ln(H_2S/$ $CH_4)$	$\ln(H_2/$ $CH_4)$			[3-15-1] $\mu^4$	0.8283	0.8539	0.8219	0.6312	26.4545	18.9433
10	10	$\ln(CO_2/$ $H_2)$	$\ln(H_2S/$ $H_2)$	$\ln(CH_4/$ $H_2)$			[3-9-1] $\mu^4$	0.8165	0.8346	0.7911	0.7047	27.2437	19.3600
11	11	$ilr_1$	$ilr_2$	$ilr_3$			[3-9-1] $\mu^4$	0.8060	0.8308	0.8236	0.5941	27.9351	19.9262
12	12	$\ln(CO_2)$	$\ln(H_2S)$	$\ln(CH_4)$	$\ln(H_2)$		[4-13-1] $\mu^{2a}$	<b>0.8379</b>	<b>0.8568</b>	<b>0.7824</b>	<b>0.7499</b>	<b>25.7650</b>	<b>17.6007</b>
13	13	$\ln(H_2S/$ $CO_2)$	$\ln(CH_4/$ $CO_2)$	$\ln(H_2/$ $CO_2)$	$\ln$ $(H_2S)$	$\ln(H_2S/$ $H_2)$	[5-9-1] $\mu^{2a}$	<b>0.8273</b>	<b>0.8457</b>	<b>0.7937</b>	<b>0.7165</b>	<b>26.4971</b>	<b>18.3239</b>

<sup>a</sup> These ANN models obtained the best correlation coefficient  $r$  in the three sets (training, validation and testing) for the prediction of BHT by using gas-phase composition of geothermal fluids.  $\mu_1$  (0.01),  $\mu_2$  (0.001),  $\mu_3$  (0.0001), and  $\mu_4$  (0.00001) are the learning rate factors used.



**Fig. 7.** Results of a comparative analysis between simulated ( $BHT_{ANN}$ ) and experimental ( $BHT_m$ ) temperatures by using the best ANN architectures tested. [A] ANN-12 and [B] ANN-13 plots generated from the WG\_SubDB<sub>1</sub>; [C] ANN-22 and [D] ANN-25 plots generated from the WG\_SubDB<sub>2</sub>; and [E] ANN-33 and [F] ANN-38 plots generated from the WG\_SubDB<sub>3</sub>.

the predicted ( $BHT_{ANN}$ ) and the target ( $BHT_m$ ) variables, which was used as a convergence criterion for evaluating the prediction efficiency of the ANN trained (see Eq. (12) in Table 9); (ii) the Root Mean Square Error (RMSE), see Eq. 13 in Table 9; and, (iii) the Mean Absolute Error (MAE), see Eq. 14 in Table 9. A comprehensive analysis of all these statistical parameters was systematically used for selecting the best two ANN architectures for the prediction of the  $BHT_{ANN}$  (from each sub-database: WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>). To assess the relative importance of the ANN input variables on the prediction of deep geothermal reservoir temperatures, the well-known Garson method (see Eq. 15 in Table 9; Fig. 4) was methodically used according to the recursive procedure depicted in Fig. 4 (Garson, 1991; Yadav et al., 2018).

#### 2.4. Geochem\_Eval module

This computational module was created to evaluate the prediction performance of the ANNs for calculating deep reservoir temperatures using an external geochemical database (Ext\_GDB with  $nD = 13$  samples), as a validation or application case (Fig. 5). The external database was created on purpose to perform an additional evaluation of the best ANNs without considering any bias effect. With this goal, the gas-phase composition of fluids collected from the Olkaria, Kenya geothermal field was used. This field corresponds to a high-temperature geothermal system with reservoir temperatures ranging from 200 to 300 °C (Kar-ingithi and Dissertation, 2002; Ofwona, 2002). The external validation was specifically assessed through the calculation of the error percentage

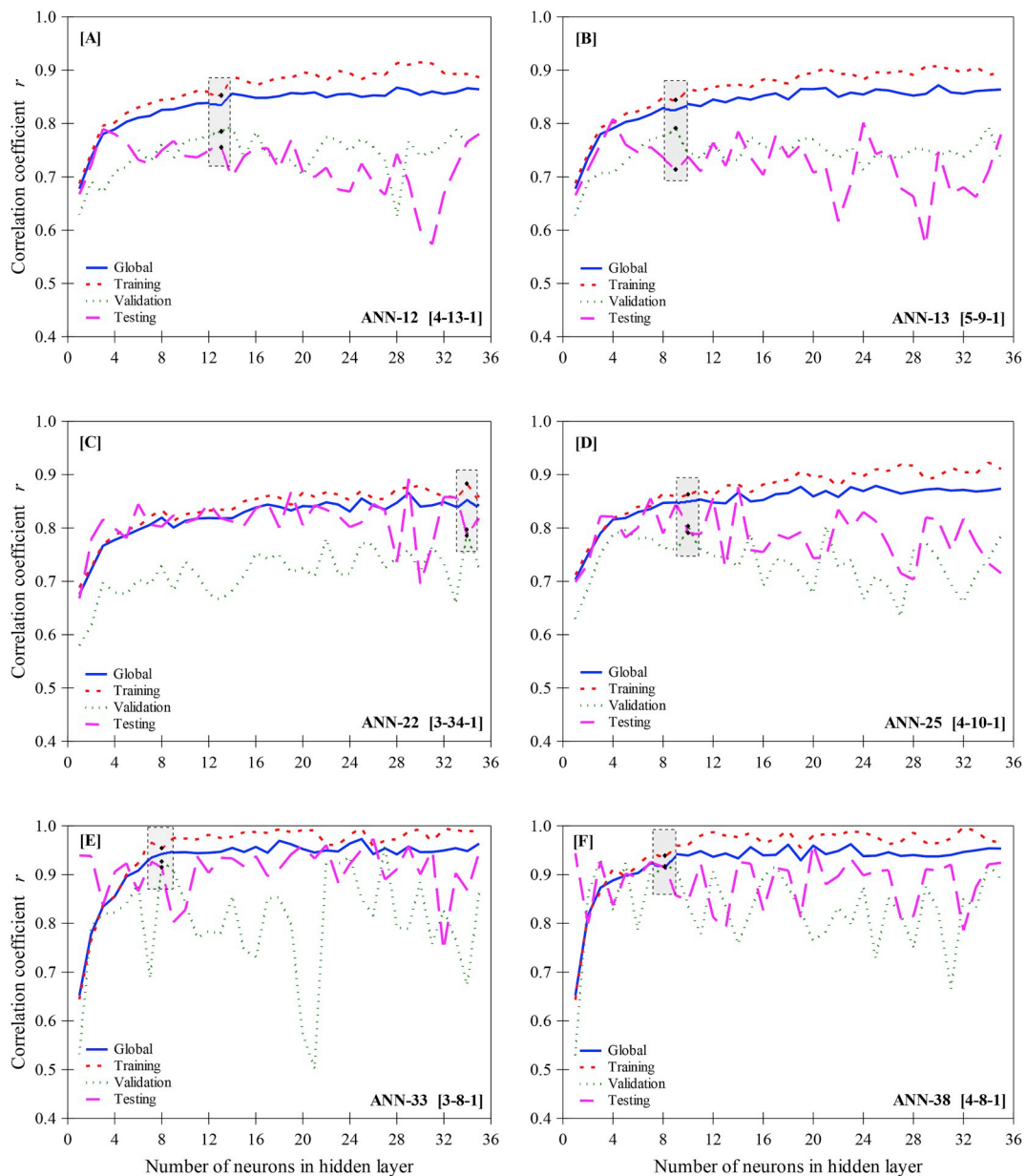


Fig. 8. Prediction performances of the best ANN architectures (ANN-33 and ANN-38) by using a close monitoring among the linear correlation coefficients and the numbers of neuron in hidden layer, and the learning rate factors.

difference ( $\%BHT_{diff}$ ) between the  $BHT_m$  and  $BHT_{ANN}$  data by using the Eq. (16), together with the calculation of the standard error of the mean.

$$\%BHT_{diff} = \left( \frac{BHT_{m(i)} - BHT_{ANN(i)}}{BHT_{m(i)}} \right) \cdot 100 \quad (16)$$

### 3. Results and discussion

Comprehensive statistical analyses applied to all the tested ANN architectures were performed by using the three working sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>). The results obtained from these analyses are described as follows.

#### 3.1. WG\_SubDB<sub>1</sub>

By considering the higher correlation coefficients ( $r$ ) obtained from the training processes of the WG\_SubDB<sub>1</sub> (global, training, validation, and testing) as well as the lower values of  $RMSE$  and  $MAE$ , the best

thirteen ANN architectures were preliminary selected and summarized in Table 10. The compilation of these evaluation results of ANN includes the number of neurons per architecture [input-hidden-output] and the initial learning rate factor ( $\mu$ ) used for obtaining the best prediction of  $BHT_{ANN}$  values.

From the trained ANN and prediction results, the ANN-12 [4-13-1] and ANN-13 [5-9-1] were identified as the best ANN architectures for predicting the  $BHT_{ANN}$  output variable with an acceptable accuracy (i.e., which closely approached to the  $BHT_m$  target variable). The ANN-12 and ANN-13 architectures used the same initial learning rate factor ( $\mu_2 = 0.001$ ) which lead to obtain  $RMSE$  values of 25.76 and 26.50, and  $MAE$  values of 17.60 and 18.32, respectively (see Table 10). The ANN-12 [4-13-1] architecture was characterized by 4 input variables [ $\ln(CO_2)$ ,  $\ln(H_2S)$ ,  $\ln(CH_4)$ , and  $\ln(H_2)$ ] for the input layer, 13 neurons for the hidden layer, and one neuron for the output layer; whereas the ANN-13 [5-9-1] with 5 input variables [ $\ln(H_2S/CO_2)$ ,  $\ln(CH_4/CO_2)$ ,  $\ln(H_2/CO_2)$ ,  $\ln(H_2S)$ , and  $\ln(H_2S/H_2)$ ] for the input layer, 9 neurons for the hidden layer, and one neuron ( $BHT_m$ ) for the output layer (see Table 10).



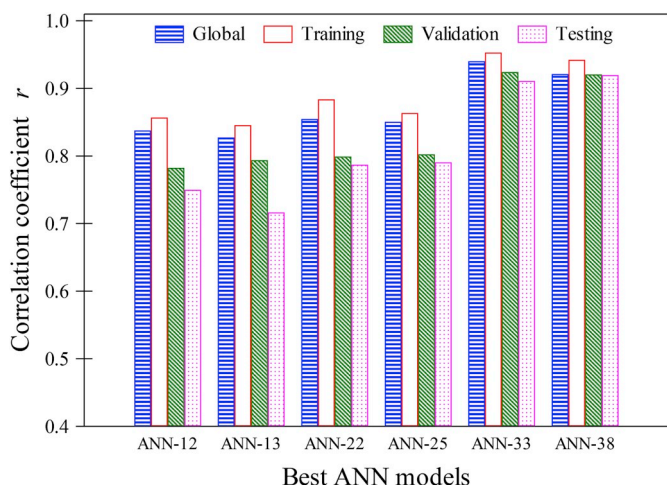


Fig. 9. Statistical comparative analysis of the linear correlation coefficients obtained from the ANN architectures tested by using the global, training, validation and testing sets.

The prediction capability of the ANN-12 and ANN-13 architectures was evaluated by using a linear regression between  $BHT_{ANN}$  and  $BHT_m$ . These results have been plotted in Fig. 7A and B, respectively. The best linear regression for the ANN-12 was obtained with a global correlation coefficient of  $r = 0.8379$  (Fig. 7A); whereas for the ANN-13, a value of  $r = 0.8273$  was achieved (Fig. 7B). In both figures, the effect of  $BHT_m$  duplicates is schematically represented by means of the dashed-squares.

To analyze the good measure for generalization of the ANN-12 and ANN-13 architectures, a trend analysis among the linear correlation coefficients ( $r$ ) and the number of neurons in the hidden layer was also plotted in Fig. 8A and B, respectively. The effect of the number of neurons in the hidden layer on the correlation coefficients obtained for the global, training, validation and testing data sets was analyzed (see Table 10). As seen in both architectures ANN-12 and ANN-13, the global and training trends clearly exhibited an increase in the correlation coefficients as the number of neurons in the hidden layer was increased. A different trend is observed for the validation and testing sets which was associated to overfitting problems in where the ANN show a good learning from the training data sets but their generalization capacity

decreased for the validation and testing data sets.

The best convergence results in each set of training (0.86 and 0.85), validation (0.78 and 0.79), and testing (0.75 and 0.72) were achieved for the ANN-12 (Fig. 8A) and ANN-13 (Fig. 8B) architectures which were characterized by 13 and 9 neurons for the hidden layer, respectively (see the shadow dashed-squares). However, it was recognized that such results do not totally solve the overfitting due to the differences still observed among the correlation coefficients, which were up to 12% and 15% for the ANN-12 and ANN-13, respectively (Fig. 9). These overfitting problems were probably due to the WG\_SubDB1 which was not filtered through the strict quality control methods given by the conditional statements Q5 to Q8 (see Fig. 3), especially those related to the effect of the duplicated bottomhole temperature data (Q8).

### 3.2. WG\_SubDB2

Following the same methodology of analysis applied to the WG\_SubDB1 and after considering the higher correlation coefficients ( $r$ ) obtained from the training simulations and their corresponding lower RMSE and MAE values, the best thirteen ANN architectures were also selected (see Table 11). These evaluation results similarly included the number of neurons per architecture, and the initial learning rate factors ( $\mu$ ) used for obtaining the best prediction of  $BHT_{ANN}$  using the WG\_SubDB2.

From the prediction results, the ANN-22 [3-34-1] and ANN-25 [4-10-1] were identified as the best ANN architectures for predicting the output  $BHT_{ANN}$  values with good accuracy. The ANN-22 and ANN-25 architectures used initial learning rate factors of  $\mu_2 = 0.001$  and  $\mu_3 = 0.0001$  which lead to obtain lower RMSE (24.91 and 24.93) and MAE (16.60 and 17.20) values, respectively (see Table 11). The ANN-22 [3-34-1] model was characterized by 3 input variables [ $\ln(CO_2/CH_4)$ ,  $\ln(H_2S/CH_4)$ , and  $\ln(H_2/CH_4)$ ] for the input layer, 34 neurons for the hidden layer, and one neuron ( $BHT_m$ ) for the output layer; whereas the ANN-25 [4-10-1] used 4 input variables [ $\ln(CO_2)$ ,  $\ln(H_2S)$ ,  $\ln(CH_4)$ , and  $\ln(H_2)$ ] for the input layer, 10 neurons for the hidden layer, and the same neuron ( $BHT_m$ ) for the output layer (Table 11).

The prediction performance of the ANN-22 and ANN-25 architectures was also evaluated by applying linear regressions between  $BHT_{ANN}$  and  $BHT_m$  (see Fig. 7C and D, respectively). A global correlation coefficient of  $r = 0.85$  was similarly obtained for both ANN-22 and ANN-25

Table 11

Summary of the best ANN architectures evaluated for the prediction of reservoir temperature using the sub-database WG\_SubDB2 ( $q_2 = 498$  samples).

ANN	Comb	Input variables					ANN Architecture	Correlation coefficient ( $r$ )				RMSE global	MAE global
		1	2	3	4	5		global	training	validation	testing		
14	1	$\ln(CO_2)$					[1-33-1] $\mu_1$	0.7035	0.7107	0.6573	0.7053	33.6842	26.3897
15	2	$\ln(H_2S)$					[1-34-1] $\mu_4$	0.7009	0.7227	0.6269	0.6319	33.8475	25.6115
16	3	$\ln(CO_2/$ $H_2)$					[1-16-1] $\mu_2$	0.6484	0.6396	0.6257	0.7345	36.0578	28.7601
17	4	$\ln(H_2S/$ $H_2)$					[1-25-1] $\mu_1$	0.4655	0.4799	0.4633	0.3757	41.9355	32.9863
18	5	$\ln(CH_4/$ $CO_2)$	$\ln(H_2S/$ $H_2)$				[2-34-1] $\mu_3$	0.5898	0.6059	0.5624	0.5436	38.2888	28.2153
19	6	$H_2S/CO_2$	$CH_4/CO_2$	$H_2/CO_2$			[3-27-1] $\mu_2$	0.8110	0.8158	0.7966	0.7984	27.7423	19.9437
20	7	$\ln(H_2S/$ $CO_2)$	$\ln(CH_4/$ $CO_2)$	$\ln(H_2/$ $CO_2)$			[3-18-1] $\mu_4$	0.8244	0.8299	0.7939	0.8260	26.8503	19.0021
21	8	$\ln(CO_2/$ $H_2S)$	$\ln(CH_4/$ $H_2S)$	$\ln(H_2/$ $H_2S)$			[3-6-1] $\mu_3$	0.7886	0.7994	0.7452	0.7462	29.1322	21.2580
22	9	$\ln(CO_2/$ $CH_4)$	$\ln(H_2S/$ $CH_4)$	$\ln(H_2/$ $CH_4)$			[3-34-1] $\mu_{2a}$	<b>0.8546</b>	<b>0.8838</b>	<b>0.7992</b>	<b>0.7872</b>	<b>24.9118</b>	<b>16.6000</b>
23	10	$\ln(CO_2/$ $H_2)$	$\ln(H_2S/$ $H_2)$	$\ln(CH_4/$ $H_2)$			[3-15-1] $\mu_3$	0.8367	0.8532	0.7732	0.7699	26.0062	18.1853
24	11	$ilr_1$	$ilr_2$	$ilr_3$			[3-14-1] $\mu_4$	0.8299	0.8452	0.7605	0.7713	26.4363	18.4547
25	12	$\ln(CO_2)$	$\ln(H_2S)$	$\ln(CH_4)$	$\ln(H_2)$		[4-10-1] $\mu_{3a}$	<b>0.8505</b>	<b>0.8635</b>	<b>0.8024</b>	<b>0.7906</b>	<b>24.9258</b>	<b>17.2014</b>
26	13	$\ln(H_2S/$ $CO_2)$	$\ln(CH_4/$ $CO_2)$	$\ln(H_2/$ $CO_2)$	$\ln$ $(H_2S)$	$\ln(H_2S/$ $H_2)$	[5-5-1] $\mu_2$	0.8087	0.8120	0.7755	0.8153	27.8618	20.5459

<sup>a</sup> These ANN models obtained the best correlation coefficient  $r$  in the three sets (training, validation and testing) for the prediction of BHT by using gas-phase composition of geothermal fluids.  $\mu_1$  (0.01),  $\mu_2$  (0.001),  $\mu_3$  (0.0001), and  $\mu_4$  (0.00001) are the learning rate factors used.

**Table 12**

Summary of the best ANN architectures evaluated for the prediction of reservoir temperature using the sub-database WG\_SubDB<sub>3</sub> ( $q_3 = 97$  samples).

ANN	Comb	Input variables					ANN Architecture	Correlation coefficient ( $r$ )				RMSE global	MAE global
		1	2	3	4	5		global	training	validation	testing		
27	1	$\ln(\text{CO}_2)$					[1-10-1] $\mu^4$	0.7613	0.7664	0.7307	0.8270	23.9468	17.3676
28	2	$\ln(\text{H}_2\text{S})$					[1-14-1] $\mu^3$	0.8103	0.8087	0.8293	0.8190	21.6177	15.6012
29	3	$\ln(\text{CO}_2/$ $\text{H}_2)$					[1-4-1] $\mu^4$	0.5599	0.5607	0.5930	0.5069	30.5086	22.7209
30	4	$\ln(\text{H}_2\text{S}/$ $\text{H}_2)$					[1-8-1] $\mu^2$	0.7086	0.7134	0.7146	0.6777	25.9815	19.7083
31	5	$\ln(\text{CH}_4/$ $\text{CO}_2)$	$\ln(\text{H}_2\text{S}/$ $\text{H}_2)$				[2-9-1] $\mu^2$	0.8419	0.8613	0.8170	0.7326	20.0529	13.4671
32	6	$\text{H}_2\text{S}/\text{CO}_2$	$\text{CH}_4/\text{CO}_2$	$\text{H}_2/\text{CO}_2$			[3-8-1] $\mu^4$	0.8902	0.9004	0.8680	0.8642	16.8836	10.6345
33	7	$\ln(\text{H}_2\text{S}/$ $\text{CO}_2)$	$\ln(\text{CH}_4/$ $\text{CO}_2)$	$\ln(\text{H}_2/$ $\text{CO}_2)$			[3-8-1] $\mu^{2a}$	<b>0.9403</b>	<b>0.9530</b>	<b>0.9243</b>	<b>0.9109</b>	<b>12.5637</b>	<b>8.6370</b>
34	8	$\ln(\text{CO}_2/$ $\text{H}_2\text{S})$	$\ln(\text{CH}_4/$ $\text{H}_2\text{S})$	$\ln(\text{H}_2/$ $\text{H}_2\text{S})$			[3-8-1] $\mu^3$	0.9154	0.9223	0.8983	0.9001	14.8325	10.9406
35	9	$\ln(\text{CO}_2/$ $\text{CH}_4)$	$\ln(\text{H}_2\text{S}/$ $\text{CH}_4)$	$\ln(\text{H}_2/$ $\text{CH}_4)$			[3-11-1] $\mu^4$	0.9239	0.9410	0.9228	0.8640	14.1776	9.6393
36	10	$\ln(\text{CO}_2/$ $\text{H}_2)$	$\ln(\text{H}_2\text{S}/$ $\text{H}_2)$	$\ln(\text{CH}_4/$ $\text{H}_2)$			[3-7-1] $\mu^3$	0.9024	0.9156	0.9067	0.7884	15.8746	11.4503
37	11	$\text{ilr}_1$	$\text{ilr}_2$	$\text{ilr}_3$			[3-9-1] $\mu^1$	0.9017	0.9249	0.9274	0.8222	16.0766	10.7413
38	12	$\ln(\text{CO}_2)$	$\ln(\text{H}_2\text{S})$	$\ln(\text{CH}_4)$	$\ln(\text{H}_2)$		[4-8-1] $\mu^{3a}$	<b>0.9211</b>	<b>0.9422</b>	<b>0.9207</b>	<b>0.9197</b>	<b>14.5553</b>	<b>10.3325</b>
39	13	$\ln(\text{H}_2\text{S}/$ $\text{CO}_2)$	$\ln(\text{CH}_4/$ $\text{CO}_2)$	$\ln(\text{H}_2/$ $\text{CO}_2)$	$\ln$ $(\text{H}_2\text{S})$	$\ln(\text{H}_2\text{S}/$ $\text{H}_2)$	[5-6-1] $\mu^3$	0.9089	0.9150	0.8979	0.8772	15.3676	10.7548

<sup>a</sup> These ANN models obtained the best correlation coefficient  $r$  in the three sets (training, validation and testing) for the prediction of BHT by using gas-phase composition of geothermal fluids.  $\mu_1$  (0.01),  $\mu_2$  (0.001),  $\mu_3$  (0.0001), and  $\mu_4$  (0.00001) are the learning rate factors used.

architectures. The effect of  $BHT_m$  duplicates was also indicated with the shadow dashed-squares drawn in both figures.

To evaluate the good measure for generalization of the ANN-22 and ANN-25 architectures, the same trend analysis among the linear correlation coefficients and the number of neurons in the hidden layer was analyzed in Fig. 8C and D, respectively. As seen in both architectures, the global and training trends exhibited an increase in the correlation coefficients with the number of neurons of the hidden layer.

For the validation and testing results, overfitting problems were still present. For the ANN-22 model, a slight improvement was achieved with a large number of neurons, whereas for the ANN-25, a good approach among the correlation coefficients was achieved for a small number of hidden neurons. The best convergence results in each set of training (0.88 and 0.86), validation (0.80 and 0.80), and testing (0.79 and 0.79) were achieved for the ANN-22 (Fig. 8C) and ANN-25 (Fig. 8D) architectures with 34 and 10 neurons in the hidden layer, respectively (see the shadow dashed-squares). Although these results were marginally improved in relation to those outcomes obtained for the ANN-12 and ANN-13 architectures, it was recognized that some overfitting issues were still present in the training simulations of the ANN-22 and ANN-25. Smaller differences among the correlation coefficients were estimated with values up to 11% and 9% for the ANN-22 and ANN-25, respectively (Fig. 9). The effect of the duplicated bottomhole temperature data ( $Q_8$ ) seems to be also reflected in the ANN training simulations.

### 3.3. WG\_SubDB<sub>3</sub>

After analyzing the same statistical parameters (higher linear correlation coefficients, and lower RMSE and MAE values) from the ANN training simulations using the WG\_SubDB<sub>3</sub>, the best thirteen ANN architectures were selected. These evaluation results are summarized in Table 12 where the number of neurons per architecture and the initial learning rate factors ( $\mu$ ) used for obtaining the best prediction of  $BHT_{ANN}$  values are also reported.

From these outcomes, the ANN-33 [3-8-1] and ANN-38 [4-8-1] resulted as the best ANN architectures to minimize the differences between the  $BHT_{ANN}$  and  $BHT_m$  variables. The ANN-33 and ANN-38 architectures used initial learning rates of  $\mu_2 = 0.001$  and  $\mu_3 = 0.0001$  which lead to achieve the lowest RMSE (12.56 and 14.56) and MAE (8.64 and 10.33) values, respectively (Table 12).

**Table 13**

Results of the sensitivity analysis obtained for the best ANN architectures using the three working sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>).

ANN	Input variables [Relative importance %]				
	1	2	3	4	5
ANN-12	$\ln(\text{CO}_2)$ [26.18%]	$\ln(\text{H}_2\text{S})$ [30.19%]	$\ln(\text{CH}_4)$ [15.37%]	$\ln(\text{H}_2)$ [28.26%]	
ANN-13	$\ln(\text{H}_2\text{S}/$ $\text{CO}_2)$ [25.13%]	$\ln(\text{CH}_4/$ $\text{CO}_2)$ [5.11%]	$\ln(\text{H}_2/$ $\text{CO}_2)$ [10.29%]	$\ln(\text{H}_2\text{S})$ [42.86%]	$\ln(\text{H}_2\text{S}/\text{H}_2)$ [16.61%]
ANN-22	$\ln(\text{CO}_2/$ $\text{CH}_4)$ [33.05%]	$\ln(\text{H}_2\text{S}/$ $\text{CH}_4)$ [34.51%]	$\ln(\text{H}_2/$ $\text{CH}_4)$ [32.44%]		
ANN-25	$\ln(\text{CO}_2)$ [27.83%]	$\ln(\text{H}_2\text{S})$ [12.81%]	$\ln(\text{CH}_4)$ [25.87%]	$\ln(\text{H}_2)$ [33.48%]	
ANN-33	$\ln(\text{H}_2\text{S}/$ $\text{CO}_2)$ [32.42%]	$\ln(\text{CH}_4/$ $\text{CO}_2)$ [31.12%]	$\ln(\text{H}_2/$ $\text{CO}_2)$ [36.47%]		
ANN-38	$\ln(\text{CO}_2)$ [22.23%]	$\ln(\text{H}_2\text{S})$ [19.55%]	$\ln(\text{CH}_4)$ [33.74%]	$\ln(\text{H}_2)$ [24.48%]	

ANN-12 and ANN-13, ANN-22 and ANN-25, and ANN-33 and ANN-38 architectures were developed using by WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>, sub-databases, respectively per pair of ANNs.

The ANN-33 [3-8-1] and ANN-38 [4-8-1] architectures were characterized by containing 3 [ $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ , and  $\ln(\text{H}_2/\text{CO}_2)$ ] and 4 [ $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$ , and  $\ln(\text{H}_2)$ ] input variables for the input layer, respectively. Both ANN architectures used 8 neurons for the hidden layer and one neuron ( $BHT_m$ ) for the output layer (Table 12).

The prediction performance of the ANN-33 and ANN-38 architectures was also assessed by applying linear regressions between  $BHT_{ANN}$  and  $BHT_m$  (see Fig. 7E and F, respectively). The best linear regression for

**Table 14**  
Summary of the descriptive statistics of the external geochemical database Ext\_GDB ( $n_D = 13$ ).

	Central tendency statistical parameters						Dispersion statistical parameters				
	Min	Max	Robust estimators			Gastwirth mean ( $E_L$ ) <sub>G</sub>	Standard deviation <i>s</i>	Robust estimators <sup>a</sup>			
			Mean $\bar{x}$	Median $\bar{x}$	Trimean ( $E_L$ ) <sub>T</sub>			Median absolute deviation MAD	$S_n$	$Q_n$	$\sigma_n$
CO <sub>2</sub>	815	989	900	897	900	890	50	56	50	59	56
H <sub>2</sub> S	3	80	39	49	43	46	25	21	18	23	27
CH <sub>4</sub>	0.8	17.5	4.2	2.9	3	3	4.4	2.5	2.4	2.5	2.5
H <sub>2</sub>	0.5	51	21.8	20.9	20.7	20.8	17.8	23.1	22.5	21.2	19.6
BHT <sub>m</sub> (°C)	220	340	277	270	274	275	36	32	32	35	39

<sup>a</sup> Dispersion robust parameters adjusted with correction factors to convert the sample estimates to population parameters (Verma et al., 2016b, 2017); All gas species are reported in mmol/mol concentration units (dry-basis). The gas analysis samples were collected from some production wells of the Olkaria, Kenya geothermal field (Karingithi and Dissertation, 2002; Ofwona, 2002).

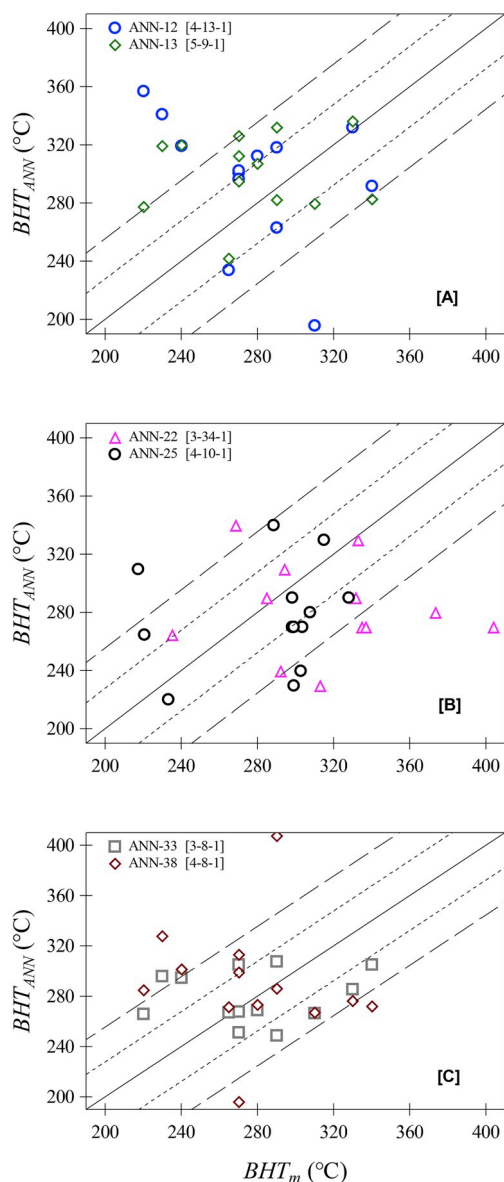
the ANN-33 was obtained with a global correlation coefficient of  $r = 0.9403$  (Fig. 7E); whereas for the ANN-38, a value of  $r = 0.9211$  was achieved (Fig. 7F). In these figures, the effect of BHT<sub>m</sub> duplicates was totally removed, and with this option, the bias and overfitting problems previously detected for the architectures ANN-12, ANN-13, ANN-22, and ANN-25 were eliminated.

To evaluate the good measure for generalization of the ANN-33 and ANN-38 architectures, a similar trend analysis among the linear correlation coefficients and the number of neurons in the hidden layer was carried out (see Fig. 8E and F, respectively). As seen in both architectures ANN-33 and ANN-38, the global and training trends exhibited the highest correlation coefficients in all these simulations. The correlation coefficients obtained for the training, validation and testing sets were nearly the same, which enabled the overfitting problems to be eliminated (see Table 12, and Fig. 8E and F). The best convergence results in each set of training (0.95 and 0.94), validation (0.92 and 0.92), and testing (0.91 and 0.92) were achieved for the ANN-33 (Fig. 8E) and ANN-38 (Fig. 8F) architectures with 8 neurons in the hidden layer (see the shadow dashed-squares in both figures). Smallest differences among the correlation coefficients were achieved with values up to 4% and 2% for the ANN-33 and ANN-38, respectively (Fig. 9). These optimum results were obtained even the small size of the working sub-database used (WG\_SubDB<sub>3</sub>). With the filtering conditions adopted for this sub-database, the effect of the duplicated bottomhole temperatures on the prediction performance of both ANN architectures totally disappeared. According to the overall output results, the ANN-33 and ANN-38 architectures clearly provided a much better prediction than those results inferred from all the ANN previously tested.

**Table 15**  
Results obtained from the application of the best ANN architectures using the external geochemical database Ext\_GDB ( $n_D = 13$ ).

No.	BHT <sub>m</sub> (°C)	BHT <sub>ANN</sub> (°C)						%BHT <sub>diff</sub>					
		ANN-12	ANN-13	ANN-22	ANN-25	ANN-33	ANN-38	ANN-12	ANN-13	ANN-22	ANN-25	ANN-33	ANN-38
1	280	313	307	373	307	269	273	12	10	33	10	-4	-2
2	330	332	337	333	315	286	276	1	2	1	-5	-13	-16
3	270	301	326	335	303	305	313	11	21	24	12	13	16
4	270	302	312	337	298	268	196	12	16	25	10	-1	-27
5	240	320	320	292	303	295	302	33	33	22	26	23	26
6	270	297	295	404	299	252	299	10	9	50	11	-7	11
7	230	341	319	313	299	296	328	48	39	36	30	29	43
8	265	234	242	235	220	*267	*271	-12	-9	-11	-17	-	-
9	310	196	280	294	217	*267	*267	-37	-10	-5	-30	-	-
10	290	263	282	332	298	249	286	-9	-3	14	3	-14	-1
11	290	318	332	285	328	308	408	10	15	-2	13	6	41
12	340	292	283	269	288	305	272	-14	-17	-21	-15	-10	-20
13	220	357	278	179	233	*266	*285	62	26	-19	6	-	-
		$\bar{x} \pm SE$						10 ± 7	10 ± 5	11 ± 6	4 ± 5	2 ± 5	7 ± 8

%BHT<sub>diff</sub> is the error percentage difference between the predicted BHT<sub>ANN</sub> and the measured BHT<sub>m</sub>. SE is the standard error of the mean =  $\left(\frac{s}{\sqrt{n}}\right)$ . \*These values represent estimated values of BHT out of the training interval of ANN-33 and ANN-38 architectures (these values were not considered for the calculation of the % BHT<sub>diff</sub> parameter). ANN-12 and ANN-13, ANN-22 and ANN-25, and ANN-33 and ANN-38 architectures were developed using by WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>, sub-databases, respectively per pair of ANNs.s



**Fig. 10.** Prediction performances of the best ANN architectures for the determination of deep reservoir temperatures by using an external geochemical database (Ext\_GDB).

Ofwona, 2002).

A summary of the descriptive statistics of the Ext\_GDB is reported in Table 14. The  $BHT_m$  measurements logged in this database ranged from 220 °C to 340 °C. The results of deep reservoir temperatures obtained from the best ANN architectures are summarized in Table 15. As a suitable evaluation tool, the percentage difference ( $\%BHT_{diff}$ ) between the  $BHT_{ANN}$  and  $BHT_m$  was estimated for all the gas-phase samples of the Ext\_GDB, together with the standard error of the mean ( $SE$ ). A graphical comparison between  $BHT_{ANN}$  and  $BHT_m$  results is shown in Fig. 10, whereas the reservoir temperature calculations are reported in Table 15. From this evaluation, it can be observed that smaller mean values of the  $\%BHT_{diff}$  ( $4 \pm 5$ ,  $2 \pm 5$  and  $7 \pm 8$ ) were obtained for the ANN-25, ANN-33, and ANN-38 architectures, respectively; whereas larger  $\%BHT_{diff}$  mean values of  $10 \pm 7$ ,  $10 \pm 5$  and  $11 \pm 6$  were achieved by applying the ANN-12, ANN-13 and ANN-22 architectures, respectively.

#### 4. Conclusions

Artificial neural networks were successfully evaluated, for the first

time as a gas geothermometry tool, for the prediction of deep geothermal reservoir temperatures. ANN was specifically used for the multivariate analysis among gas-phase compositional data and  $BHT$  measurements. Up to 455 ANN architectures based on three-layer feed-forward networks were effectively evaluated. As original computational methodology, a comprehensive ANN modeling algorithm was successfully applied using three working geochemical sub-databases (WG\_SubDB<sub>1</sub>, WG\_SubDB<sub>2</sub>, and WG\_SubDB<sub>3</sub>).

From a geochemometric analysis based on the estimation of statistical parameters ( $RMSE$ ,  $MAE$ , linear regressions, and  $\%BHT_{diff}$  and  $SE$ , with an external validation database), six ANN architectures (ANN-12, ANN-13, ANN-22, ANN-25, ANN-33, and ANN-38) resulted as the best prediction tools for deep geothermal reservoir temperatures (with error percentage differences ranging from 2 to 11%). The best ANN architectures (ANN-33 and ANN-38) were characterized by containing 3 [ $\ln(H_2S/CO_2)$ ,  $\ln(CH_4/CO_2)$ , and  $\ln(H_2/CO_2)$ ] and 4 [ $\ln(CO_2)$ ,  $\ln(H_2S)$ ,  $\ln(CH_4)$ , and  $\ln(H_2)$ ] input variables for the input layer, respectively, and the same number of neurons in the hidden and output layers (8 and 1, respectively).

With these results, it was again demonstrated the capability of artificial neural network techniques in solving complicated (multivariate) Earth Science problems. ANN techniques may become an important gas geothermometry tool to be widely applied for the prediction of deep reservoir temperatures of geothermal systems under exploration and exploitation.

#### Computer code availability

Six computer scripts or codes were written in Matlab<sup>®</sup> programming language for the evaluation of all the ANN architectures. The computer scripts are available for downloading from the following public server or repository (<https://github.com/ANNGroup/GasG-Scripts-ANNs.git>); which may be directly accessed by the users. A quick User's Manual for running the ANN simulations is also included, together with a README file, and the open source code license (MIT). Hardware specifications to install and effectively run the Matlab<sup>®</sup> computer scripts are also described in the User's Manual, together with the typical CPU times elapsed in all the ANN simulations.

#### Authorship statement

D. Pérez-Zarate: Computational methodology, programming and numerical simulations; E. Santoyo: Leader of Research, computational methodology (design), fluid geochemistry and geochemometrics; A. Acevedo-Anicasio: Computational algorithms and optimization of ANN; L. Díaz-González: Statistics (univariate and multivariate analysis); C. García-López: Creation of the World Geochemical Database.

#### Acknowledgments

This work was partially funded by the P09 CeMIE-Geo research project (207032 CONACyT-SENER). The authors also express their gratitude to Mirna Guevara and Esther O. García for their valuable comments and assistance in the literature compilation. We also thank Veda Lome for drawing some figures, and Alfredo Quiroz for his technical support to upload the computer codes in the public repository. The second author also wants to thank to the Physics and Mathematics Department of the Universidad Iberoamericana (Mexico) for the Professorship support in a sabbatical leave program carried out on 2018–19. Finally, the authors want to thank to the Editor-in-Chief and the three anonymous reviewers for their comments which helped to improve the final version of the manuscript.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cog.2019.04.001>.

org/10.1016/j.cageo.2019.05.004.

## References

- Álvarez del Castillo, A., Santoyo, E., García-Valladares, O., 2012. A new void fraction correlation inferred from artificial neural networks for modeling two-phase flows in geothermal wells. *Comput. Geosci.* 41, 25–39.
- Arellano, V.M., García, A., Barragán, R.M., Izquierdo, G., Aragón, A., Nieva, D., 2003. An updated conceptual model of the Los Humeros geothermal reservoir (Mexico). *J. Volcanol. Geotherm. Res.* 124, 67–88.
- Arellano, V.M., Torres, M.A., Barragán, R.M., 2005. Thermodynamic Evolution of the Los Azufres, Mexico, Geothermal Reservoir from 1982 to 2002. *Geothermics* 34, 592–616.
- Arellano, V., Barragán, R.M., Aragón, A., Aguilar, K., García, P., 2006. Exploitation Response from the Los Humeros Reservoir. Unpublished Research Project Report IIE/11/13121/02/F. Instituto de Investigaciones Eléctricas, Mexico, 100 pp [in Spanish].
- Arnórsson, S., 1985. Gas pressures in geothermal systems. *Chem. Geol.* 49, 319–328.
- Arnórsson, S., 2000. The quartz- and Na/K geothermometers: I. New thermodynamic calibration. In: *Proceedings of the World Geothermal Congress. Kyushu-Tohoku, Japan*, pp. 929–934.
- Arnórsson, S., D'Amore, F., 2000. Isotopic and Chemical Techniques in Geothermal Exploration Development and Use. Book - International Atomic Energy Agency, Vienna, Austria, 351 pp.
- Arnórsson, S., Gunnlaugsson, E., 1985. New gas geothermometers for geothermal exploration - calibration and application. *Geochem. Cosmochim. Acta* 49, 1307–1325.
- Arnórsson, S., Bjarnason, J.Ö., Giroud, N., Gunnarsson, I., Stefánsson, A., 2006. Sampling and analysis of geothermal fluids. *Geofluids* 6, 203–216.
- Arnórsson, S., Angcoy, E., Bjarnarson, J.O., Giroud, N., Gunnarsson, I., Kasalainen, H., Karingithi, C., Stefánsson, A., 2010. Gas chemistry of volcanic geothermal system. In: *Proceedings of the World Geothermal Congress, Bali, Indonesia*, pp. 1–12.
- Attoh-Okiné, N.O., 1999. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Adv. Eng. Software* 30, 291–302.
- Barragán, R.M., Nieva, D., Santoyo, E., González, E., Verma, M., López-Mendiola, J.M., 1991. Geoquímica de fluidos del campo geotérmico de Los Humeros, Puebla (México). *Geoterminia. Revista Mexicana de Geoenergía* 7, 23–48 [in Spanish].
- Barragán, R.M., Gómez, V.A., Portugal, E., Sandoval, F., Segovia, N., 2005. Gas geochemistry for the los azufres (michoacán) geothermal reservoir, México. *Ann. Geophys.* 48, 145–157.
- Barragán, R.M., Arellano, V.M., Armenta, M., Aguado, R., 2008. Cambios químicos en fluidos de pozos del campo geotérmico de Los Humeros: evidencia de recarga profunda. *Geoterminia Revista Mexicana de Geoenergía* 11, 11–20 [in Spanish].
- Barragán, R.M., Núñez, J., Arellano, V.M., Nieva, D., 2016. EQUILGAS: program to estimate temperatures and in situ two-phase conditions in geothermal reservoirs using three combined FT-HSH gas equilibria models. *Comput. Geosci.* 88, 1–8.
- Bassam, A., Santoyo, E., Andaverde, J., Hernández, J.A., Espinoza-Ojeda, O.M., 2010. Estimation of static formation temperatures in geothermal wells by using an artificial neural network approach. *Comput. Geosci.* 36, 1191–1199.
- Bassam, A., Álvarez del Castillo, A., García-Valladares, O., Santoyo, E., 2015. Determination of pressure drops in flowing geothermal wells by using artificial neural networks and wellbore simulation tools. *Appl. Therm. Eng.* 75, 1217–1228.
- Bayram, A.F., 2001. Applications of an artificial neural network model to a Na-K geothermometer. *J. Volcanol. Geotherm. Res.* 112, 75–81.
- Bertrami, R., Cioni, R., Coraza, E., D'Amore, F., Marini, L., 1985. Carbon monoxide in geothermal gases: reservoir temperature calculations at Larderello (Italy). *Trans. Geoth. Resour. Council.* 9, 299–303.
- Bhanot, K., Peddoju, S.K., Bhardwaj, T., 2018. A model to find optimal percentage of training and testing data for efficient ECG analysis using neural network. *Int. J. Syst. Assur. Eng. Manag.* 9, 12–17.
- Blamey, N.J.F., 2006. H<sub>2</sub>S concentrations in geothermal and hydrothermal fluids – a new gas geothermometer. In: *Proceedings of Thirty-First Workshop Geothermal Reservoir Engineering. Stanford University, CA, USA, SGP-TR-179*, pp. 403–407.
- Can, I., 2002. A new improved Na/K geothermometer y artificial neural networks. *Geothermics* 31, 751–760.
- Cortés, S.O., Urquiza, G., Hernández, J.A., 2009. Optimization of operating conditions for compressor performance by means of neural network inverse. *Appl. Energy* 86, 2487–2493.
- D'Amore, F., Nuti, S., 1977. Notes on the chemistry of geothermal gases. *Geothermics* 6, 39–45.
- D'Amore, F., Panichi, C., 1980. Evaluation of deep temperatures in geothermal systems by a new gas geothermometer. *Geochem. Cosmochimica Acta* 44, 549–556.
- D'Amore, F., Mejía, J.T., 1999. Chemical and physical reservoir parameters at initial conditions in Berlin geothermal field, El Salvador: a first assessment. *Geothermics* 28, 45–73.
- D'Amore, F., Celati, R., Calore, C., Bertrami, R., 1983. Effects of natural recharge on gas composition in the Larderello - castelnuovo area. In: *Proceedings Ninth Workshop Geothermal Reservoir. SCP-TR-74, Stanford, CA, USA*, pp. 221–229.
- Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr.* 25, 80–108.
- Demuth, H., Beale, M., 2003. *Neural Network Toolbox for Use with Matlab-User's Guide Version 4.1*. MathWorks, Inc., USA.
- Díaz-González, L., Santoyo, E., Reyes-Reyes, J., 2008. Three new improved Na/K geothermometers using computational and geochemometric tools: application to the temperature prediction of geothermal systems. *Rev. Mex. Ciencias Geol.* 24, 465–482 [in Spanish].
- Díaz-González, L., Hidalgo-Dávila, C.A., Santoyo, E., Hermsillo-Valadez, J., 2013. Evaluation of training techniques of artificial neural networks for geothermometric studies of geothermal systems. *Rev. Mex. Ingen. Quim.* 12, 105–120 [in Spanish].
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric log ratio transformations for compositional data analysis. *Math. Geol.* 35, 279–300.
- Furuya, S., Aoki, M., Gotoh, H., Takenaka, T., 2000. Takigami geothermal system, northeastern Kyushu, Japan. *Geothermics* 29, 91–211.
- García, A., Santoyo, E., Espinosa, G., Hernández, I., Gutierrez, H., 1998. Estimation of temperatures in geothermal wells during circulation and shut-in in the presence of lost circulation. *Transport Porous Media* 33, 103–127.
- García-López, C.G., Pandarinath, K., Santoyo, E., 2014. Solute and gas geothermometry of geothermal wells: a geochemometrics study for evaluating the effectiveness of geothermometers to predict deep reservoir temperatures. *Int. Geol. Rev.* 56, 2015–2049.
- Garson, G.D., 1991. Interpreting neural-network connection weights. *AI Expet.* 6 (4), 47–51.
- Ghritlahre, H.K., Prasad, R.K., 2018. Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique. *J. Environ. Manag.* 223, 566–575.
- Giggenbach, W.F., 1980. Geothermal gas equilibria. *Geochem. Cosmochim. Acta* 44, 2021–2032.
- Giggenbach, W.F., 1987. Redox processes governing the chemistry of fumarolic gas discharges from White Island, New Zealand. *Appl. Geochem.* 2, 143–161.
- Giggenbach, W.F., 1988. Geothermal solute equilibria. Derivation of Na-K-Mg-Ca geothermometers. *Geochem. Cosmochim. Acta* 52, 2749–2765.
- Giggenbach, W.F., 1991. Chemical techniques in geothermal exploration. In: D'Amore (Ed.), *Application of Geochemistry in Geothermal Reservoir Development. UNITAR/UNDP publication, Rome*, pp. 119–142.
- González-Partida, E., Carrillo-Chávez, A., Levresse, G., Tello-Hinojosa, E., Venegas-Salgado, S., Ramirez-Silva, G., Verma, M.P., Tritlla, J., Camprubi, A., 2005. Hydrogeochemical and isotopic fluid evolution of the Los Azufres geothermal field, Central Mexico. *Appl. Geochem.* 20, 23–39.
- Gudmundsson, B.T., Arnórsson, S., 2002. Geochemical monitoring of the Krafla and Námafjall geothermal areas, N-Iceland. *Geothermics* 31, 195–243.
- Guo, Q., Pang, Z., Wang, Y., Tian, J., 2017. Fluid geochemistry and geothermometry application of the Kangding high-temperature geothermal system in eastern Himalayas. *Appl. Geochem.* 81, 63–75.
- Hagan, M.T., Menhaj, M., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5, 989–993.
- Hagan, M.T., Demuth, H.B., Beale, M.H., 2014. *Neural Network Design*, second ed. PWS Publishing, Boston. 802 pp.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Network.* 2, 359–366.
- Johnson, L.M., Rezaee, R., Kadkodaie, A., Smith, G., Yu, H., 2018. Geochemical property modelling of a potential shale reservoir in the canning basin (western Australia), using artificial neural networks and geostatistical tools. *Comput. Geosci.* 120, 73–81.
- Kacandes, G.H., Grandstaff, D.E., 1989. Differences between geothermal and experimentally derived fluids: how well do hydrothermal experiments model the composition of geothermal reservoir fluids? *Geochem. Cosmochim. Acta* 53, 343–358.
- Karingithi, C.W., 2002. *Hydrothermal Mineral Buffers Controlling Reactive Gases Concentrations in the Greater Olkaria Geothermal System, Kenya*. M.Sc. Dissertation. University of Iceland, Iceland, 51 pp.
- Koga, A., Kita, I., Nitta, K., Hikino, E., Taguchi, S., 1995. New gas geothermometer using CO<sub>2</sub>/H<sub>2</sub> and CH<sub>4</sub>/H<sub>2</sub> ratios. *J. Geotherm. Res. Soc. Jpn.* 17, 201–211.
- López-Mendiola, J.M., Munguía, F., 1989. Geochemical evidences of the boiling phenomenon in the geothermal field of Los Humeros, Puebla, Mexico. *Geoterminia. Revista Mexicana de Geoenergía* 5, 89–106 [in Spanish].
- Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* 11, 431–441.
- MathWorks, 2017. *SIMULINK: Dynamic System Simulation for Matlab v.9.3.0.713579 (R2017b)*. The MathWorks, Inc, Natick, MA. <http://www.mathworks.com/products/Simulink>.
- Nehring, N., D'Amore, F., 1984. Gas chemistry and thermometry of the Cerro Prieto, Mexico, geothermal field. *Geothermics* 13, 75–89.
- Neupane, G., Mattson, E.D., McLing, T.L., Palmer, C.D., Smith, R.W., Wood, T.R., Podgorney, R.K., 2016. Geothermometric evaluation of geothermal resources in southeastern Idaho. *Geotherm. Energy Sci.* 4, 11.
- Nicholson, K., 1993. *Geothermal Fluids: Chemistry and Exploration Techniques*. Springer-Verlag, New York, 263 pp.
- Nieva, D., Fausto, J., González, J., Garibaldi, F., 1982. Flow of vapor into the production zone of Cerro Prieto I wells. In: *Proceedings of the Fourth Symposium on the Cerro Prieto Geothermal Field, vol. 2. Baja California, México*, pp. 455–461.
- Nieva, D., Quijano, L., Garfias, A., Barragán, R.M., Laredo, F., 1983. Heterogeneity of the liquid phase, and vapor separation in Los Azufres (Mexico) geothermal reservoir. In: *Proceedings of the Ninth Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA, USA, SGP-TR-74*, pp. 253–260.
- Nieva, D., Verma, M., Santoyo, E., Barragán, R.M., Portugal, E., 1987. Chemical and isotopic evidence of steam upflow and partial condensation in Los Azufres reservoir. In: *Proceedings of the Twelfth Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA, USA, SGP-TR-109*, pp. 253–260.

- Ofwona, C.A., 2002. A Reservoir Study of Olkaria East Geothermal System, Kenya. M.Sc. Dissertation. University of Iceland, Iceland, 74 pp.
- Opondo, K.M., 2008. The fluid characteristics of three exploration wells drilled at Olkaria-domes field, Kenya. In: Proceedings of the Thirty-Third Workshop on Geothermal Reservoir Engineering Stanford University Stanford, CA, USA, pp. 1–6.
- Pang, Z., 2001. Isotope and chemical geothermometry and its applications. *Sci. China E* 44, 16–20.
- Pang, Z.H., Reed, M., 1998. Theoretical chemical thermometry on geothermal waters: problems and methods. *Geochem. Cosmochim. Acta* 62, 1083–1091.
- Ping, H., Ármannsson, H., 1996. Gas geothermometry in selected Icelandic geothermal fields with comparative examples from Kenya. *Geothermics* 25, 307–347.
- Porkhial, S., Salehpour, M., Ashraf, H., Jamali, A., 2015. Modeling and prediction of geothermal reservoir temperature behavior using evolutionary design of neural networks. *Geothermics* 53, 320–327.
- Portugal, E., Santoyo, E., Verma, S.P., Nieva, D., 1984. Composición química de fluidos geotérmicos del campo de Los Azufres, Mich. In: *Memorias de la Reunión de la Unión Geofísica Mexicana. Baja California Sur, La Paz*, pp. 12–16 de Noviembre, 8pp. [in Spanish].
- Poulton, M.M., 2001. Computational Neural Networks for Geophysical Data Processing, vol. 30. Pergamon Press, Elsevier, Amsterdam, 352 pp.
- Powell, T., Cumming, W., 2010. Spreadsheets for geothermal water and gas geochemistry. In: Proceedings of the Thirty-Fifth Workshop on Geothermal Reservoir Engineering: Stanford University, vol. 188. Stanford, CA, USA, SGP-TR, pp. 1–10.
- Pradeep, T., Srinivasu, P., Avadhani, P.S., Murthy, Y.V.S., 2011. Comparison of variable learning rate and Levenberg-Marquardt back-propagation training algorithms for detecting attacks in intrusion detection systems. *Int. J. Comput. Sci. Eng.* 3, 3572–3581.
- Quijano, J.L., Torres, M., 1995. The Los Humeros geothermal reservoir, a case of very high temperature system. In: Proceedings of the World Geothermal Congress 1995, Florence, Italy, vol. 3, pp. 1569–1573.
- Quijano-León, J.L., 1985. Comportamiento de los gases en la zona sur del campo geotérmico de Los Azufres. *Geotermia. Revista Mexicana de Geoenergía* 1, 45–55 [in Spanish].
- Reed, M., Spycher, N., 1984. Calculation of pH and mineral equilibria in hydrothermal waters with application to geothermometry and studies of boiling and dilution. *Geochem. Cosmochim. Acta* 48, 1479–1492.
- Robles, J., Munguía, F., 1993. Reactivación del sistema geotérmico de Los Humeros, Pue. etapas del metamorfismo. *Geotermia Revista Mexicana de Geoenergía* 9, 193–204.
- Sahoo, S., Jha, M.K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeol. J.* 21, 1865–1887.
- Santoyo, E., Verma, S.P., Nieva, D., Portugal, E., 1991. Variability in the gas phase composition of fluids discharged from Los Azufres geothermal field, Mexico. *J. Volcanol. Geotherm. Res.* 47, 161–181.
- Serban, I.V., Sordoni, A., Bengio, Y., Courville, A.C., Pineau, J., 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. *AAAI* 16, 3776–3784.
- Serpen, G., Palabiyik, Y., Serpen, U., 2010. An artificial neural network model for Na/K geothermometers. In: Proceedings of the World Geothermal Congress. Bali, Indonesia, pp. 1–10.
- Smith, J.S., Wu, B., Wilamowski, B.M., 2018. Neural network training with Levenberg-Marquardt and adaptable weight compression. *EEE Trans. Neural Netw. Learn. Syst.* 99, 1–8.
- Stefánsson, A., Arnórsson, S., 2002. Gas pressures and redox reaction in geothermal fluids in Iceland. *Chem. Geol.* 190, 251–271.
- Stefánsson, A., Arnórsson, S., Gunnarsson, I., Kaasalainen, H., Gunnlaugsson, E., 2011. The geochemistry and sequestration of H<sub>2</sub>S into the geothermal system at Hellisheidi, Iceland. *J. Volcanol. Geotherm. Res.* 202, 179–188.
- Stefánsson, A., Stefánssdóttir, G., Keller, N.S., Barsotti, S., Sigurdsson, A., Thorlákssdóttir, S.B., Peffer, M.A., Eiríksdóttir, E.S., Jónasdóttir, E.B., Löwis, S., Gíslason, S.R., 2017. Major impact of volcanic gases on the chemical composition of precipitation in Iceland during the 2014–2015 Holuhraun eruption. *J. Geophys. Res.: Atmosphere* 122, 1971–1982.
- Supranto, S., Budiarto, T., Djoko, W., Idrus, A., 1996. Proposed empirical gas geothermometer using multidimensional approach. In: Proceedings of the Twenty-First Workshop on Geothermal Reservoir Engineering Stanford University, CA, USA, SGP-TR-151, pp. 195–197.
- Tello, E., 1992. Características geoquímicas e isotópicas de fluidos producidos por pozos del campo geotérmico de Los Humeros, Puebla. *Geotermia. Revista Mexicana de Geoenergía* 8, 3–48 [in Spanish].
- Tello, E., Verma, M.P., Tovar, R., 2000. Origin of acidity in the Los Humeros, Mexico, geothermal reservoir. In: Proceedings World Geothermal Congress, Japan, pp. 2959–2966.
- Trépanier, S., Mathieu, L., Daigneult, R., Faure, S., 2016. Precursors predicted by artificial neural networks for mass balance calculations: quantifying hydrothermal alteration in volcanic rocks. *Comput. Geosci.* 89, 32–43.
- Truesdell, A.H., Nehring, N.L., 1978. Gases and water isotopes in a geochemical section across the Larderello, Italy, geothermal field. *Pure Appl. Geophys.* 117, 276–289.
- Van der Baan, M., Jutten, C., 2000. Neural networks in geophysical applications. *Geophysics* 65, 1032–1047.
- Verma, S.P., Santoyo, E., 1997. New improved equations for Na/K, Na/Li and SiO<sub>2</sub> geothermometers by outlier detection and rejection. *J. Volcanol. Geotherm. Res.* 70, 9–23.
- Verma, S.P., Cruz-Huicochea, R., Díaz-González, L., 2013. Univariate data analysis system: deciphering mean composition of island and continental arc magmas, and influence of the underlying crust. *Int. Geol. Rev.* 55, 1922–1940.
- Verma, S.P., Rivera-Gómez, M.A., Díaz-González, L., Quiroz-Ruiz, A., 2016. Log-ratio transformed major-element based multidimensional classification for altered high-Mg igneous rocks. *Geochem. Geophys. Geosyst.* 17, 1–18.
- Verma, S.P., Díaz-González, L., Pérez-Garza, J.A., Rosales-Rivera, M., 2016. Quality control in geochemistry from a comparison of four central tendency and five dispersion estimators and example of a geochemical reference material. *Arab. J. Geosci.* 9, 1–14. <https://doi.org/10.1007/s1251701627644>. Art. 740.
- Verma, S.P., Díaz-González, L., Pérez-Garza, J.A., Rosales-Rivera, M., 2017. Erratum to: quality control in geochemistry from a comparison of four central tendency and five dispersion estimators and example of a geochemical reference material. *Arab. J. Geosci.* 10, 1–3. <https://doi.org/10.1007/s12517-016-2764-4>. Art. 24.
- Wagh, V., Panaskar, D., Muley, A., Mukate, S., Gaikwad, S., 2018. Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. *Groundwater for Sustainable Development* Available online 8 January 2018 (in press), Corrected Proof. <https://doi.org/10.1016/j.gsd.2017.12.012>.
- Wang, L., Zeng, Y., Chen, T., 2015. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Syst. Appl.* 42, 855–863.
- Yadav, A.M., Chaurasia, R.C., Suresh, N., Gajbhiye, P., 2018. Application of artificial neural networks and response surface methodology approaches for the prediction of oil agglomeration process. *Fuel* 220, 826–836.
- Yan-guang, L., Bing, L., Chuan, L., Xi, Z., Guiling, W., 2017. Reconstruction of deep fluid chemical constituents for estimation of geothermal reservoir temperature using chemical geothermometers. *J. Groundw. Sci. Eng.* 5, 173–181.
- Ye, Z., Kim, M.K., 2018. Predicting electricity consumption in a building using an optimized back-propagation and levenberg-marquardt back-propagation neural network: case study of a shopping mall in China. *Sustain. Cities Soc.* 42, 176–183.

# ANEXO 2

RESEARCH

Open Access



# GaS\_GeoT: A computer program for an effective use of newly improved gas geothermometers in predicting reliable geothermal reservoir temperatures

A. Acevedo-Anicasio<sup>1</sup>, E. Santoyo<sup>2\*</sup> , D. Pérez-Zárte<sup>3</sup>, Kailasa Pandarinath<sup>2</sup>, M. Guevara<sup>2</sup> and L. Díaz-González<sup>4</sup>

\*Correspondence:

esg@ier.unam.mx

<sup>2</sup> Instituto de Energías

Renovables, Universidad

Nacional Autónoma de

México, Priv. Xochicalco S/N,

Temixco, 62580 Morelos,

México

Full list of author information

is available at the end of the article

## Abstract

A geochemometric study based on a multi-criteria decision analysis was applied, for the first time, for the optimal evaluation and selection of artificial neural networks, and the prediction of geothermal reservoir temperatures. Eight new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) were derived from this study. For an effective and practical application of these geothermometers, a new computer program GaS\_GeoT was developed. The prediction efficiency of the new geothermometers was compared with temperature estimates inferred from twenty-five existing geothermometers using gas-phase compositions of fluids from liquid- (LIQDR) and vapour-dominated (VAPDR) reservoirs. After applying evaluation statistical metrics (DIFF%, RMSE, MAE, MAPE, and the Theil's U test) to the temperature estimates obtained by using all the geothermometers, the following inferences were accomplished: (1) the new eight gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) provided reliable and systematic temperature estimates with performance wise occupying the first eight positions for LIQDR; (2) the GasG<sub>3</sub> and GasG<sub>1</sub> geothermometers exhibited consistency as the best predictor models by occupying the first two positions over all the geothermometers for VAPDR; (3) the GasG<sub>3</sub> geothermometer exhibited a wider applicability, and a better prediction efficiency over all geothermometers in terms of a large number of samples used (up to 96% and 85% for LIQDR and VAPDR, respectively), and showed the smallest differences between predicted and measured temperatures in VAPDR and LIQDR; and lastly (4) for the VAPDR, the existing geothermometers ND84c, A98c, and ND98b sometimes showed a better prediction than some of the new gas geothermometers, except for GasG<sub>3</sub> and GasG<sub>1</sub>. These results indicate that the new gas geothermometers may have the potential to become one of the most preferred tools for the estimation of the reservoir temperatures in geothermal systems.

**Keywords:** Geothermal energy, Renewable energy, Geochemometrics, Fluid geochemistry, Artificial intelligence



## Introduction

Geothermal energy has emerged as a clean alternative source of renewable energy for electric power generation and other direct uses (Wu and Li 2020). The enormous amount of energy stored in geothermal systems makes them an important renewable and sustainable energy source (Nieva et al. 2018; Gutiérrez-Negrín et al. 2020). Among the geothermal systems today exploited, hydrothermal systems stand out for their storage capability of hot fluids, which are used for the generation of electricity (Thien et al. 2015). The chemical composition of liquid and steam (gas) phases of geothermal fluids provides useful information on hydrogeological processes, thermal and recharge conditions of reservoirs, and underground flow patterns (Nicholson 1993). Within these applications, the reliable estimation of reservoir temperatures is a crucial task to evaluate the energy potential of geothermal resources (Gutiérrez-Negrín 2019). To carry out this task, several chemical geothermometers have been proposed for the prediction of deep equilibrium temperatures in geothermal systems (Guo et al. 2017). Chemical geothermometers are low-cost tools used for predicting reservoir temperatures in the early exploration and exploitation stages (Yan-guang et al. 2017). Solute geothermometers are mostly recommended for the prediction of reservoir temperatures in liquid-dominated reservoirs, LIQDR (Verma et al. 2008), whereas gas geothermometers are predominantly suggested for the calculation of reservoir temperatures in vapour-dominated (VAPDR) reservoirs (García-López et al. 2014).

The gas chemistry of geothermal fluids has been applied not only for the estimation of geothermal reservoir temperatures but also to elucidate water–rock interaction processes (Pang, 2001). The direct escape from deep magmatic sources, and the occurrence of gas–gas and gas–mineral reactions may explain the consumption and production of gases in these systems (Minissale et al. 1997).

The equilibria attainment among gas species is considered as a fundamental thermodynamic basis for gas geothermometry. This physicochemical process has been widely studied for the prediction of reservoir temperatures in LIQDR and VAPDR using the composition of vapour (or gas) samples collected from wells and fumaroles. To our knowledge, Ellis (1957) was the first geochemist to point out that gases in magmatic steam might be used to estimate reservoir temperatures. The first gas geothermometer is attributed to Tonani (1973) who found some relationships among geothermometric models, and the composition of geothermal gases.

From these studies, numerous gas geothermometers have been proposed based upon the analysis of gas-phase compositions, and the thermodynamic study of gas–gas and gas–mineral equilibria reactions. A comprehensive review on gas geothermometers has not been published yet in indexed journals, although some efforts have been conducted in compiling some of the geothermometers most commonly used (e.g. Henley et al. 1985; Nicholson 1993; Powell 2000; Powell and Cumming 2010). In this work, an update compilation of the most commonly used gas geothermometers is reported in Table 1. Most of these geothermometers may be roughly grouped as follows:

- Simple equations calibrated with databases of gas-phase compositions of fluids collected from geothermal wells;

**Table 1 Updated compilation of gas geothermometers reported in the literature for determining geothermal reservoir temperatures**

No	Gas geothermometer	Acronym	Gas-Mineral equilibria reactions	Gas concentration units	Geothermal reservoir-type <sup>a</sup>	Calculation method	Temperature interval (°C)	References
1	CO <sub>2</sub> -H <sub>2</sub> S-CH <sub>4</sub> -H <sub>2</sub>	DP80	CaSO <sub>4</sub> +FeS <sub>2</sub> +3H <sub>2</sub> O+CO <sub>2</sub> ↔ CaCO <sub>3</sub> +1/3Fe <sub>3</sub> O <sub>4</sub> +3H <sub>2</sub> S+7/3CO <sub>2</sub> C+CO <sub>2</sub> +H <sub>2</sub> ↔ 2CH <sub>4</sub> +2H <sub>2</sub> O	% volume	LIQDR and VAPDR	Analytical	145–300	D'Amore and Panichi (1980)
2	H <sub>2</sub> O-CO <sub>2</sub> -CH <sub>4</sub> -H <sub>2</sub>	G80a	CO <sub>2</sub> +4H <sub>2</sub> ↔ CH <sub>4</sub> +2H <sub>2</sub> O 2NH <sub>3</sub> ↔ N <sub>2</sub> +3H <sub>2</sub>	mmol/mol	LIQDR	Numerical <sup>b</sup>	100–340	Giggenbach (1980)
3	NH <sub>3</sub> -H <sub>2</sub> -N <sub>2</sub>	G80b	CO <sub>2</sub> +4H <sub>2</sub> ↔ CH <sub>4</sub> +2H <sub>2</sub> O 2NH <sub>3</sub> ↔ N <sub>2</sub> +3H <sub>2</sub>	mmol/mol	LIQDR	Numerical <sup>b</sup>	100–340	Giggenbach (1980)
4	CO <sub>2</sub> -CH <sub>4</sub> -H <sub>2</sub>	ND84a	CO <sub>2</sub> +4H <sub>2</sub> ↔ CH <sub>4</sub> +2H <sub>2</sub> O	Mole fraction	LIQDR	Analytical	150–350	Nehring and D'Amore (1984)
5	CO <sub>2</sub> -H <sub>2</sub>	ND84b	H <sub>2</sub> +1/2O <sub>2</sub> ↔ H <sub>2</sub> O C+O <sub>2</sub> ↔ CO <sub>2</sub>	Mole fraction	LIQDR	Analytical	150–350	Nehring and D'Amore (1984)
6	CO <sub>2</sub> -H <sub>2</sub> S	ND84c	3FeS <sub>2</sub> +2H <sub>2</sub> +4H <sub>2</sub> O ↔ Fe <sub>3</sub> O <sub>4</sub> +6H <sub>2</sub> S C+O <sub>2</sub> ↔ CO <sub>2</sub>	Mole fraction	LIQDR	Analytical	150–350	Nehring and D'Amore (1984)
7	CO <sub>2</sub>	AG85a	2clinozoisite+2calcite+3quartz+2H <sub>2</sub> O ↔ 3prehnite+2CO <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
8	CO <sub>2</sub> -H <sub>2</sub>	AG85b	2clinozoisite+2calcite+3quartz+2H <sub>2</sub> O ↔ 3prehnite+2CO <sub>2</sub> 4pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+2pyrite+3H <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
9	CO <sub>2</sub> -H <sub>2</sub>	AG85c	2clinozoisite+2calcite+3quartz+2H <sub>2</sub> O ↔ 3prehnite+2CO <sub>2</sub> 4pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+2pyrite+3H <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
10	H <sub>2</sub>	AG85d	4pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+2pyrite+3H <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
11	H <sub>2</sub>	AG85e	4pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+2pyrite+3H <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
12	H <sub>2</sub> S	AG85f	pyrite+pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+3H <sub>2</sub> S	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
13	H <sub>2</sub> S	AG85g	pyrite+pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+3H <sub>2</sub> S	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
14	H <sub>2</sub> S-H <sub>2</sub>	AG85h	pyrite+pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+3H <sub>2</sub> S 4pyrrhotite+2prehnite+2H <sub>2</sub> O ↔ 2epidote+2pyrite+3H <sub>2</sub>	mmol/kg	LIQDR and VAPDR	Analytical	200–300	Arnósson and Gumlaugsson (1985)
15	CO <sub>2</sub> -CH <sub>4</sub> -CO	B85	CO <sub>2</sub> +H <sub>2</sub> ↔ CO+H <sub>2</sub> O CH <sub>4</sub> +3CO <sub>2</sub> ↔ 4CO+2H <sub>2</sub> O	Mole fraction	LIQDR and VAPDR	Numerical <sup>b</sup>	n.r	Bertrami et al. (1985)
16	CO <sub>2</sub> -N <sub>2</sub>	A87a	n.a	mmol/kg	LIQDR	Analytical	200–300	Arnósson (1987)
17	CO <sub>2</sub> -N <sub>2</sub>	A87b	n.a	mmol/kg	LIQDR	Analytical	200–300	Arnósson (1987)

**Table 1 (continued)**

No	Gas geothermometer	Acronym	Gas-Mineral equilibria reactions	Gas concentration units	Geothermal reservoir-type <sup>a</sup>	Calculation method	Temperature interval (°C)	References
18	CO <sub>2</sub> -H <sub>2</sub> S-CH <sub>4</sub> -H <sub>2</sub> -CO	SD89	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O H <sub>2</sub> + 1/2O <sub>2</sub> ↔ H <sub>2</sub> O H <sub>2</sub> + 1/2S <sub>2</sub> ↔ H <sub>2</sub> S H <sub>2</sub> + CO <sub>2</sub> ↔ CO + H <sub>2</sub> O 1/3Fe <sub>3</sub> O <sub>4</sub> + S <sub>2</sub> ↔ FeS <sub>2</sub> + 2/3O <sub>2</sub>	% mol	VAPDR	Numerical <sup>b</sup>	140–370	Saracco and D'Amore (1989)
19	CO <sub>2</sub> -CH <sub>4</sub>	G91a	3CO <sub>2</sub> + CH <sub>4</sub> ↔ 4CO + 2H <sub>2</sub> O	mmol/mol	LIQDR	Analytical	100–350	Giggenbach (1991)
20	H <sub>2</sub> -Ar	G91b	2Fe <sub>3</sub> O <sub>4</sub> + H <sub>2</sub> O ↔ 3Fe <sub>2</sub> O <sub>3</sub> + H <sub>2</sub> + O <sub>2</sub>	mmol/mol	LIQDR	Analytical <sup>b</sup>	100–350	Giggenbach (1991)
21	CO <sub>2</sub> -Ar	G91c	n.a	mmol/mol	LIQDR	Analytical <sup>b</sup>	100–350	Giggenbach (1991)
22	H <sub>2</sub> -Ar	GG92a	FeS <sub>2</sub> + FeO + 2H <sub>2</sub> O ↔ Fe <sub>2</sub> O <sub>3</sub> + 2H <sub>2</sub> S	mmol/mol	LIQDR	Analytical <sup>b</sup>	100–350	Giggenbach and Glover (1992)
23	CO <sub>2</sub> -CO	GG92b	FeS <sub>2</sub> + FeO + 2H <sub>2</sub> O ↔ Fe <sub>2</sub> O <sub>3</sub> + 2H <sub>2</sub> S	mmol/mol	LIQDR	Analytical <sup>b</sup>	100–350	Giggenbach and Glover (1992)
24	CO <sub>2</sub> -H <sub>2</sub>	K95a	Empirical geothermometer. Regression approach	% volume	LIQDR	Analytical	n.r	Koga et al. (1995)
25	CH <sub>4</sub> -H <sub>2</sub>	K95b	Empirical geothermometer. Regression approach	% volume	LIQDR	Analytical	n.r	Koga et al. (1995)
26	CO <sub>2</sub> -H <sub>2</sub> S	S96a	Empirical geothermometer. Multidimensional approach	% volume	LIQDR	Analytical	n.r	Supranto et al. (1996)
27	CO <sub>2</sub> -H <sub>2</sub> S-CH <sub>4</sub>	S96b	Empirical geothermometer. Multidimensional approach	% volume	LIQDR	Analytical	n.r	Supranto et al. (1996)
28	CO <sub>2</sub>	A98a	2clinozoisite + 2calcite + 3quartz + 2H <sub>2</sub> O ↔ 3prehnite + 2CO <sub>2</sub>	mmol/kg	LIQDR	Analytical	> 230	Arnórisson et al. (1998)
29	CO <sub>2</sub>	A98b	2clinozoisite + 2calcite + 3quartz + 2H <sub>2</sub> O ↔ 3prehnite + 2CO <sub>2</sub>	mmol/kg	LIQDR	Analytical	> 230	Arnórisson et al. (1998)
30	H <sub>2</sub> S	A98c	pyrite + pyrrhotite + 2prehnite + 2H <sub>2</sub> O ↔ 2epidote + 3H <sub>2</sub> S	mmol/kg	LIQDR	Analytical	> 150	Arnórisson et al. (1998)
31	H <sub>2</sub>	A98d	4pyrrhotite + 2prehnite + 2H <sub>2</sub> O ↔ 2epidote + 2pyrite + 3H <sub>2</sub>	mmol/kg	LIQDR	Analytical	> 150	Arnórisson et al. (1998)
32	CO <sub>2</sub> -N <sub>2</sub>	A98e	n.r	mmol/kg	LIQDR	Analytical	n.r	Arnórisson et al. (1998)
33	H <sub>2</sub> -Ar	A98f	n.r	mmol/kg	LIQDR	Analytical <sup>b</sup>	n.r	Arnórisson et al. (1998)
34	H <sub>2</sub> -Ar	A98g	n.r	mmol/kg	LIQDR	Analytical <sup>b</sup>	n.r	Arnórisson et al. (1998)
35	H <sub>2</sub> S	B06	2Fe <sub>3</sub> O <sub>4</sub> + 12H <sub>2</sub> S + CO <sub>2</sub> ↔ 6FeS <sub>2</sub> + 10H <sub>2</sub> O + CH <sub>4</sub>	% mol	LIQDR	Analytical	200–300	Blamey (2006)
36	CO <sub>2</sub> -H <sub>2</sub> -Ar	PC10a	2Fe <sub>3</sub> O <sub>4</sub> + H <sub>2</sub> O ↔ 3Fe <sub>2</sub> O <sub>3</sub> + H <sub>2</sub> + O <sub>2</sub> 2H <sub>2</sub> O + C ↔ 2H <sub>2</sub> + CO <sub>2</sub>	% mol	LIQDR and VAPDR	Grid <sup>b</sup>	100–350	Powell and Cumming (2010)
37	CO <sub>2</sub> -CO-CH <sub>4</sub>	PC10b	CaCO <sub>3</sub> + K-mica ↔ CaAl <sub>2</sub> -silicate + K-feldspar + CO <sub>2</sub> CO <sub>2</sub> + H <sub>2</sub> O ↔ CO + H <sub>2</sub> O 3CO <sub>2</sub> + CH <sub>4</sub> ↔ 4CO + 2H <sub>2</sub> O	% mol	Liquid—and VAPDR	Grid <sup>b</sup>	100–350	Powell and Cumming (2010)

**Table 1 (continued)**

No	Gas geothermometer	Acronym	Gas-Mineral equilibria reactions	Gas concentration units	Geothermal reservoir-type <sup>a</sup>	Calculation method	Temperature interval (°C)	References
38	CO <sub>2</sub> -CH <sub>4</sub> -H <sub>2</sub> -H <sub>2</sub> S	PC10c	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O 3FeS <sub>2</sub> + 2H <sub>2</sub> + 4H <sub>2</sub> O ↔ Fe <sub>3</sub> O <sub>4</sub> + 6H <sub>2</sub> S FeS <sub>2</sub> + H <sub>2</sub> ↔ FeS + H <sub>2</sub> S	% mol	LIQDR and VAPDR	Grid <sup>b</sup>	100–350	Powell and Cumming (2010)
39	CO <sub>2</sub> -CH <sub>4</sub> -H <sub>2</sub>	PC10d	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O CaCO <sub>3</sub> + K-mica ↔ CaAl <sub>2</sub> -silicate + K-feldspar + CO <sub>2</sub>	% mol	LIQDR and VAPDR	Grid <sup>b</sup>	100–350	Powell and Cumming (2010)
40	CO <sub>2</sub> -CH <sub>4</sub> -H <sub>2</sub> -H <sub>2</sub> S	PC10e	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O 3FeS <sub>2</sub> + 2H <sub>2</sub> + 4H <sub>2</sub> O ↔ Fe <sub>3</sub> O <sub>4</sub> + 6H <sub>2</sub> S FeS <sub>2</sub> + H <sub>2</sub> ↔ FeS + H <sub>2</sub> S	% mol	LIQDR and VAPDR	Grid <sup>b</sup>	200–350	Powell and Cumming (2010)
41	CO <sub>2</sub> -H <sub>2</sub> S-H <sub>2</sub>	B16a	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O H <sub>2</sub> + 2H <sub>2</sub> O + 3/2FeS <sub>2</sub> ↔ 3H <sub>2</sub> S + 1/2Fe <sub>3</sub> O <sub>4</sub>	% volume	LIQDR	Numerical <sup>b</sup>	125–350	Barragán et al. (2016)
42	CO <sub>2</sub> -H <sub>2</sub> S-H <sub>2</sub>	B16b	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O 5/4H <sub>2</sub> + 3/4Fe <sub>2</sub> O <sub>3</sub> + 3/2FeS <sub>2</sub> + 7/4H <sub>2</sub> O ↔ 3H <sub>2</sub> S + Fe <sub>3</sub> O <sub>4</sub>	% volume	LIQDR	Numerical <sup>b</sup>	150–350	Barragán et al. (2016)
43	CO <sub>2</sub> -H <sub>2</sub> S-H <sub>2</sub>	B16c	CO <sub>2</sub> + 4H <sub>2</sub> ↔ CH <sub>4</sub> + 2H <sub>2</sub> O H <sub>2</sub> + FeS <sub>2</sub> ↔ H <sub>2</sub> S + FeS	% volume	LIQDR	Numerical <sup>b</sup>	100–350	Barragán et al. (2016)

<sup>a</sup> Type of geothermal reservoir used to calibrate the gas geothermometers

<sup>b</sup> Gas geothermometers that were not considered in this work

n.r. = not reported; n.a. = not available. Analytical method refers to a gas geothermometer that provides a direct mathematical function for the calculation of the reservoir temperatures, i.e., BHT = f(gas concentration); whereas Numerical method refers to a more complex function that correlate multiple variables (i.e., gas concentration, pressure, the fraction of water, temperature, among others); Grid geothermometric method refers to a gas geothermometer that provides a complex grid-numerical algorithm for the estimation of temperatures and other key parameters of the reservoir (i.e., steam fraction, distribution coefficients, and the steam/gas ratio)

**Table 2 A listing of software reported in the technical literature for computing geothermal reservoir temperatures**

No	Open-source computer programs	Computational platform/language	Number of gas geothermometers	Working temperature interval (°C)	Gas concentration units	Geothermal reservoir-type <sup>a</sup>	References
1	CO2B	Fortran	1	140–370	% mol	VAPDR	Saracco and D'Amore (1989)
2	Spreadsheet application	Excel	5	100–350	% mol	LIQDR and VAPDR	Powell and Cumming (2010)
3	GasGeo	Visual Basic	21	100–350	mmol/kg; mole fraction; mmol/mol; %mol; %vol	LIQDR and VAPDR	Pandarinath et al. (2011)
4	EQUILGAS	Visual Fortran	3	100–350	% vol	LIQDR	Barragán et al. (2016)
5	GasGeoPlus	Fortran	21	100–350	mmol/kg; mole fraction; mmol/mol; %mol; %vol	LIQDR and VAPDR	Pérez-Zárate et al. (2019)
6	SYS_GASCHEM	Perl	21	100–350	mmol/kg; mole fraction; mmol/mol; %mol; %vol	LIQDR and VAPDR	García-Mandujano (2019)
7	GaS_GeoT	Java	8	170–374	mmol/mol	LIQDR and VAPDR	This work

<sup>a</sup> Type of geothermal reservoir used to calibrate the gas geothermometer

- Complex grid-numerical geothermometers that calculate reservoir temperatures and other key parameters (e.g. the steam excess, the distribution coefficients of gases between liquid and steam, etc.); and
- Geothermometric equations derived from fluid–rock interaction experiments.

Although these geothermometers have been proposed for the prediction of geothermal temperatures, their generalized application has been limited by the following issues:

1. The significant statistical differences among the temperature estimates predicted by a group of gas geothermometers;
2. The scarcity of the gas-phase compositions at lower temperature levels (between 90 and 150 °C) which has hindered the proposal of new improved equations;
3. The limited intervals of the gas composition and temperature for the application of the geothermometric equations;
4. The scarcity of geochemometric studies for the evaluation of the temperature estimates and their uncertainties; and
5. The lack of practical computer programs to estimate reservoir temperatures by using a wide variety of concentration units, and the complicated calculations involved in some equations.

To address the referred issues (1–3), geochemical databases with representative compositions of gas and/or steam phases from LIQDR and VAPDR are required.

Considering the complex nature of the gas–mineral equilibria processes, it is also necessary to explore new regression tools based on artificial intelligence techniques for calibrating multivariate gas geothermometers, and so, to predict reservoir temperatures with a better accuracy. Among these techniques, the artificial neural networks (ANN) have been used in solving multivariate problems in Earth sciences (Poulton, 2001). For example, in geothermal studies, the use of ANN has been applied for (i) the development of Na–K geothermometers (Díaz-González et al. 2008; and Serpen et al. 2009); (ii) the prediction of mass and heat transport in geothermal wells (Bassam et al. 2010; Álvarez del Castillo et al. 2012; Porkhial et al. 2015); and (iii) the optimization of geothermal power plants (Arslan and Yetik 2011), among others.

One of the latest ANN applications reported for geothermal studies was conducted by Pérez-Zárate et al. (2019) who performed a preliminary study to evaluate ANNs for the prediction of geothermal temperatures using gas-phase compositions from which the present research work has been comprehensively completed.

To evaluate the uncertainties of gas geothermometers (the issue referred as 4), a limited number of geochemometric studies have been reported (e.g. D'Amore and Panichi 1980; Arnórsson et al. 2006; García-López et al. 2014). These studies stated that the prediction efficiency of gas geothermometers is affected by several error sources, such as gas sampling errors, analytical errors, coefficient errors, and the total propagated errors, associated with the calculation of temperatures (e.g. Kacandes and Grandstaff 1989; García-López et al. 2014).

The availability of new computer programs to calculate reservoir temperatures (the issue referred as 5) still constitutes a current necessity for the geochemometric studies. Computer programs to apply solute geothermometers in estimating temperatures are widely available for the study of geothermal systems (e.g. Verma et al. 2008; Spycher et al. 2016), whereas for the gas geothermometry, these programs are rarely shared in the literature. A short listing of open-source programs commonly used in gas geothermometry is included in Table 2. To fulfil the constraints (1–5), the development of new improved gas geothermometers and computer programs are still claimed by the geothermal industry.

To address a reliable prediction of geothermal reservoir temperatures, new improved gas geothermometers and a practical computer program called GaS\_GeoT have been developed in this work. GaS\_GeoT was calibrated for the reliable prediction of reservoir temperatures in LIQDR and VAPDR for a temperature interval between 170 and 374 °C.

To evaluate the prediction efficiency of the new gas geothermometers in geothermal wells, an updated worldwide geochemical database containing gas-phase compositions of fluids were compiled. The prediction efficiency of these geothermometers was compared against twenty-five existing gas geothermometers (listed in Table 3). Details of this geochemometric study are also outlined in this work.

### Work methodology

A comprehensive computational methodology was developed to achieve the following research objectives (Fig. 1a, b): (i) to evaluate ANNs based on a Multi-Criteria Decision Analysis (MCDA) for selecting optimal prediction models that enable new improved gas

**Table 3 Existing gas geothermometers reported in the geothermal literature which were used in the present work**

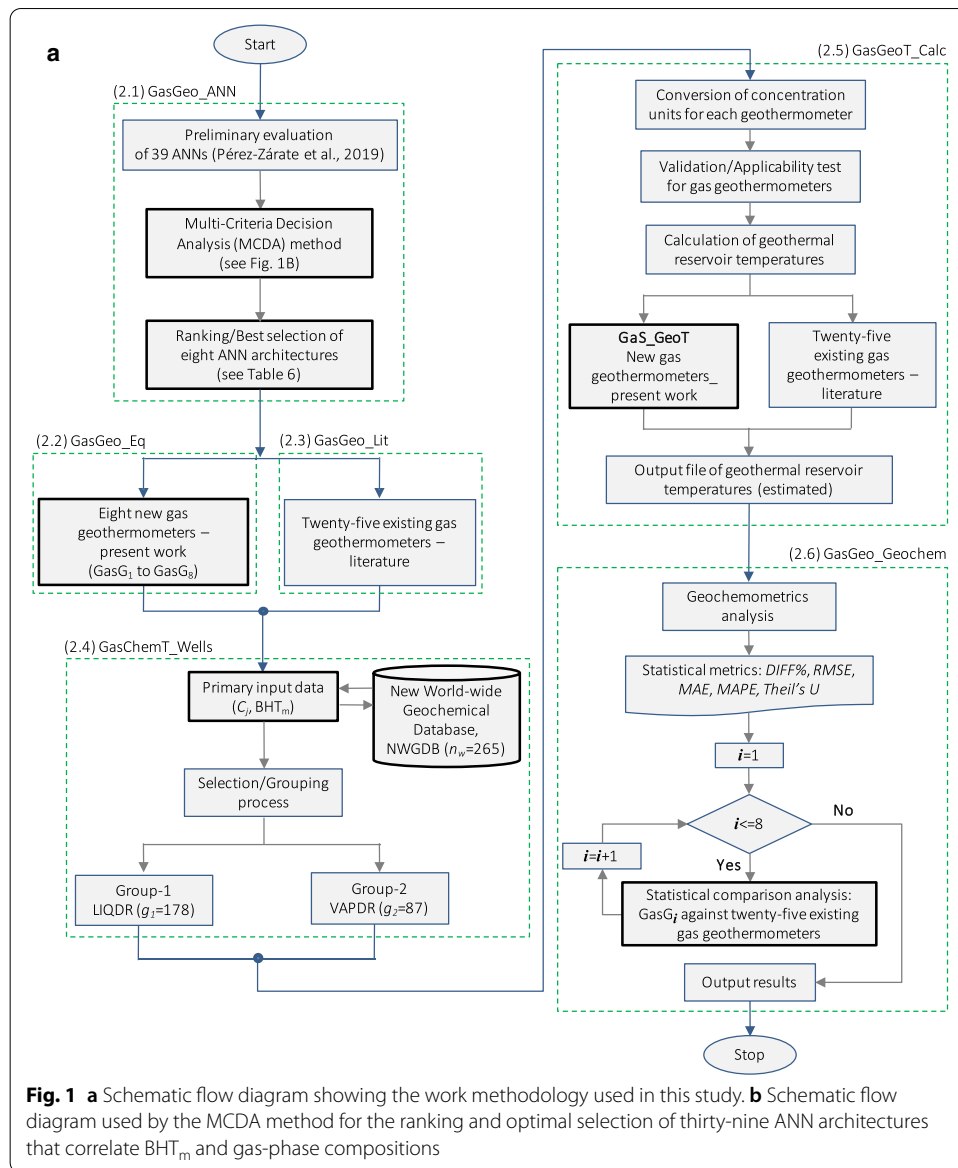
No.	Acronym	Temperature function (°C)	Gas concentration units	References
1	DP80	If %CO <sub>2</sub> ≥ 75 $\frac{24775}{2 \log \left( \frac{CH_4}{CO_2} \right) - 6 \log \left( \frac{H_2}{CO_2} \right) - 3 \log \left( \frac{HS}{CO_2} \right) + 36.05} - 273.15$ If %CO <sub>2</sub> < 75 $\frac{24775}{2 \log \left( \frac{CH_4}{CO_2} \right) - 6 \log \left( \frac{H_2}{CO_2} \right) - 3 \log \left( \frac{HS}{CO_2} \right) - 7 \log(0.1) + 36.05} - 273.15$	% volume	D'Amore and Panichi (1980)
2	ND84a	$\sqrt{\frac{(5.5697 - \beta) + (R - 5.5697)^2 + 1056.34368}{0.039396}} - 273.15$	Mole fraction	Nehring and D'Amore (1984)
3	ND84b	$R = \log(X_{CH_4}) - 4 \log(X_{H_2}) - \log(X_{CO_2})$	Mole fraction	Nehring and D'Amore (1984)
4	ND84c	$\sqrt{\frac{(5 + 1.89499) + (S + 1.89499)^2 + 128.90864}{0.01432}} - 273.15$ $S = \log(X_{H_2}) + \frac{1}{2} \log(X_{CO_2})$ $5.78356V^2 + 110.212V + 662.814$ $V = \log(X_{H_2S}) + \frac{1}{6} \log(X_{CO_2})$	Mole fraction	Nehring and D'Amore (1984)
5	AG85a	$-44.1 + 269.25 [\log(CO_2)] - 76.88 [\log(CO_2)]^2 + 9.52 [\log(CO_2)]^3$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
6	AG85b	$341.7 - 28.57 \log \left( \frac{CO_2}{H_2} \right)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
7	AG85c	$311.7 - 66.72 \log \left( \frac{CO_2}{H_2} \right)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
8	AG85d	$277.2 + 20.99 \log(H_2)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
9	AG85e	$212.2 + 38.59 \log(H_2)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
10	AG85f	$246.7 + 44 \log(H_2S)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
11	AG85g	$173.2 + 65.04 \log(H_2S)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
12	AG85h	$304.1 - 39.48 \log \left( \frac{H_2S}{H_2} \right)$	mmol/kg	Arnórsón and Gunnlaugsson (1985)
13	A87a	$148.5 + 64.35 \log \left( \frac{CO_2}{N_2} \right) + 5.239 \left[ \log \left( \frac{CO_2}{N_2} \right) \right]^2 - 1.832 \left[ \log \left( \frac{CO_2}{N_2} \right) \right]^3$	mmol/kg	Arnórsón (1987)
14	A87b	$135.9 + 63.14 \log \left( \frac{CO_2}{N_2} \right) + 6.241 \left[ \log \left( \frac{CO_2}{N_2} \right) \right]^2 - 1.813 \left[ \log \left( \frac{CO_2}{N_2} \right) \right]^3$	mmol/kg	Arnórsón (1987)
15	G91a	$\frac{4625}{10.4 + \log \left( \frac{CH_4}{CO_2} \right)} - 273.15$	mmol/mol	Giggenbach (1991)

**Table 3 (continued)**

No.	Acronym	Temperature function (°C)	Gas concentration units	References
16	K95a	$352.45 - 53.36 \log \left( \frac{CO_2}{H_2} \right)$	% volume	Koga et al. (1995)
17	K95b	$228.95 - 24.06 \left( \frac{CH_4}{H_2} \right)$	% volume	Koga et al. (1995)
18	S96a	$\frac{738.3091}{(\%CO_2^{0.0008})(\%H_2S)^{0.1037}} - 273.15$	% volume	Supranto et al. (1996)
19	S96b	$\frac{1138.9501(\%Gas)^{0.0037}(\%H_2S)^{0.1076}}{(\%CO_2^{0.2019})(\%CH_4)^{0.0056}} - 273.15$	% volume	Supranto et al. (1996)
20	A98a	$121.8 + 72.012 \log(CO_2) + 11.068 [\log(CO_2)]^2 - 4.724 [\log(CO_2)]^3$	mmol/kg	Arnórsson et al. (1998)
21	A98b	$129.2 + 88.299 \log(CO_2) + 21.946 [\log(CO_2)]^2 - 10.103 [\log(CO_2)]^3$	mmol/kg	Arnórsson et al. (1998)
22	A98c	$177.6 + 66.152 \log(H_2S) + 4.811 [\log(H_2S)]^2$	mmol/kg	Arnórsson et al. (1998)
23	A98d	$227.1 + 56.168 \log(H_2) + 5.836 [\log(H_2)]^2 + 6.630 [\log(H_2)]^3$	mmol/kg	Arnórsson et al. (1998)
24	A98e	$173.2 + 48.751 \log \left( \frac{CO_2}{H_2} \right) + 7.599 \left[ \log \left( \frac{CO_2}{H_2} \right) \right]^2 + 1.793 \left[ \log \left( \frac{CO_2}{H_2} \right) \right]^3$	mmol/kg	Arnórsson et al. (1998)
25	B06	For H <sub>2</sub> S concentrations ≥ 0.00175 For H <sub>2</sub> S concentrations < 0.00175 $64.68 \log(H_2S) + 427.8$	%mol	Blamey (2006)

All the gas geothermometers use concentration units in wet-basis, except DP80, G91a, S96a, S96b, and B06 which use dry-basis

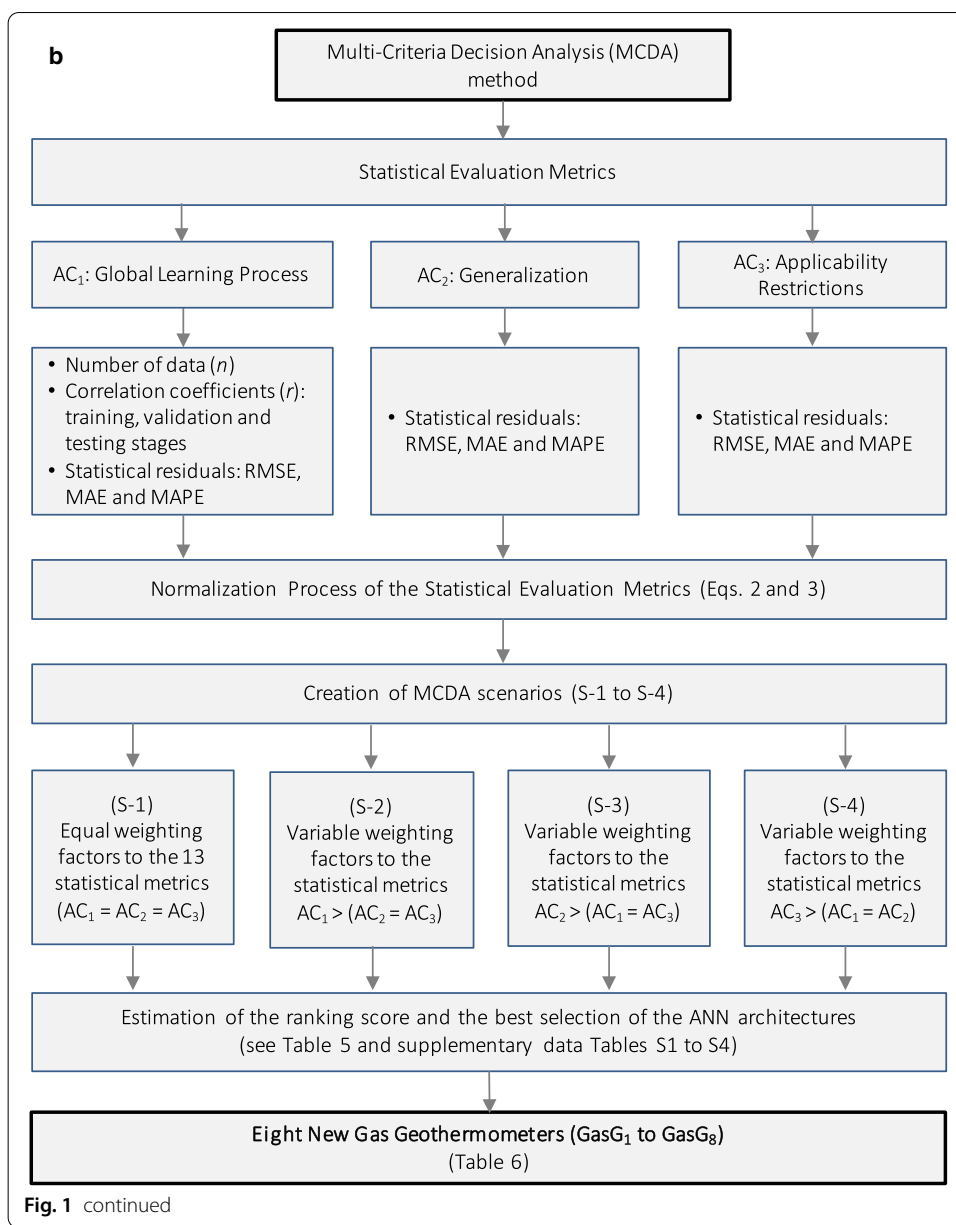




geothermometers to be developed; (ii) to create a new computer program for the effective use of the new gas geothermometers; (iii) to evaluate the prediction efficiency of the new gas geothermometers using the gas-phase compositions from well samples collected in LIQDR and VAPDR; and (iv) to compare the temperature estimates inferred from new gas geothermometers with those temperatures predicted by some existing geothermometers. In order to address these research goals, six computational modules were structured:

*GasGeo\_ANN*: To carry out the optimal evaluation and selection of ANN architectures by using a novel application of the MCDA method;

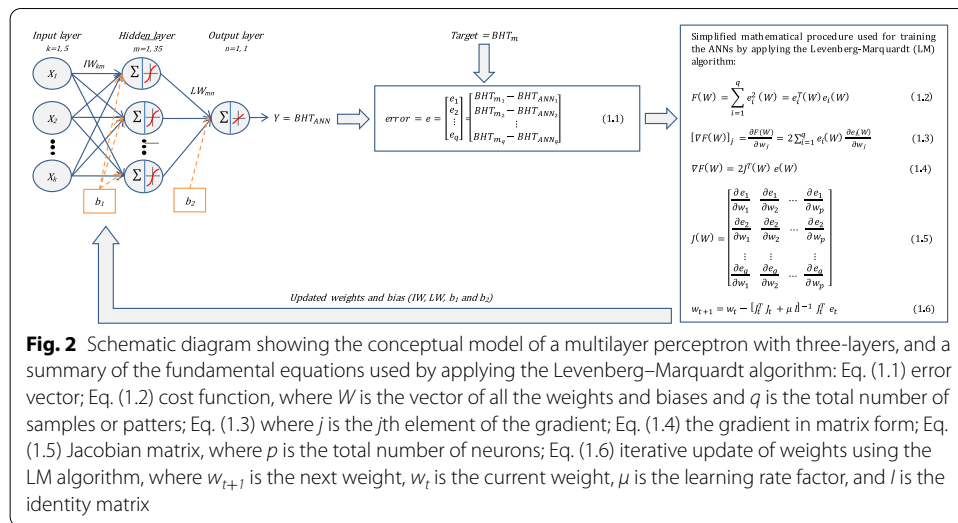
*GasGeo\_Eq*: To describe the new gas geothermometer equations developed and their applicability conditions;



*GasGeo\_Lit*: To select twenty-five existing gas geothermometers for carrying out statistical comparison analyses;

*GasChemT\_Wells*: To compile a new Worldwide Geochemical Database (NWGDB) containing gas-phase compositions, and bottom-hole temperatures (BHT<sub>m</sub>) measured in geothermal wells;

*GasGeoT\_Calc*: To describe the calculation of geothermal reservoir temperatures by using the new and existing gas geothermometers, and the operation of the new computer program (GaS\_GeoT); and



*GasGeo\_Geochem*: To perform a comprehensive geochemometric analysis among the temperature estimates (predicted from all the geothermometers) and the bottom-hole temperature measurements ( $BHT_m$ ).

A summary of these computational modules is described as follows:

### GasGeo\_ANN

A previous evaluation study of ANNs to predict geothermal reservoir temperatures was preliminarily conducted by Pérez-Zárate et al. (2019), which constitutes the foundational basis of the present research work. 455 ANN architectures were originally trained to predict reservoir temperatures using three Worldwide Geochemical Subdatabases ( $WG\_SubDB_1$ :  $q_1 = 527$ ;  $WG\_SubDB_2$ :  $q_2 = 498$ ; and  $WG\_SubDB_3$ :  $q_3 = 97$ ) containing gas-phase ( $CO_2$ ,  $H_2S$ ,  $CH_4$ , and  $H_2$ ) compositions of geothermal fluids as input variables. According to Fig. 2, a multilayer perceptron model was used for the design of these ANNs using the conventional matrix notation (Eq. 1) for the determination of the ANN output (or target):

$$Y = BHT_{ANN} = LW \cdot f[IW \cdot X + b_1] + b_2, \tag{1}$$

where  $X$  is the input variable, and  $IW$  and  $b_1$  represent the matrix and the vector of coefficients between the input and hidden layers.  $LW$  and  $b_2$  are the layer weighting coefficients and the vector of coefficients between the hidden and output layers, and  $f$  corresponds to an activation function for the hidden layer.

A full description of the multilayer perceptron model is reported by Pérez-Zárate et al. (2019), and schematically summarized in Fig. 2 by listing the fundamental equations used for adjusting weighting and bias coefficients in the learning process of each ANN layer (referred as Eqs. 1.1–1.6). Such a perceptron model was solved by creating several Matlab numerical scripts that were previously reported (<https://github.com/ANNGroup/GasG-Scripts-ANNs.git>).

For the learning process of ANNs, the well-known Levenberg–Marquardt algorithm (LM), the hyperbolic tangent function, and the linear function were applied. All the ANNs were characterized by an input layer, one hidden layer, and an output layer. For the input and output layers, the data were normalized between -1 and 1, whereas for the hidden layer, the number of neurons varied from 1 to 35. Input data sets were randomly divided into training (80%), validation (10%), and testing (10%). From this study, thirty-nine ANNs were preliminarily reported as the most acceptable prediction models to correlate multivariate relationships among gas-phase compositions and BHT data (referred as ANN-1 to ANN-39 in Pérez-Zárate et al. 2019).

By using traditional evaluation metrics based on small differences and acceptable correlation coefficients between measured ( $BHT_m$ ) and simulated ( $BHT_{ANN}$ ) temperatures, six ANNs were proposed as the '*better prediction models*' (cited as ANN-12, ANN-13, ANN-22, ANN-25, ANN-33, and ANN-38 in the same paper). Although acceptable results were obtained from these earlier prediction models, some applicability problems were later identified in two ANNs that used a small data set for the learning process ( $WG\_SubDB_3$ ;  $q_3 = 97$ , which was used for the ANN-33 and ANN-38 models). The low representativeness of this sub-data set was lastly reflected on very limited intervals of applicability.

A generalization problem was therefore identified in the final development stage of new gas geothermometers, which may affect the future application of these tools in geothermometric studies for the geothermal prospection and exploitation. To correct these problems, a new evaluation methodology based on a novel MCDA method was applied for the optimal selection of the most reliable ANNs among the thirty-nine architectures previously proposed (Fig. 1b). The MCDA method was used, for the first time in the ANN literature, to comprehensively evaluate the efficiency of these ANN prediction models by utilizing the evaluation results derived from the following:

- I. The analysis Case 1 ( $AC_1$ ) described by the global learning processes of three sub-databases ( $WG\_SubDB_1$ ,  $WG\_SubDB_2$ , and  $WG\_SubDB_3$ ) applying the thirty-nine ANNs (pre-selected) and the following statistical metrics: the number of data used by each sub-database ( $q_1 = 527$ ;  $q_2 = 498$  and  $q_3 = 97$ ), the correlation coefficients ( $r$ ) obtained between measured ( $BHT_m$ ) and simulated ( $BHT_{ANN}$ ) temperatures for the training, validation, and testing stages, and the residuals calculated between measured and predicted temperatures (RMSE, MAE, and MAPE);
- II. The analysis Case 2 ( $AC_2$ ) described by the application of the thirty-nine ANNs to the largest sub-database ( $q_1 = 527$ ), as a generalization case, using the same statistical residuals between measured and predicted temperatures (RMSE, MAE, and MAPE); and
- III. The analysis Case 3 ( $AC_3$ ) described by the application of the thirty-nine ANNs to the same sub-database ( $q_1 = 527$ ) but strongly restricted by the applicability conditions of each ANN prediction model, using the same residuals between measured and predicted temperatures.

**Table 4 A summary of the thirty-nine ANNs (proposed by Pérez-Zárate et al., 2019) for determining reservoir temperatures**

ANN	Input variables					ANN architecture	Statistical evaluation metrics												
	Original learning process <sup>a</sup>						Full data <sup>b</sup>					After applicability conditions <sup>c</sup>							
	n data	(r)	(r)	(r)	(r)		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	
WG_SubDB <sub>1</sub> ; q <sub>11</sub> =527																			
1	ln(CO <sub>2</sub> )					[1-32-1]	527	0.69	0.58	0.62	34.83	27.08	10.29	34.83	27.08	10.29	34.83	27.08	10.29
2	ln(H <sub>2</sub> S)					[1-3-1]	527	0.67	0.59	0.67	35.45	27.58	10.46	35.45	27.58	10.46	35.45	27.58	10.46
3	ln(CO <sub>2</sub> /H <sub>2</sub> )					[1-13-1]	527	0.68	0.64	0.55	35.47	27.85	10.31	35.47	27.85	10.31	35.47	27.85	10.31
4	ln(H <sub>2</sub> S/H <sub>2</sub> )					[1-26-1]	527	0.48	0.35	0.36	42.06	32.65	12.27	42.06	32.65	12.27	42.06	32.65	12.27
5	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )				[2-28-1]	527	0.69	0.57	0.31	38.21	27.94	10.44	38.21	27.94	10.44	36.68	27.36	10.25
6	H <sub>2</sub> S/CO <sub>2</sub>	CH <sub>4</sub> /CO <sub>2</sub>	H <sub>2</sub> /CO <sub>2</sub>			[3-12-1]	527	0.80	0.74	0.74	29.26	20.98	7.99	29.26	20.98	7.99	29.12	20.88	7.94
7	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )			[3-18-1]	527	0.85	0.80	0.64	26.84	18.39	6.87	26.84	18.39	6.87	26.46	18.17	6.78
8	ln(CO <sub>2</sub> /H <sub>2</sub> S)	ln(CH <sub>4</sub> /H <sub>2</sub> S)	ln(H <sub>2</sub> /H <sub>2</sub> S)			[3-9-1]	527	0.83	0.80	0.61	28.22	19.45	7.31	28.22	19.45	7.31	28.22	19.45	7.31
9	ln(CO <sub>2</sub> /CH <sub>4</sub> )	ln(H <sub>2</sub> S/CH <sub>4</sub> )	ln(H <sub>2</sub> /CH <sub>4</sub> )			[3-15-1]	527	0.85	0.82	0.63	26.45	18.94	7.14	26.45	18.94	7.14	26.43	18.91	7.12
10	ln(CO <sub>2</sub> /H <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )	ln(CH <sub>4</sub> /H <sub>2</sub> )			[3-9-1]	527	0.83	0.79	0.70	27.24	19.36	7.27	27.24	19.36	7.27	27.24	19.36	7.27
11	ilr <sub>1</sub>	ilr <sub>2</sub>	ilr <sub>3</sub>			[3-9-1]	527	0.83	0.82	0.59	27.94	19.93	7.40	27.94	19.93	7.40	27.92	19.89	7.38
12	ln(CO <sub>2</sub> )	ln(H <sub>2</sub> S)	ln(CH <sub>4</sub> )	ln(H <sub>2</sub> )		[4-13-1]	527	0.86	0.78	0.75	25.77	17.60	6.66	25.76	17.60	6.66	25.74	17.57	6.64
13	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )	ln(H <sub>2</sub> /H <sub>2</sub> )	[5-9-1]	527	0.85	0.79	0.72	26.50	18.32	6.92	26.50	18.32	6.92	26.50	18.32	6.92

**Table 4 (continued)**

ANN	Input variables					ANN architecture	Statistical evaluation metrics												
	Original learning process <sup>a</sup>						Full data <sup>b</sup>						After applicability conditions <sup>c</sup>						
	1	2	3	4	5		n data	(r)	(r)	(r)	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WG_SubDB <sub>2</sub> ; q <sub>2</sub> =498																			
14	ln(CO <sub>2</sub> )					[1-33-1]	498	0.71	0.66	0.71	33.68	26.39	10.10	40.13	30.11	11.37	34.40	26.82	10.19
15	ln(H <sub>2</sub> S)					[1-34-1]	498	0.72	0.63	0.63	33.85	25.61	9.68	35.03	26.64	9.96	35.01	26.52	9.94
16	ln(CO <sub>2</sub> /H <sub>2</sub> )					[1-16-1]	498	0.64	0.63	0.73	36.06	28.76	10.69	37.05	29.63	10.93	36.68	29.17	10.75
17	ln(H <sub>2</sub> S/H <sub>2</sub> )					[1-25-1]	498	0.48	0.46	0.38	41.94	32.99	12.41	42.30	33.26	12.43	42.38	33.33	12.48
18	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )				[2-34-1]	498	0.61	0.56	0.54	38.29	28.22	10.62	39.54	29.38	10.96	38.62	28.46	10.64
19	H <sub>2</sub> S/CO <sub>2</sub>	CH <sub>4</sub> /CO <sub>2</sub>	H <sub>2</sub> /CO <sub>2</sub>			[3-27-1]	498	0.82	0.80	0.80	27.74	19.94	7.64	146.81	41.14	14.80	28.59	20.39	7.74
20	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )			[3-18-1]	498	0.83	0.79	0.83	26.85	19.00	7.27	29.22	20.66	7.78	27.46	19.30	7.29
21	ln(CO <sub>2</sub> /H <sub>2</sub> S)	ln(CH <sub>4</sub> /H <sub>2</sub> S)	ln(H <sub>2</sub> /H <sub>2</sub> S)			[3-6-1]	498	0.80	0.75	0.75	29.13	21.26	8.11	30.64	22.25	8.41	29.90	21.72	8.21
22	ln(CO <sub>2</sub> /CH <sub>4</sub> )	ln(H <sub>2</sub> S/CH <sub>4</sub> )	ln(H <sub>2</sub> /CH <sub>4</sub> )			[3-34-1]	498	0.88	0.80	0.79	24.91	16.60	6.33	48.58	22.76	8.38	23.76	16.33	6.12
23	ln(CO <sub>2</sub> /H <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )	ln(CH <sub>4</sub> /H <sub>2</sub> )			[3-15-1]	498	0.85	0.77	0.77	26.01	18.19	6.84	31.94	20.69	7.68	26.54	18.47	6.87
24	ilr <sub>1</sub>	ilr <sub>2</sub>	ilr <sub>3</sub>			[3-14-1]	498	0.85	0.76	0.77	26.44	18.45	7.00	35.17	21.90	8.12	27.16	18.89	7.10
25	ln(CO <sub>2</sub> )	ln(H <sub>2</sub> S)	ln(CH <sub>4</sub> )	ln(H <sub>2</sub> )		[4-10-1]	498	0.86	0.80	0.79	24.93	17.20	6.47	59.96	22.64	8.55	25.80	17.66	6.58
26	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S)	ln(H <sub>2</sub> /H <sub>2</sub> )	[5-5-1]	498	0.81	0.78	0.82	27.86	20.55	7.81	323.58	69.74	24.67	28.34	20.86	7.87
WG_SubDB <sub>3</sub> ; q <sub>3</sub> =97																			
27	ln(CO <sub>2</sub> )					[1-10-1]	97	0.77	0.73	0.83	23.95	17.37	6.31	126.01	71.75	25.33	44.39	36.45	12.60

**Table 4 (continued)**

ANN	Input variables					ANN architecture	Statistical evaluation metrics												
	Original learning process <sup>a</sup>						Full data <sup>b</sup>						After applicability conditions <sup>c</sup>						
	1	2	3	4	5		n data	(r)	(r)	(r)	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
28	ln(H <sub>2</sub> S)					[1-14-1]	97	0.81	0.83	0.82	21.62	15.60	5.81	57.35	39.24	14.34	41.09	32.97	12.02
29	ln(CO <sub>2</sub> /H <sub>2</sub> )					[1-4-1]	97	0.56	0.59	0.51	30.51	22.72	8.25	49.59	41.80	14.66	50.63	42.93	14.62
30	ln(H <sub>2</sub> S/H <sub>2</sub> )					[1-8-1]	97	0.71	0.71	0.68	25.98	19.71	7.01	52.31	40.83	15.11	49.80	38.88	14.12
31	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )				[2-9-1]	97	0.86	0.82	0.73	20.05	13.47	4.80	107.09	61.29	22.34	52.88	38.80	13.19
32	H <sub>2</sub> S/CO <sub>2</sub>	CH <sub>4</sub> /CO <sub>2</sub>	H <sub>2</sub> /CO <sub>2</sub>			[3-8-1]	97	0.90	0.87	0.86	16.88	10.63	3.86	172.56	67.79	24.89	36.21	26.83	9.45
33	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )			[3-8-1]	97	0.95	0.92	0.91	12.56	8.64	3.08	93.54	57.17	20.69	44.15	31.54	11.23
34	ln(CO <sub>2</sub> /H <sub>2</sub> S)	ln(CH <sub>4</sub> /H <sub>2</sub> S)	ln(H <sub>2</sub> /H <sub>2</sub> S)			[3-8-1]	97	0.92	0.90	0.90	14.83	10.94	3.88	63.15	42.59	16.23	40.54	29.64	10.57
35	ln(CO <sub>2</sub> /CH <sub>4</sub> )	ln(H <sub>2</sub> S/CH <sub>4</sub> )	ln(H <sub>2</sub> /CH <sub>4</sub> )			[3-11-1]	97	0.94	0.92	0.86	14.18	9.64	3.50	124.97	65.67	23.30	50.39	37.10	12.78
36	ln(CO <sub>2</sub> /H <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )	ln(CH <sub>4</sub> /H <sub>2</sub> )			[3-7-1]	97	0.92	0.91	0.79	15.87	11.45	4.14	88.95	53.02	19.17	45.02	33.73	11.70
37	ilr <sub>1</sub>	ilr <sub>2</sub>	ilr <sub>3</sub>			[3-9-1]	97	0.92	0.93	0.82	16.08	10.74	3.78	63.24	40.40	15.03	35.66	25.94	9.08
38	ln(CO <sub>2</sub> )	ln(H <sub>2</sub> S)	ln(CH <sub>4</sub> )	ln(H <sub>2</sub> )		[4-8-1]	97	0.94	0.92	0.92	14.56	10.33	3.68	102.77	64.26	22.86	43.31	32.94	11.29
39	ln(H <sub>2</sub> S/CO <sub>2</sub> )	ln(CH <sub>4</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> /CO <sub>2</sub> )	ln(H <sub>2</sub> S/H <sub>2</sub> )		[5-6-1]	97	0.92	0.90	0.88	15.37	10.75	3.84	66.45	44.07	15.72	39.57	29.27	10.21

<sup>a</sup> The global learning process was reported in Pérez-Zárate et al. (2019): n data (number of samples used in each sub-database); (r) training, validation and testing (the correlation coefficient calculated between the bottom-hole temperatures (measured) and the temperature estimates (predicted) by ANNs; RMSE, MAE and MAPE (residuals calculated between the bottom-hole temperatures and temperature estimates by ANNs)

<sup>b</sup> Full data: RMSE, MAE and MAPE (residuals calculated using the largest representative sub-database: q<sub>1</sub> = 527)

<sup>c</sup> After applicability conditions: RMSE, MAE and MAPE (residuals calculated using the samples that fulfil the applicability conditions of the ANN architectures (reported in Table 8))

To apply the MCDA, it was necessary to normalize the resulting thirteen statistical metrics (Table 4). The normalization was performed either to minimize or maximize the evaluation metrics using the methodology proposed by Dincer and Acar (2015).

For example, in the particular case of the residual errors (RMSE, MAE, and MAPE), the best efficiency of the ANN prediction models would be given when these metrics achieve lower values, whereas for the linear correlation coefficient ( $r$ ), higher values would be desirable. For the statistical residual ( $SR_i$ ), or the minimization obtained during the global learning and application processes of the ANNs, normalized data were calculated by using the RMSE, MAE, and MAPE results and the following equation:

$$\text{Normalized Data}_i = \frac{SR_{\max} - SR_i}{SR_{\max} - SR_{\min}} \times 10 \tag{2}$$

whereas for the maximization, the normalized data was determined by using the following equation:

$$\text{Normalized Data}_i = \frac{SP_i - SP_{\min}}{SP_{\max} - SP_{\min}} \times 10, \tag{3}$$

where  $SP_i$  are the statistical regression parameters:  $n$  (the data number used by each database) and  $r$ ; the correlation coefficients, which were obtained from the training, validation, and testing of the ANNs.

The MCDA is suggested as an optimized method for the integral evaluation of various statistical metrics among predictor models using different scenarios to achieve a specific objective function (Adem and Geneletti 2018). To apply the MCDA, the Multi-Attribute Value Theory (MAVT) algorithm was used (Santoyo-Castelazo 2011; Santoyo-Castelazo et al. 2011; Santoyo-Castelazo and Azapagic 2014; Estévez et al. 2018). This algorithm required the determination of partial value functions, and the estimation of weighting factors for each evaluation metric. The global value function  $V(s)$  which represents the total score to be obtained by any ANN architecture in each scenario  $s$  is calculated as follows:

$$V(s) = \sum_{i=1}^E w_i u(s), \tag{4}$$

where  $w_i$  and  $u(s)$  are the weighting factors used by the evaluation metric, and the value function which provides the metric efficiency for each ANN architecture and scenario, respectively; and  $E$  is the total number of the evaluation metrics.

Wang et al. (2009) suggested the use of variable weighting factors by considering the importance of the scenarios assumed on the predictor model efficiency. To apply the MCDA method, Santoyo-Castelazo and Azapagic (2014) suggested the use of a sensitivity analysis to determine the ranking score of prediction models based on different scenarios that enable indicators (or statistical metrics) to be varied under certain weighting criterion. In this work, the sensitivity analysis was applied for evaluating the prediction efficiency of each ANN model in the calculation of reservoir temperatures using four different scenarios (S), where thirteen statistical metrics were comprehensively analysed by assigning the following weighting criteria:



- (S-1) Equal weighting factors assigned to the statistical metrics computed for the three analysis cases already postulated: AC<sub>1</sub> (global learning processes); AC<sub>2</sub> (the generalization case); and AC<sub>3</sub> (the applicability restrictions imposed by each ANN model), i.e. AC<sub>1</sub> = AC<sub>2</sub> = AC<sub>3</sub>, Fig. 1b; see Additional file 1: Table S1.
- (S-2) Greater weighting factors assigned to the seven statistical metrics computed for the AC<sub>1</sub> (global learning processes) in comparison with equal weighting values assumed for the metrics of the AC<sub>2</sub> and AC<sub>3</sub>, i.e. AC<sub>1</sub> > (AC<sub>2</sub> = AC<sub>3</sub>), Fig. 1b; see Additional file 1: Table S2.
- (S-3) Greater weighting factors assigned to the three statistical metrics computed for the AC<sub>2</sub> (the generalization case) in comparison with equal weighting values assumed for the metrics of the AC<sub>1</sub> and AC<sub>3</sub>, i.e. AC<sub>2</sub> > (AC<sub>1</sub> = AC<sub>3</sub>), Fig. 1b; see Additional file 1: Table S3.
- (S-4) Greater weighting factors assigned to the three statistical metrics computed from the AC<sub>3</sub> (the restricted case of the applicability conditions) in comparison with equal weighting values assumed for the metrics of the AC<sub>1</sub> and AC<sub>2</sub>, i.e. AC<sub>3</sub> > (AC<sub>1</sub> = AC<sub>2</sub>), Fig. 1b; see Additional file 1: Table S4.

After applying the MCDA (by means of the algorithm represented in Fig. 1b), the best ANN architectures were optimally selected and used for the development of the new gas geothermometers. A summarized description of the ranking score obtained by each ANN together with the optimal selection results are reported in Table 5. A full report containing all MCDA calculations obtained for the thirty-nine ANNs is reported in Additional file 1: Tables S1 to S4.

**GasGeo\_Eq**

This module was created to describe the new gas geothermometer equations inferred from the best ANNs (from here referred as GasG<sub>1</sub> to GasG<sub>8</sub>), and their applicability conditions. Table 6 summarizes the eight ANNs that were optimally selected for the development of the new gas geothermometer equations. The number of neurons used for the three layers of the ANNs (input, hidden, and output) are included. The input variables used by each ANN are also reported, including their relative contribution (in %) obtained from the sensitivity analysis (Garson 1991). Weighting and bias coefficients of the optimal ANNs are reported in Table 7. These coefficients were used for the development of each gas geothermometer (GasG<sub>*i*</sub>, *i* = 1 to 8) by using the following general equation (Eq. 5):

$$BHT_{GasG_i} (^{\circ}C) = \left\{ \left[ \frac{\left( \sum_{j=1}^m (LW_{(m,n)} \cdot \left( \frac{2}{1 + \exp(-2 \cdot (\sum_k IW_{(m,k)} \cdot X_{(k)} + b_{1(m)})} - 1 \right)) + b_{2(n)} \right) + 1 \right]}{2} \right] \cdot \alpha \right\} + 170, \tag{5}$$

where *X* is the input variable; *IW* and *b*<sub>1</sub> represent the associated coefficients to the input–hidden layers of the ANN; and *LW* and *b*<sub>2</sub> are the coefficients for the hidden–output layers. The subscripts *k*, *m*, and *n* refer to input, hidden, and output neurons, respectively, whereas *α* represents a normalization factor equal to 214 for the GasG<sub>1</sub>, GasG<sub>2</sub>, GasG<sub>6</sub>, and GasG<sub>8</sub>, and 193 for the remaining geothermometers (GasG<sub>3</sub>, GasG<sub>4</sub>, GasG<sub>5</sub>, and GasG<sub>7</sub>).

**Table 5 Ranking score and optimal selection of the eight ANN architectures using the MCDA method**

Ranking/ best selection	Neural network model	ANN architecture	Score				Score average
			Scenario 1	Scenario 2	Scenario 3	Scenario 4	
1	ANN-22 <sup>a</sup>	[3-34-1]	8.45	8.12	8.58	8.74	8.47
2	ANN-12 <sup>a</sup>	[4-13-1]	8.38	7.98	8.68	8.57	8.40
3	ANN-25 <sup>a</sup>	[4-10-1]	8.23	7.94	8.37	8.45	8.25
4	ANN-13 <sup>a</sup>	[5-9-1]	8.18	7.78	8.50	8.36	8.20
5	ANN-23 <sup>a</sup>	[3-15-1]	8.12	7.76	8.39	8.31	8.15
6	ANN-7 <sup>a</sup>	[3-18-1]	8.11	7.68	8.45	8.32	8.14
7	ANN-9 <sup>a</sup>	[3-15-1]	8.04	7.63	8.37	8.21	8.06
8	ANN-20 <sup>a</sup>	[3-18-1]	8.02	7.68	8.31	8.15	8.04
9	ANN-24	[3-14-1]	7.97	7.62	8.23	8.14	7.99
10	ANN-10	[3-9-1]	7.95	7.55	8.29	8.10	7.97
11	ANN-8	[3-9-1]	7.76	7.33	8.13	7.92	7.78
12	ANN-11	[3-9-1]	7.74	7.32	8.11	7.90	7.77
13	ANN-6	[3-12-1]	7.49	7.09	7.87	7.60	7.51
14	ANN-21	[3-6-1]	7.33	6.96	7.71	7.42	7.36
15	ANN-37	[3-9-1]	7.22	7.44	7.12	7.04	7.21
16	ANN-19	[3-27-1]	6.89	6.86	6.68	7.16	6.90
17	ANN-39	[5-6-1]	6.86	7.22	6.75	6.50	6.83
18	ANN-34	[3-8-1]	6.84	7.24	6.75	6.43	6.82
19	ANN-33	[3-8-1]	6.46	7.19	6.05	5.95	6.41
20	ANN-38	[4-8-1]	6.06	6.82	5.55	5.61	6.01
21	ANN-15	[1-34-1]	5.95	5.53	6.47	5.95	5.97
22	ANN-32	[3-8-1]	5.98	6.58	5.24	5.97	5.94
23	ANN-36	[3-7-1]	5.96	6.54	5.76	5.44	5.93
24	ANN-14	[1-33-1]	5.87	5.52	6.31	5.87	5.89
25	ANN-28	[1-14-1]	5.88	6.06	6.08	5.45	5.87
26	ANN-1	[1-32-1]	5.69	5.23	6.25	5.70	5.72
27	ANN-2	[1-3-1]	5.62	5.18	6.18	5.61	5.65
28	ANN-3	[1-13-1]	5.56	5.10	6.13	5.56	5.59
29	ANN-26	[5-5-1]	5.56	5.95	4.56	6.06	5.53
30	ANN-16	[1-16-1]	5.42	5.02	5.97	5.36	5.44
31	ANN-35	[3-11-1]	5.49	6.44	5.02	4.79	5.43
32	ANN-5	[2-28-1]	5.06	4.47	5.71	5.15	5.10
33	ANN-18	[2-34-1]	4.98	4.47	5.61	4.98	5.01
34	ANN-31	[2-9-1]	4.67	5.37	4.47	4.00	4.63
35	ANN-30	[1-8-1]	4.42	4.57	4.86	3.79	4.41
36	ANN-27	[1-10-1]	4.20	4.74	3.83	3.89	4.16
37	ANN-17	[1-25-1]	3.51	2.90	4.31	3.46	3.54
38	ANN-4	[1-26-1]	3.47	2.81	4.30	3.47	3.51
39	ANN-29	[1-4-1]	3.30	3.23	3.96	2.73	3.31

<sup>a</sup> Neural network models that represent the new gas geothermometer equations proposed in this work

Considering the complexity of the Eq. (5), it is pertinent to assume that a future application of the eight geothermometric equations could be discouraged by users when compared with simple equations already proposed from the existing gas geothermometers (Table 3). To overcome this limitation, a computer program (GaS\_GeoT) codified

**Table 6 New improved gas geothermometers proposed in this research work**

Neural network model	New gas geothermometer nomenclature (GasG <sub>i</sub> )	Input variables					Number of inputs neurons	Number of hidden neurons	Number of outputs neurons	ANN architecture
		1	2	3	4	5				
ANN-7	GasG <sub>1</sub>	ln(H <sub>2</sub> S/CO <sub>2</sub> ) 46.32%	ln(CH <sub>4</sub> /CO <sub>2</sub> ) 24.84%	ln(H <sub>2</sub> /CO <sub>2</sub> ) 28.84%			3	18	1	[3-18-1]
ANN-9	GasG <sub>2</sub>	ln(CO <sub>2</sub> /CH <sub>4</sub> ) 42%	ln(H <sub>2</sub> S/CH <sub>4</sub> ) 31.02%	ln(H <sub>2</sub> /CH <sub>4</sub> ) 26.97%			3	15	1	[3-15-1]
ANN-20	GasG <sub>3</sub>	ln(H <sub>2</sub> S/CO <sub>2</sub> ) 39.12%	ln(CH <sub>4</sub> /CO <sub>2</sub> ) 37.68%	ln(H <sub>2</sub> /CO <sub>2</sub> ) 23.2%			3	18	1	[3-18-1]
ANN-22	GasG <sub>4</sub>	ln(CO <sub>2</sub> /CH <sub>4</sub> ) 33.05%	ln(H <sub>2</sub> S/CH <sub>4</sub> ) 34.51%	ln(H <sub>2</sub> /CH <sub>4</sub> ) 32.44%			3	34	1	[3-34-1]
ANN-23	GasG <sub>5</sub>	ln(CO <sub>2</sub> /H <sub>2</sub> ) 24.6%	ln(H <sub>2</sub> S/H <sub>2</sub> ) 50.2%	ln(CH <sub>4</sub> /H <sub>2</sub> ) 25.21%			3	15	1	[3-15-1]
ANN12	GasG <sub>6</sub>	ln(CO <sub>2</sub> ) 26.18%	ln(H <sub>2</sub> S) 30.19%	ln(CH <sub>4</sub> ) 15.37%	ln(H <sub>2</sub> ) 28.26%		4	13	1	[4-13-1]
ANN-25	GasG <sub>7</sub>	ln(CO <sub>2</sub> ) 27.83%	ln(H <sub>2</sub> S) 12.81%	ln(CH <sub>4</sub> ) 25.87%	ln(H <sub>2</sub> ) 33.48%		4	10	1	[4-10-1]
ANN-13	GasG <sub>8</sub>	ln(H <sub>2</sub> S/CO <sub>2</sub> ) 25.13%	ln(CH <sub>4</sub> /CO <sub>2</sub> ) 5.11%	ln(H <sub>2</sub> /CO <sub>2</sub> ) 10.29%	ln(H <sub>2</sub> S/ H <sub>2</sub> ) 42.86%	ln(H <sub>2</sub> S/ H <sub>2</sub> ) 16.61%	5	9	1	[5-9-1]

According to the Garson's method (Garson 1991), the relative contribution for each input variable was determined in subsequent calculations of the geothermal reservoir temperatures. This value appears in percentage units below each input variable

**Table 7 Weighting and bias coefficients of the best ANN models (new gas geothermometers: GasG<sub>1</sub> to GasG<sub>8</sub>)**

Neuron	IW <sub>(m,k)</sub>					b1 <sub>(m,1)</sub>	LW <sub>(m,n)</sub>	b2 <sub>(n)</sub>
	Input variable 1	Input variable 2	Input variable 3	Input variable 4	Input variable 5	Bias	Weight	Bias
GasG <sub>1</sub> (k=3, m=18, n=1)								
1	-10.4966	-1.1378	0.8463			9.5785	-1.7731	-1.0356
2	22.0807	-12.9076	19.0550			-10.6989	-0.4489	
3	-14.8833	4.3913	-0.1612			8.5887	4.2903	
4	-51.9879	-0.3416	5.9872			19.3727	-0.2385	
5	16.7261	-4.8153	-0.0926			-9.4315	4.2683	
6	4.3149	3.8523	-5.4022			-0.5974	0.8856	
7	10.5391	-20.9557	33.7107			2.5871	-0.2241	
8	-2.0890	1.1073	20.2039			7.9015	0.3611	
9	0.5418	1.3637	-2.3029			0.0489	-3.3014	
10	-5.6519	0.9427	4.8160			1.3659	-10.5466	
11	1.1402	-3.2332	2.6054			-4.1995	1.5774	
12	5.2141	-0.5295	-4.0756			-1.3200	-7.5987	
13	3.6485	-19.4886	-0.8257			12.7539	5.0275	
14	-4.4414	0.0312	0.4327			4.3220	6.0399	
15	-3.5400	-0.7734	3.5304			1.4321	-2.7057	
16	-4.0591	19.9584	0.6652			-13.0292	5.0045	
17	5.5351	-1.0741	-5.2122			-1.3846	-4.8823	
18	2.7430	1.2762	0.3413			1.2202	-0.9442	
GasG <sub>2</sub> (k=3, m=15, n=1)								
1	-74.7200	-29.3826	71.2961			-9.1189	0.1740	-0.4764
2	-51.6547	21.8776	5.1177			13.0007	-0.3961	
3	-0.5704	-8.5959	11.2749			1.0424	-11.5175	
4	-2.8170	-5.8255	2.3507			6.9290	0.1195	
5	138.9491	-164.2573	14.7326			-2.4335	-0.2194	
6	10.0731	10.0604	3.4971			-0.7834	-10.0376	
7	8.2489	8.4971	2.0266			-0.4685	20.6834	
8	-6.9094	4.3475	2.2922			0.6182	0.5344	
9	-0.6145	-8.8942	11.4466			0.9874	11.3964	
10	-6.5749	-7.2235	-1.1089			0.2658	10.8816	
11	-109.4214	3.9406	18.2203			-19.5511	-0.2835	
12	68.6847	-37.4115	-78.8355			0.2217	0.1522	
13	-7.8109	1.3878	-6.8636			-9.5630	-14.2420	
14	8.0752	-1.1217	7.0651			10.1010	-13.8272	
15	-20.7201	10.9331	-12.9707			-8.4363	-0.2044	
GasG <sub>3</sub> (k=3, m=18, n=1)								
1	0.2605	-2.8324	0.7405			-4.6395	0.0965	-0.6741
2	-4.1589	5.9679	0.2775			5.0844	0.4591	
3	1.2094	13.6531	3.8001			-6.2841	0.1207	
4	-18.5398	0.6195	0.5262			10.0722	-0.5122	
5	2.7557	-5.8167	2.0639			-6.4099	0.0843	
6	-2.5005	-0.3730	-2.0823			-0.7033	-2.6769	
7	-8.7356	3.4340	1.9566			5.5936	-0.4428	
8	-3.9373	-0.2286	3.8324			1.0464	-0.6335	
9	2.1539	0.4060	3.8254			1.0421	-3.2267	
10	0.5414	12.1314	0.1324			-0.6745	0.5703	

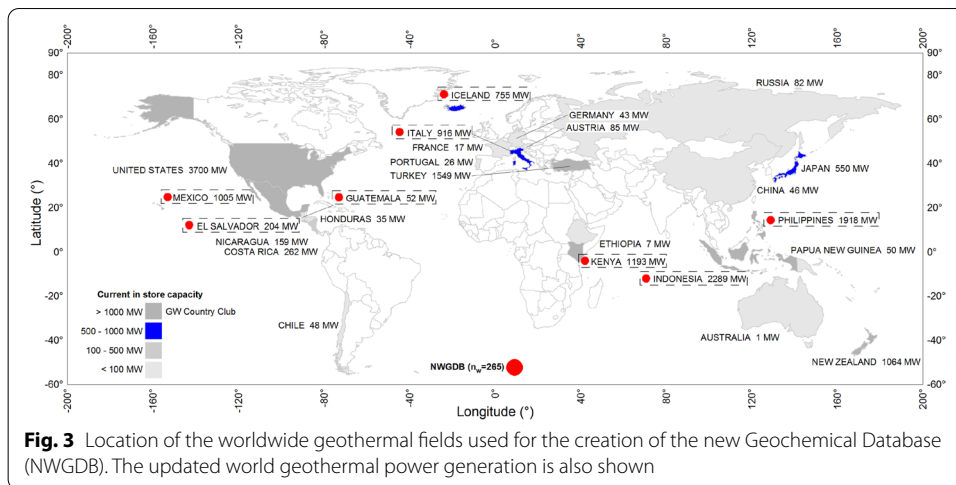
**Table 7 (continued)**

Neuron	$IW_{(m,k)}$					$b1_{(m,1)}$	$LW_{(m,n)}$	$b2_{(n)}$
	Input variable 1	Input variable 2	Input variable 3	Input variable 4	Input variable 5	Bias	Weight	Bias
11	2.4237	9.6744	0.8198			-1.0205	-0.4172	
12	2.1196	0.1036	-0.6293			-1.2896	-1.8327	
13	-3.1159	-1.4763	0.4641			-5.7187	-0.1605	
14	0.8259	1.6177	9.0616			1.9310	1.2339	
15	3.6085	-2.0623	-2.2694			5.3824	-0.1270	
16	-13.5662	1.6842	-2.0641			3.1308	-0.3221	
17	1.0129	4.8087	0.1825			4.8000	-0.2732	
18	4.7110	6.9895	-4.6717			-2.6020	-0.2882	
GasG <sub>4</sub> (k=3, m=34, n=1)								
1	0.8413	1.0481	4.5492			-4.3604	0.5430	0.7175
2	1.6852	-1.1809	4.0247			-4.2626	-0.6184	
3	-3.1846	3.0345	0.6129			4.2989	0.4547	
4	-2.4474	-2.2656	3.1111			3.6910	-0.1940	
5	-1.7082	4.3553	1.0333			3.4071	-2.2762	
6	-4.8797	6.0548	0.1265			2.8825	1.9591	
7	0.4419	4.7801	0.6585			-1.5392	2.1423	
8	1.9559	3.0935	-4.7555			-1.7757	3.0085	
9	-0.3619	-5.2932	6.4612			1.1353	3.5850	
10	1.5416	-1.9525	4.7684			-2.3951	2.4549	
11	-4.0310	-2.8368	8.6461			2.0696	-3.2083	
12	1.3875	2.8019	2.1562			-0.2151	-2.1271	
13	-6.2492	-0.8381	4.9963			-0.5368	2.3720	
14	-5.4812	-2.5508	0.2280			1.4242	2.2546	
15	1.2885	3.7342	-3.5705			-1.3129	-3.2678	
16	0.6120	6.8702	1.6604			0.4849	5.1290	
17	-8.0473	7.0992	3.3462			-0.6557	6.2486	
18	-6.5842	4.8933	4.9639			-0.5687	-5.3591	
19	2.0680	-2.0030	3.2922			0.5423	1.8140	
20	-2.7406	9.8262	8.5937			2.1876	0.6365	
21	7.7279	2.7843	-10.4992			0.5845	1.5671	
22	-10.3807	11.2299	1.2403			-0.8739	-3.2182	
23	-2.5572	-7.6695	-0.8698			-1.1968	2.8984	
24	6.7225	-3.9473	2.0077			2.1102	-4.8149	
25	12.5036	3.5952	-5.8924			2.4635	0.5574	
26	1.1006	-0.6048	-4.1736			2.9012	-0.9523	
27	-3.2534	5.5943	-2.5071			-0.9127	2.2162	
28	6.1765	2.4106	-1.4982			5.4936	7.1628	
29	-8.6020	4.5660	-3.5287			-3.1652	-3.1148	
30	2.2915	-3.0145	-2.4441			3.5085	0.0843	
31	-5.0789	-1.0713	-4.5243			-7.5724	-3.3251	
32	-1.0222	2.7998	-5.5370			-3.9064	-2.1729	
33	-10.9591	-4.7211	4.3928			-8.6022	3.8019	
34	6.5221	1.8689	2.3093			8.0946	-5.0940	
GasG <sub>5</sub> (k=3, m=15, n=1)								
1	4.6610	9.3354	-2.9977			-10.1278	1.2400	-1.9220
2	-2.0011	3.1019	4.6559			1.5880	2.0416	
3	-1.3652	-6.8953	0.2085			-4.3337	-3.5757	
4	2.2086	6.9970	-0.7490			4.9651	-3.5151	

**Table 7 (continued)**

Neuron	$IW_{(m,k)}$					$b1_{(m,1)}$	$LW_{(m,n)}$	$b2_{(n)}$
	Input variable 1	Input variable 2	Input variable 3	Input variable 4	Input variable 5	Bias	Weight	Bias
5	15.0925	12.5673	-1.6365			-2.8091	-0.6210	
6	-0.2917	-4.4285	-4.8459			-2.2553	1.9381	
7	-1.4554	-3.8719	0.6722			0.5524	-3.7513	
8	14.6274	-18.1442	1.1720			2.4447	-0.3234	
9	-4.1866	2.2442	6.8280			-2.1283	0.5290	
10	1.2033	-3.7615	-7.3399			1.4214	0.4381	
11	-0.9754	-5.1881	0.8466			0.5475	2.5457	
12	-4.1395	-8.2389	14.8899			-5.2272	-0.1067	
13	4.3414	15.7990	7.7251			-8.5051	-0.5169	
14	3.0032	0.5427	0.7922			2.7043	2.2648	
15	-0.2435	3.8179	0.3571			-2.9697	-1.0737	
GasG <sub>6</sub> (k=4, m=13, n=1)								
1	-7.0424	1.0592	0.5872	-4.9261		-0.9909	0.6853	-1.4850
2	-7.7923	0.7611	2.4586	11.5728		3.6960	1.0193	
3	-4.0733	-1.6240	1.7513	-3.4984		-0.9385	-0.7838	
4	-2.3369	-9.6032	0.3973	-0.2256		4.4306	-0.6460	
5	4.6053	3.2060	-6.2870	2.5629		0.3976	-1.5127	
6	4.7627	2.8469	-6.2893	3.3543		0.4991	1.4433	
7	-0.7276	5.4561	1.5209	4.0939		2.8770	3.4748	
8	-6.3221	5.2401	-0.0758	-1.8561		-3.3356	-0.7253	
9	0.0618	3.3627	0.2201	2.8292		-5.0749	-3.2557	
10	1.3088	1.8037	0.9426	6.5565		2.9480	-2.2063	
11	5.6895	3.8142	-3.7349	0.3184		3.0027	-0.6407	
12	3.5135	-3.4591	2.5028	-4.2444		5.3923	-1.8108	
13	-1.2860	-6.6278	-2.7831	-3.4112		-2.6695	2.0116	
GasG <sub>7</sub> (k=4, m=10, n=1)								
1	7.9032	-0.6015	0.7257	-0.8062		-4.9130	-0.4283	-0.6240
2	-2.9711	-3.8961	-0.3714	5.0930		2.1358	-0.4423	
3	0.6463	-1.2954	0.1709	7.7610		-0.2102	6.7284	
4	4.6109	-0.1063	-18.6311	-1.7928		-3.7415	-4.8884	
5	-8.3235	2.7237	-6.0207	-4.6909		-1.0212	0.7313	
6	5.7018	-4.4436	7.5091	-5.5924		7.9518	7.8252	
7	-5.6660	3.7079	-8.5863	7.8219		-8.8848	7.7244	
8	-0.5289	1.7097	-0.2783	-8.7773		0.0500	6.2932	
9	5.7362	-0.4683	-22.2945	-1.7934		-4.4083	4.7232	
10	8.8670	-2.9926	-0.6986	8.4312		6.8119	0.9991	
GasG <sub>8</sub> (k=5, m=9, n=1)								
1	22.4842	-33.9535	-29.5633	15.8975	48.5322	20.3047	0.1346	-1.3893
2	3.4291	0.5420	1.0172	-4.9756	4.5659	0.9671	1.6130	
3	-14.4388	3.7895	-7.2506	79.5516	5.1701	-33.4122	0.2263	
4	-43.2597	-1.2331	-24.8234	50.0978	-9.2286	-14.9195	-0.6028	
5	-2.2463	0.2702	0.2465	2.9529	1.1975	-1.8629	-19.6929	
6	48.8397	-3.6155	19.3550	-55.8381	20.4814	10.5299	-0.2055	
7	-2.9234	-0.2702	-0.2371	5.0167	-2.1591	-0.3359	9.1898	
8	-3.2910	-0.3027	-0.4428	6.2711	-2.9274	-0.4075	-7.1448	
9	1.7139	-0.3069	1.3356	-4.0765	-0.0239	2.9979	-17.7178	

The input variables of each new gas geothermometer are reported in Table 6. The subscripts  $k$ ,  $m$  and  $n$  refer to input, hidden and output neurons, respectively



in Java object-oriented programming was created. This program was developed for the effective use of the new gas geothermometer equations for the calculation of geothermal temperatures from gas-phase compositions ( $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{CH}_4$ , and  $\text{H}_2$ ). GaS\_GeoT and a quick user manual are available in the public server ([https://github.com/ANNGroup/GaS\\_GeoT.git](https://github.com/ANNGroup/GaS_GeoT.git)).

The applicability conditions of the new gas geothermometers ( $\text{GasG}_1$  to  $\text{GasG}_8$ ) are reported in Table 8. These limits are the min–max values of concentration and temperature used during the ANN learning processes. The gas concentrations must be given in mmol/mol units (dry-basis), whereas the temperature in °C.

The input variables for the  $\text{GasG}_1$ – $\text{GasG}_5$  and  $\text{GasG}_8$  are given by the natural logarithm values (dimensionless) because these equations use gas ratios, whereas for the  $\text{GasG}_6$  and  $\text{GasG}_7$ , the input variables are directly given in mmol/mol units.

### GasGeo\_Lit

This module was created to carry out a comparison of prediction efficiencies among the new ( $\text{GasG}_1$  to  $\text{GasG}_8$ ) and existing gas geothermometers using their temperature estimates. Twenty-five existing gas geothermometers have been included in Table 3. The fundamental criteria for selecting these geothermometers were as follows: (i) the determination of temperatures as a direct function of gas concentrations; and (ii) the use of gas concentrations ( $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{CH}_4$ ,  $\text{H}_2$ , and  $\text{N}_2$ ) most commonly applied in geothermometric studies.

### GasChemT\_Wells

This module was created for compiling the NWGDB ( $n_w = 265$ ) with data that were not used in the ANN learning process. NWGDB contains compositions (in mmol/mol, dry-basis) of major ( $\text{CO}_2$  and  $\text{H}_2\text{S}$ ) and trace ( $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{CH}_4$ ) geothermal gases,

and BHT measurements (in °C) logged in wells from thirteen world geothermal fields (Fig. 3): (1) Berlin geothermal field (BGF), El Salvador (compiled from: Renderos 2002;  $n=41$ : collected from a LIQDR); (2) Zunil geothermal field (ZGF), Guatemala (Giggenbach et al., 1992;  $n=4$ : LIQDR); (3) Krafla geothermal field (KGF), Iceland (Stefánsson 2017;  $n=8$ : LIQDR); (4) Kamojang geothermal field (KaGF), Indonesia (Laksminingpuri and Martinus 2013;  $n=10$ : VAPDR); (5) Sibayak geothermal field (SGF), Indonesia (Abidin et al. 2005;  $n=9$ : LIQDR); (6) Amiata geothermal field (AGF), Italy (Chiodini and Marini 1998;  $n=4$ : VAPDR); (7) Larderello geothermal field (LGF), Italy (Chiodini and Marini 1998;  $n=30$ : VAPDR); (8) Olkaria geothermal field (OGF), Kenya (Karingithi et al. 2010; Wamalwa 2015;  $n=29$ : 13 data collected from a LIQDR and 16 from a VAPDR); (9) Cerro Prieto geothermal field (CPGF), Mexico (Nieva et al. 1982; Nehring and D'Amore 1984;  $n=9$ : LIQDR); (10) Las Tres Virgenes geothermal field (LTVGF), Mexico (Verma et al. 2006;  $n=4$ : LIQDR); (11) Los Azufres geothermal field (LAGF), Mexico (Nieva et al. 1985 and 1987; Santoyo et al., 1991; Arellano et al., 2005; Barragán et al., 2005, 2008, 2012;  $n=36$ : 21 data collected from a LIQDR and 15 from a VAPDR); (12) Los Humeros geothermal field (LHGF), Mexico (López-Mendiola and Munguía, 1989; Tello et al., 2000; Arellano et al. 2003, 2015;  $n=72$ : 60 data compiled from a LIQDR and 12 from a VAPDR); and (13) Palinpinon geothermal field (PGF), Philippines (D'Amore et al. 1993;  $n=9$ : LIQDR).

To perform the analysis of prediction efficiency in new and existing gas geothermometers, the input data compiled in NWGDB were classified in two major groups: Group-1, LIQDR ( $g_1=178$ ), and Group-2, VAPDR ( $g_2=87$ ). A complete version of the NWGDB is reported in Additional file 1: Table S5.

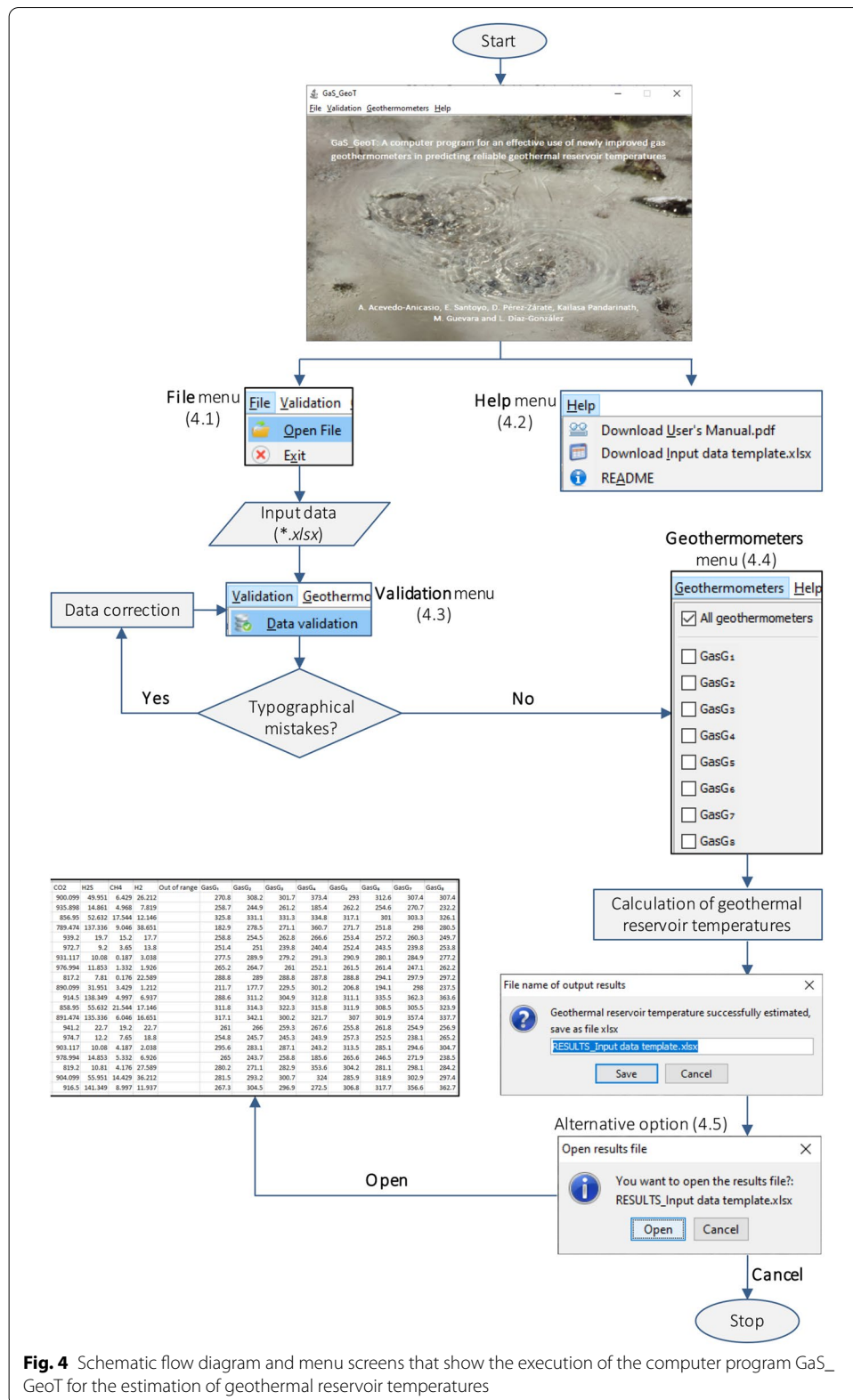
### GasGeoT\_Calc

This module was created to describe the new computer program GaS\_GeoT and the effective use of the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>). The user interface and a calculation routine of geothermal reservoir temperatures by using GaS\_GeoT are also described. Before running GaS\_GeoT and to apply the twenty-five geothermometer equations (Table 3), the concentration units of the gas-phase compositions and the applicability conditions of each gas geothermometer were verified (Table 8).

To proceed with the calculation of the reservoir temperatures, all gas geothermometers were executed (see Fig. 1a). As one of the goals of this work was focused on the effective use of the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>), the calculation process used by GaS\_GeoT is schematically referred in five major sections of Fig. 4 and summarized as follows.

To open or upload the input data through the *File* option (menu 4.1 in Fig. 4), an Excel spreadsheet (with a filename extension: *xlsx*) is required. This file must contain eight columns: Column 1, Sample ID (numerical data: integer); Columns 2–4: Geothermal Well, Geothermal Field, and Country information, respectively (in alphanumeric format); and the Columns 5–8, to enter the gas-phase compositions of the gases CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub> in mmol/mol (dry-basis), respectively (numerical data in floating format).





**Fig. 4** Schematic flow diagram and menu screens that show the execution of the computer program GaS\_GeoT for the estimation of geothermal reservoir temperatures

In the *Help* option (menu 4.2 in Fig. 4), a template file with some input data examples is available as a useful query. A User's Manual is also included in the *Help* option of the program, and the public server ([https://github.com/ANNGroup/GaS\\_GeoT.git](https://github.com/ANNGroup/GaS_GeoT.git)).

Two additional menu options appear in the main screen of GaS\_GeoT: *Validation* and *Geothermometers*. The *Validation* menu option performs an input data validation for checking the existence of typographical mistakes in the CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub> compositions (menu 4.3 in Fig. 4). The *Geothermometers* menu option invokes a computer subroutine which will make either the total or partial selection of the gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) by checking their corresponding boxes (menu 4.4 in Fig. 4). After selecting the geothermometers, the calculation of temperatures is performed, and the GaS\_GeoT will prompt the user to provide an output file name to print a report with the gas-phase compositions, and the temperature estimates calculated by the new gas geothermometers. An alternative output option is also requested by GaS\_GeoT for displaying the output results on screen (option 4.5 in Fig. 4).

After applying GaS\_GeoT and twenty-five existing geothermometers to the gas-phase composition of NWGDB (LIQDR  $g_1=178$  and VAPDR  $g_2=87$ ), an output file with all the temperature estimates is generated both to carry out the geochemometric evaluation, and to estimate the prediction efficiencies of the thirty-three gas geothermometers (Fig. 1a). A complete version of the output file is reported in Additional file 1: Tables S6 and S7, which contain the results obtained from the new and existing gas geothermometers, respectively.

### GasGeo\_Geochem

This module was developed to perform the geochemometric analyses using the temperature estimates obtained from new and existing gas geothermometers. To carry out these analyses, the following statistical metrics were applied: (1) Percent Difference, DIFF%; (2) Root Mean Square Error, RMSE; (3) Mean Absolute Error, MAE; (4) Mean Absolute Percentage Error, MAPE; and (5) the Difference Coefficient Ratio or statistical Theil's U test. As these metrics require the knowledge of actual bottom-hole temperatures, most of these were used as accuracy measures. The calculation equations used by these metrics are summarized in Table 9. RMSE, MAE, and MAPE metrics are interpreted as statistical residuals obtained between the bottom-hole temperatures (BHT<sub>m</sub>) and the temperature estimates predicted by any gas geothermometer, whereas the metrics DIFF% and Theil's U require a previous analysis. For example, when the DIFF% value is positive, the calculated temperature is greater than the BHT<sub>m</sub> of the respective geothermal well (which means an overestimation) and vice versa. Differences (DIFF%) either positive or negative  $\leq 20\%$  are assumed as acceptable estimates by considering the total propagated errors that are commonly quantified in some solute geothermometers (Verma and Santoyo 1997; García-López et al. 2014).

On the other hand, the statistical Theil's U is recommended for evaluating the prediction efficiency among several predictor models (Theil 1961). This metric was therefore used to evaluate the efficiency of the new geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) among other existing gas geothermometers. If the calculated values for the

Theil's coefficient ratios are less than 1, it means that the errors obtained by the new geothermometers are lower than that those obtained from the existing geothermometers and vice versa (Álvarez del Castillo et al. 2012).

## Results and discussion

After applying thirteen evaluation metrics and the MCDA, the thirty-nine ANNs were ranked (Table 5). Based on these optimization results, eight new gas geothermometers were successfully developed (Table 6): (1) GasG<sub>1</sub> [ $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ ,  $\ln(\text{H}_2/\text{CO}_2)$ ]; (2) GasG<sub>2</sub> [ $\ln(\text{CO}_2/\text{CH}_4)$ ,  $\ln(\text{H}_2\text{S}/\text{CH}_4)$ ,  $\ln(\text{H}_2/\text{CH}_4)$ ]; (3) GasG<sub>3</sub> [ $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ ,  $\ln(\text{H}_2/\text{CO}_2)$ ]; (4) GasG<sub>4</sub> [ $\ln(\text{CO}_2/\text{CH}_4)$ ,  $\ln(\text{H}_2\text{S}/\text{CH}_4)$ ,  $\ln(\text{H}_2/\text{CH}_4)$ ]; (5) GasG<sub>5</sub> [ $\ln(\text{CO}_2/\text{H}_2)$ ,  $\ln(\text{H}_2\text{S}/\text{H}_2)$ ,  $\ln(\text{CH}_4/\text{H}_2)$ ]; (6) GasG<sub>6</sub> [ $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$ ,  $\ln(\text{H}_2)$ ]; (7) GasG<sub>7</sub> [ $\ln(\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{CH}_4)$ ,  $\ln(\text{H}_2)$ ]; and (8) GasG<sub>8</sub> [ $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ ,  $\ln(\text{H}_2/\text{CO}_2)$ ,  $\ln(\text{H}_2\text{S})$ ,  $\ln(\text{H}_2\text{S}/\text{H}_2)$ ].

By considering the gas-phase compositions and the BHT<sub>m</sub> measurements compiled in the NWGDB ( $n_w = 265$ ), thirteen geothermal fields of the world (Berlin, Zunil, Krafla, Kamojang, Sibayak, Amiata, Larderello, Olkaria, Cerro Prieto, Las Tres Virgenes, Los Azufres, Los Humeros, and Palinpinon) were used for the geochemometric evaluation (see Additional file 1: Table S5). The geothermal reservoir temperatures estimated by applying the new gas geothermometers (Gas\_GeoT) along with twenty-five existing geothermometers are reported in Additional file 1: Tables S6 and S7, respectively.

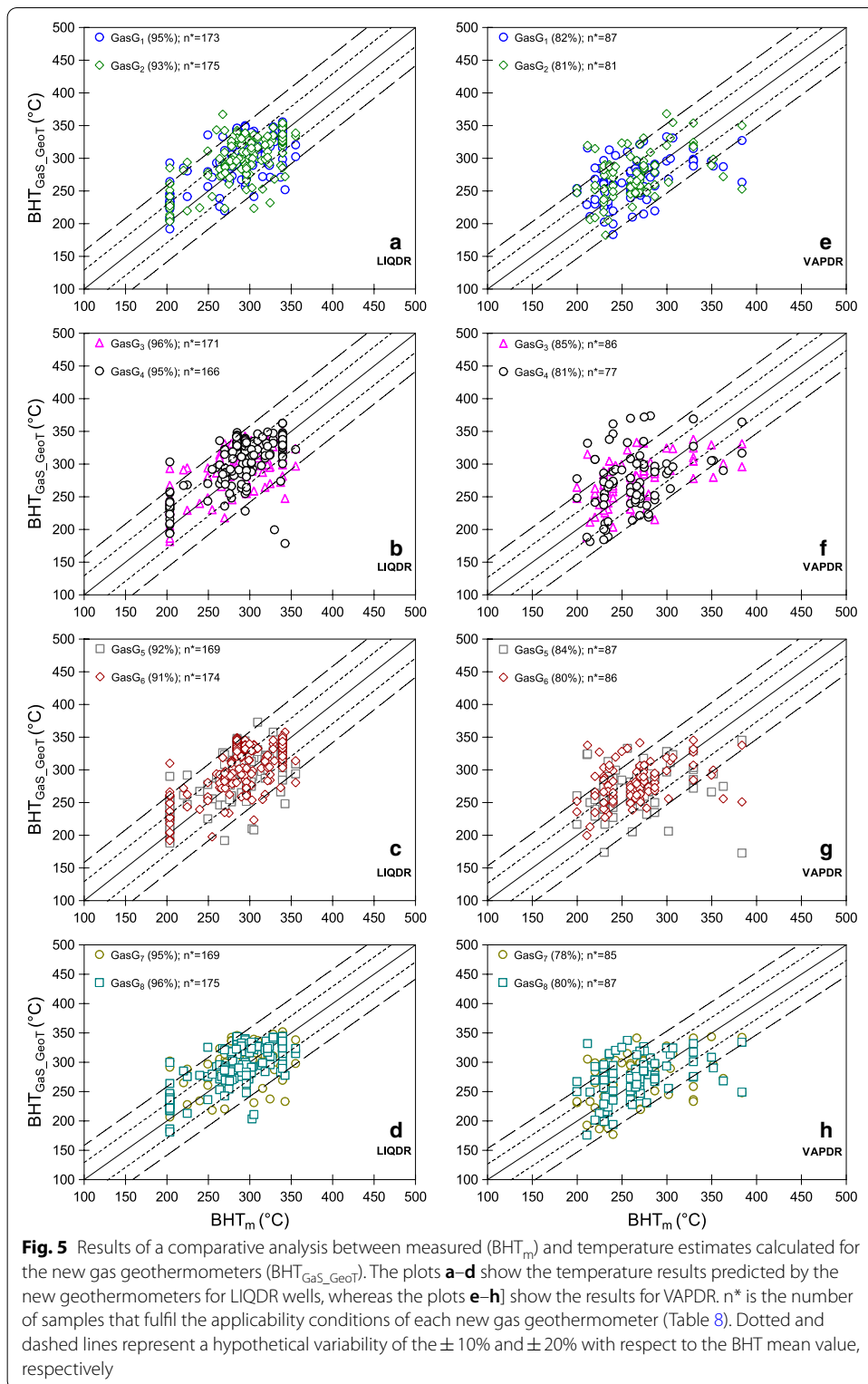
The prediction efficiency for all the gas geothermometers to determine reservoir temperatures was compared with measured BHT<sub>m</sub> values using the evaluation metrics (DIFF%, RMSE, MAE, MAPE, and Theil's U). A summary of the prediction efficiency results is reported in Table 10. A full version of this statistical comparison is also reported in Additional file 1: Tables S8 to S13.

### Results for wells located in LIQDR

The reservoir temperatures estimated for the gas phase compositions from LIQDR using the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) and twenty-five existing gas geothermometers were statistically compared with the corresponding BHTs, and the obtained results are presented as rounded-off values in the following sections:

#### Analysis of the DIFF% metric

The percent difference (DIFF%) calculated from the temperature estimates by all the thirty-three gas geothermometers (new and existing) and BHTs shows better efficiencies by the newly gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) with 91 to 96% of the predicted temperatures falling within the limits of acceptance (DIFF%  $\pm$  20%; Table 10). This behaviour is clearly observed when the temperature estimates are plotted against the measured BHTs (Fig. 5a–d). However, only four existing geothermometers from the literature (ND84c, AG85b, AG85d, and AG85f) predicted 85 to 88% of the temperatures in the same acceptable limits (Table 10 and Fig. 6a–d). About 55–63% of the estimated reservoir temperatures are overestimated (with DIFF%  $\leq$  20%), 30–34% of



the temperatures are underestimated (with  $DIFF\% \leq 20\%$ ), and 5–12% are equal (with  $DIFF\% \leq 1\%$ ), by the eight new gas geothermometers ( $GasG_1$  to  $GasG_8$ ), when compared to the BHTs, respectively (Table 10).

**Table 8 Applicability conditions defined for the eight new gas geothermometers developed**

New gas geothermometers (GasG <sub>i</sub> )	Gas concentrations (mmol/mol dry-basis)				Bottom-hole temperature °C
	CO <sub>2</sub>	H <sub>2</sub> S	CH <sub>4</sub>	H <sub>2</sub>	
	Concentration intervals [Min, Max]				BHT <sub>m</sub> [Min, Max]
GasG <sub>1</sub> , GasG <sub>2</sub> , GasG <sub>6</sub> , and GasG <sub>8</sub>	[311, 3206]	[0.1032, 357]	[0.000059, 57]	[0.1084, 477]	[170, 374]
GasG <sub>3</sub> , GasG <sub>4</sub> , GasG <sub>5</sub> , and GasG <sub>7</sub>	[699, 994]	[0.1032, 197]	[0.000059, 57]	[0.1084, 128]	[170, 374]
	Input variable 1	Input variable 2	Input variable 3	Input variable 4	Input variable 5
GasG <sub>1</sub>	[-9.17, -0.21]	[-16.56, -2.67]	[-9.07, 0.43]	-	-
GasG <sub>2</sub>	[2.67, 16.56]	[-0.82, 13.95]	[-2.99, 11.13]	-	-
GasG <sub>3</sub>	[-9.17, -1.39]	[-16.56, -2.67]	[-9.07, -1.77]	-	-
GasG <sub>4</sub>	[2.67, 16.56]	[-0.82, 13.95]	[-2.99, 11.13]	-	-
GasG <sub>5</sub>	[1.77, 9.07]	[-2.90, 6.14]	[-11.13, 2.99]	-	-
GasG <sub>6</sub>	[5.74, 8.07]	[-2.27, 5.88]	[-9.73, 4.04]	[-2.22, 6.17]	-
GasG <sub>7</sub>	[6.55, 6.90]	[-2.27, 5.28]	[-9.73, 4.04]	[-2.22, 4.85]	-
GasG <sub>8</sub>	[-9.17, -0.21]	[-16.56, -2.67]	[-9.07, 0.43]	[-2.27, 5.88]	[-2.90, 6.14]

The input variables of each new gas geothermometer are reported in Table 6

**Table 9 Statistical metrics used for evaluating the prediction efficiency of gas geothermometers (new and existing)**

Statistical metric	Calculation equation	Equation number	Reference
Percent difference (DIFF%)	$DIFF\% = \left[ \left( \frac{BHT_{CALC(i)} - BHT_{m(i)}}{BHT_{m(i)}} \right) \times 100 \right]$	(6)	García-López et al. (2014)
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (BHT_{m(i)} - BHT_{CALC(i)})^2}$	(7)	Willmott et al. (2009)
Mean absolute error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n  BHT_{m(i)} - BHT_{CALC(i)} $	(8)	Wang and Lu (2018)
Mean absolute percentage error (MAPE)	$MAPE = \left( \frac{1}{n} \sum_{i=1}^n \left  \frac{BHT_{m(i)} - BHT_{CALC(i)}}{BHT_{m(i)}} \right  \right) \times 100$	(9)	Li and Shi (2010)
Difference coefficient (Theil's U)	$Theil's U = \frac{\sqrt{\sum_{i=1}^n (BHT_{m(i)} - BHT_{GasGeoT(i)})^2}}{\sqrt{\sum_{i=1}^n (BHT_{m(i)} - BHT_{GasGeoLit(i)})^2}}$	(10)	Álvarez del Castillo et al. (2012)

$BHT_{m(i)}$  is the bottom-hole temperature measured in a geothermal well;  $BHT_{CALC(i)}$  is the temperature calculated by any gas geothermometer;  $BHT_{GasGeoT(i)}$  is the temperature calculated by the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>);  $BHT_{GasGeoLit(i)}$  is the temperature calculated by the existing gas geothermometers (Table 3), and  $n$  is the total number of gas samples

Arnórsson and Gunnlaugsson (1985) reported reservoir temperatures in some geothermal fields where the gas geothermometers may yield both under- and over-estimates for reservoir temperatures. Powell (2000) suggested that temperature underestimates may be due to the addition of un-equilibrated biogenic methane, whereas Barragán et al. (2000) stated that an overestimation of temperatures may derive from an excess of gases in total discharge of the wells.

Based on the DIFF% metric, the resulting ranking between 1st and 12th positions for the gas geothermometers under evaluation is occupied by GasG<sub>8</sub>, GasG<sub>3</sub>, GasG<sub>1</sub>, GasG<sub>7</sub>, GasG<sub>4</sub>, GasG<sub>2</sub>, GasG<sub>5</sub>, GasG<sub>6</sub>, AG85d, AG85b, ND84c, and AG85f, respectively.

**Analysis of RMSE, MAE, and MAPE metrics**

The RMSE values of the geothermometers vary between 32 and 136 (see Additional file 1: Table S9). The lowest values of RMSE ranging from 32 to 35 were obtained for the eight new gas geothermometers: GasG<sub>1</sub> to GasG<sub>8</sub> (Table 10). This systematic behaviour of RMSE replicates the better efficiency provided by the new gas geothermometers in comparison to those existing geothermometers, in predicting reservoir temperatures comparable to BHT<sub>m</sub>, which is clearly observed in Fig. 7a.

Moreover, all new gas geothermometers were characterized by the lowest values of MAE (ranging from 25 to 27), which also demonstrated that their temperature estimates have a better accuracy (Fig. 7b). With respect to the MAPE metric, small temperature differences between the temperature estimates and BHTs were also provided by the new gas geothermometers, which varied from 9 to 10 (Table 10). As MAPE represents an average value of absolute percentage errors, the lowest values calculated by the new gas geothermometers suggest a small better efficiency in predicting the reservoir temperatures, which is observed in Fig. 7c.

Based on the RMSE values, the resulting ranking between 1<sup>st</sup> and 12<sup>th</sup> positions for all the gas geothermometers was GasG<sub>3</sub>, GasG<sub>8</sub>, GasG<sub>5</sub>, GasG<sub>6</sub>, GasG<sub>7</sub>, GasG<sub>1</sub>, GasG<sub>2</sub>,

**Table 10 Summary of statistical results obtained from the geochemometric evaluation study of gas geothermometers**

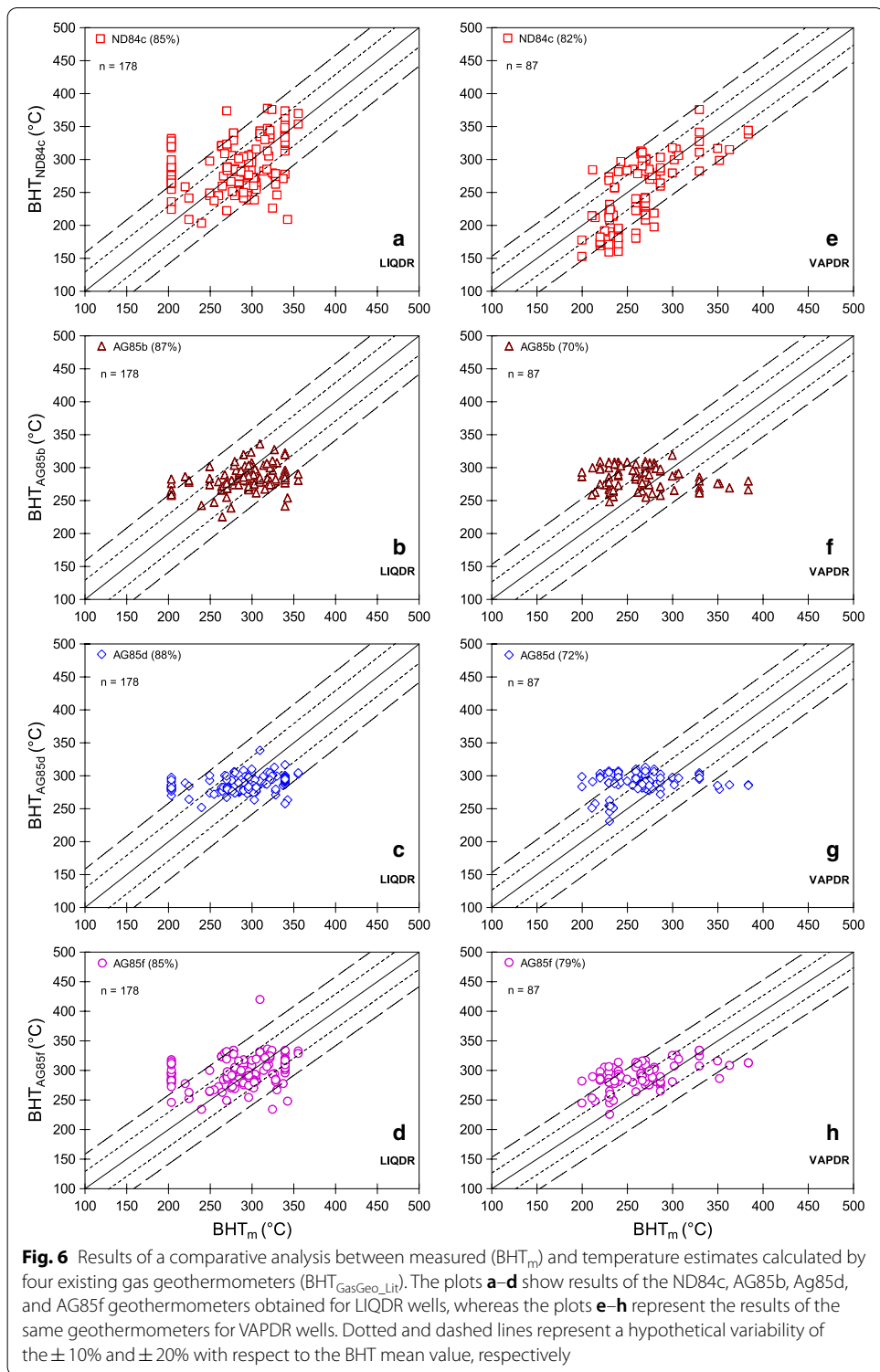
Evaluation metrics		Gas Geothermometers (new and existing) <sup>b</sup> ; NWGDB $n_w = 265$ data (LIQDR: $g_1 = 178$ ; VAPDR: $g_2 = 87$ )											
Geothermal reservoir-type <sup>a</sup>		$G_{as}G_1$	$G_{as}G_2$	$G_{as}G_3$	$G_{as}G_4$	$G_{as}G_5$	$G_{as}G_6$	$G_{as}G_7$	$G_{as}G_8$	ND84c	AG85b	AG85d	AG85f
Percent difference		Number of Samples (%—applicability)											
Over-estimation (DIFF% ≤ 20%)	LIQDR	103(60)	111(63)	99(58)	105(63)	93(55)	101(58)	97(58)	101(58)	70(39)	57(32)	55(31)	67(38)
	VAPDR	38(44)	42(52)	43(50)	37(48)	47(54)	51(59)	48(57)	42(48)	27(31)	62(71)	67(77)	69(79)
Under-estimation (DIFF% ≤ 20%)	LIQDR	56(32)	56(32)	52(30)	50(30)	57(34)	54(31)	53(31)	56(32)	94(53)	104(58)	117(66)	88(49)
	VAPDR	43(49)	32(39)	38(44)	36(47)	34(39)	29(34)	30(35)	39(45)	53(61)	20(23)	16(18)	13(15)
Equal to BHT (DIFF% ≤ 1%)	LIQDR	14(8)	8(5)	20(12)	11(7)	19(11)	19(11)	19(11)	18(10)	14(8)	17(10)	6(3)	23(13)
	VAPDR	6(7)	7(9)	5(6)	4(5)	6(7)	6(7)	7(8)	6(7)	7(8)	5(6)	4(5)	5(6)
Ranking <sup>c</sup>		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
DIFF% (± 20%) <sup>d</sup>	LIQDR	$G_{as}G_8$	$G_{as}G_3$	$G_{as}G_1$	$G_{as}G_7$	$G_{as}G_4$	$G_{as}G_2$	$G_{as}G_5$	$G_{as}G_6$	AG85d	AG85b	ND84c	AG85f
	VAPDR	168(96)	164(96)	164(95)	160(95)	157(95)	163(93)	156(92)	159(91)	156(88)	154(87)	152(85)	152(85)
RMSE (°C)	LIQDR	A98c	$G_{as}G_3$	$G_{as}G_1$	$G_{as}G_7$	ND84c	$G_{as}G_2$	$G_{as}G_4$	$G_{as}G_5$	ND84b	$G_{as}G_6$	AG85f	$G_{as}G_7$
	VAPDR	74(85)	73(85)	73(84)	71(82)	71(82)	66(82)	62(81)	70(81)	70(81)	69(80)	69(79)	66(78)
MAE (°C)	LIQDR	$G_{as}G_5$	$G_{as}G_8$	$G_{as}G_2$	$G_{as}G_6$	$G_{as}G_7$	$G_{as}G_1$	$G_{as}G_2$	$G_{as}G_4$	AG85b	AG85d	AG85f	ND84c
	VAPDR	31.6	31.8	32.6	33.2	33.4	33.5	33.9	34.7	36.9	38.7	38.8	45.2
MAPE (%)	LIQDR	$G_{as}G_1$	$G_{as}G_1$	ND84c	A98c	$G_{as}G_2$	ND84b	AG85f	$G_{as}G_6$	$G_{as}G_8$	$G_{as}G_7$	$G_{as}G_5$	$G_{as}G_4$
	VAPDR	35.6	38.0	38.9	39.8	39.8	40.1	41.0	41.9	42.1	43.4	44.8	49.0
Thell's U test	LIQDR	$G_{as}G_8$	$G_{as}G_5$	$G_{as}G_2$	$G_{as}G_7$	$G_{as}G_4$	$G_{as}G_1$	$G_{as}G_6$	$G_{as}G_2$	AG85f	AG85b	AG85d	ND84c
	VAPDR	25.2	25.2	25.3	25.8	26.3	26.5	26.6	27.3	27.7	28.3	30.6	33.9
	VAPDR	$G_{as}G_1$	$G_{as}G_1$	$G_{as}G_2$	$G_{as}G_6$	$G_{as}G_1$	A98c	$G_{as}G_8$	ND84c	ND84b	$G_{as}G_7$	AG85f	$G_{as}G_4$
	VAPDR	27.6	29.3	30.2	30.4	31.3	31	32.1	32.7	33.8	33.8	34.0	39.2
	VAPDR	$G_{as}G_5$	$G_{as}G_3$	$G_{as}G_8$	$G_{as}G_7$	$G_{as}G_4$	$G_{as}G_1$	$G_{as}G_6$	$G_{as}G_2$	AG85b	AG85f	AG85d	ND84c
	VAPDR	9.1	9.1	9.1	9.3	9.4	9.5	9.6	9.8	10.0	10.6	11.3	12.6
	VAPDR	$G_{as}G_1$	$G_{as}G_1$	$G_{as}G_2$	$G_{as}G_6$	$G_{as}G_1$	A98c	$G_{as}G_8$	ND84c	$G_{as}G_7$	AG85g	ND84b	$G_{as}G_4$
	VAPDR	10.6	11.2	11.5	11.8	11.9	11.9	12.4	12.7	12.8	12.9	13.1	15.4
	VAPDR	$G_{as}G_5$	$G_{as}G_8$	$G_{as}G_2$	$G_{as}G_7$	$G_{as}G_6$	$G_{as}G_1$	$G_{as}G_4$	$G_{as}G_2$	AG85b	AG85d	AG85f	ND84c
	VAPDR	$G_{as}G_5$	$G_{as}G_1$	ND84c	A98c	ND84b	AG85f	$G_{as}G_2$	$G_{as}G_6$	$G_{as}G_8$	$G_{as}G_7$	$G_{as}G_5$	$G_{as}G_4$

<sup>a</sup> Type of geothermal reservoir used to calibrate the gas geothermometers

<sup>b</sup> For the existing gas geothermometers (literature), we have selected only the best four geothermometers

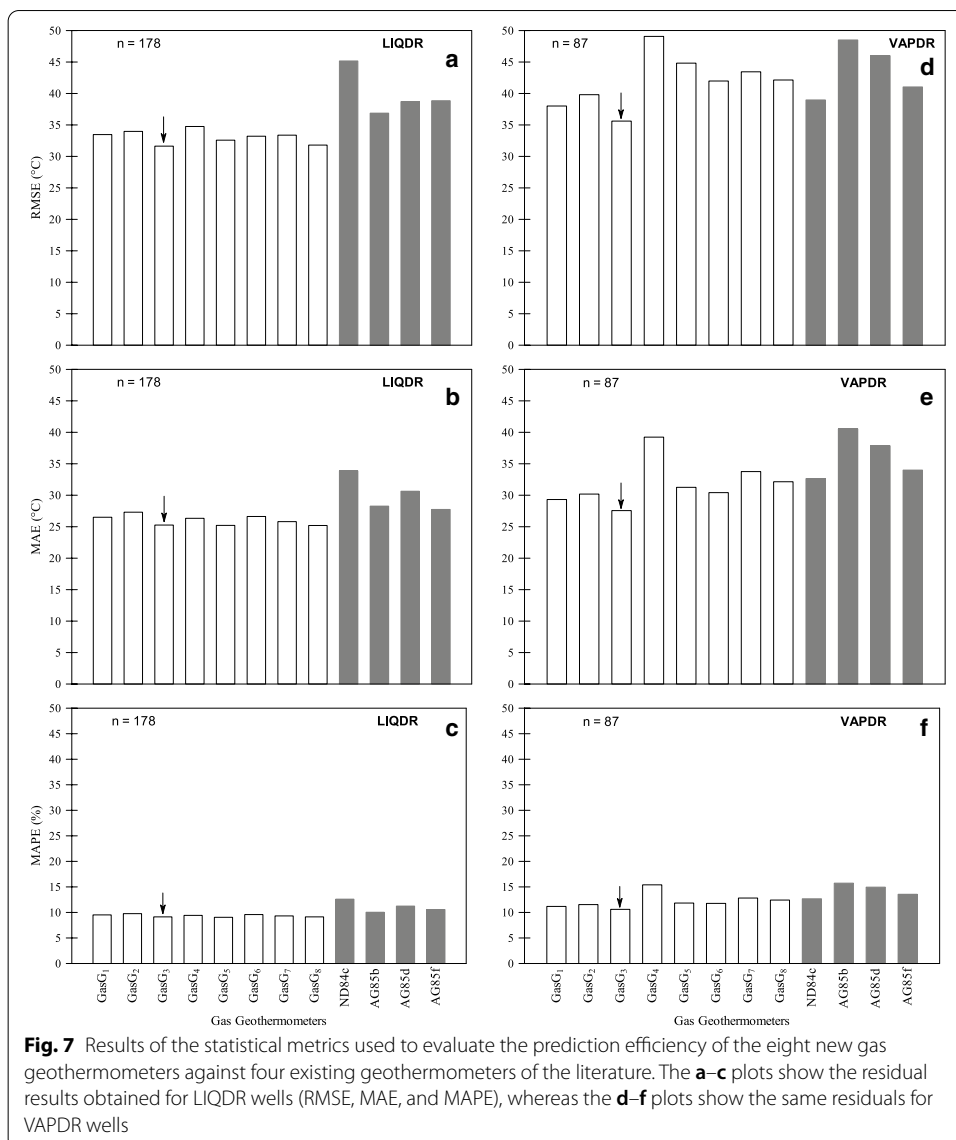
<sup>c</sup> The full ranking can be checked in the Additional file 1: Tables S8 to S13

<sup>d</sup> Within the limits of acceptance. The new gas geothermometers were typed in italic font



GasG<sub>4</sub>, AG85b, AG85d, AG85f, and ND84c, respectively, whereas for the MAE and MAPE metrics, the first five positions corresponded to the new gas geothermometers with an order slightly different but consistent as the best geothermometric tools: GasG<sub>8</sub>,





GasG<sub>5</sub>, GasG<sub>3</sub>, GasG<sub>7</sub>, and GasG<sub>4</sub>; and GasG<sub>5</sub>, GasG<sub>3</sub>, GasG<sub>8</sub>, GasG<sub>7</sub>, and GasG<sub>4</sub>, respectively (Table 10).

**Analysis of the statistical Theil’s U test (LIQDR)**

After analysing the Theil’s U results, the following inferences were accomplished: (1) the GasG<sub>3</sub> geothermometer may be considered with confidence as the best predictor model among all other geothermometers because the Theil’s U values were systematically lower than 1; (2) it was also observed that the new geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) systematically provided the lowest errors in comparison with twenty-five existing geothermometers (Table 10); and (3) after comparing the prediction efficiency among the new gas geothermometers, the GasG<sub>3</sub> exhibited Theil’s U lower values than 1 in comparison with those obtained for the remaining geothermometers (GasG<sub>1</sub>–GasG<sub>2</sub> and GasG<sub>4</sub>–GasG<sub>8</sub>). A full description of the LIQDR results obtained for all the

gas geothermometers under evaluation using the statistical evaluation metrics (DIFF%, RMSE, MAE, MAPE, and Theil's U) is reported in Additional file 1: Tables S8 to S10.

### Results for wells located in VAPDR

The prediction efficiency of the reservoir temperatures calculated in VAPDR samples using the thirty-three gas geothermometers (new and existing) was also compared with the actual BHT measurements.

### Analysis of the DIFF% metric

Based on the percent difference (DIFF%) values calculated by the new gas geothermometers, a good prediction efficiency was observed with 78 to 85% of the predicted temperatures falling within the limits of acceptance ( $\text{DIFF}\% \pm 20\%$ ), as it is observed in Fig. 5e–h. Out of the total twenty-five gas geothermometers, four geothermometers (ND84b, ND84c, AG85f, and A98c) from the literature have shown statistical differences comparable to these new geothermometers with 79 to 85% (Table 10, and Fig. 6e–h). About 44–59% of the estimated reservoir temperatures are overestimated (with  $\text{DIFF}\% \leq 20\%$ ), 34–49% of the temperatures are underestimated (with  $\text{DIFF}\% \leq 20\%$ ), and 5–9% are equal (with  $\text{DIFF}\% \leq 1\%$ ), by the eight new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>), when compared to the BHTs, respectively (Table 10).

According to these results, the new ranking between 1st and 12th positions for the gas geothermometers under evaluation was given by A98c, GasG<sub>3</sub>, GasG<sub>5</sub>, GasG<sub>1</sub>, ND84c, GasG<sub>2</sub>, GasG<sub>4</sub>, GasG<sub>8</sub>, ND84b, GasG<sub>6</sub>, AG85f, and GasG<sub>7</sub>, respectively. In this ranking, the prediction efficiency was actually very close between the geothermometers A98c and GasG<sub>3</sub>, differing in only two decimals of the applicability percentage (i.e. 85.1% for A98c, and 84.9% for GasG<sub>3</sub>; both percentages rounded-off as 85%). With these results, the gas geothermometer GasG<sub>3</sub> systematically shows a high prediction efficiency similar to those results obtained for LIQDR systems.

### Analysis of the RMSE, MAE, and MAPE metrics

The RMSE values calculated in VAPDR samples for all geothermometers showed a wider variability between 36 and 181 in comparison with those values estimated for LIQDR samples (see Additional file 1: Table S12). For the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>), seven out of the eight (except GasG<sub>4</sub>) predict low values of RMSE ranging from 36 to 45 (Table 10 and Fig. 7d), whereas for the existing gas geothermometers, a wider interval of RMSE was calculated (from 39 to 181).

In relation to the MAE metric, seven out of the eight new geothermometers were characterized by the lowest values ranging from 28 to 34, which also demonstrated that the temperature estimates predicted by the seven new geothermometers (except GasG<sub>4</sub>) are more accurate, as it is shown in Fig. 7e. Regarding the differences calculated in the MAPE values of the new gas geothermometers varied from 11 to 15 (Table 10 and Fig. 7f); which also suggest that seven out of eight geothermometers (except GasG<sub>4</sub>) provided a slight better efficiency.

Based on the integrated calculations obtained for all the metrics (RMSE, MAE, and MAPE), it was found that seven out of the eight new gas geothermometers show a better prediction efficiency to determine the reservoir temperatures in VAPDR gas samples.

The resulting ranking between 1st and 12th positions according to RMSE was given by GasG<sub>3</sub>, GasG<sub>1</sub> ND84c, A98c, GasG<sub>2</sub>, ND84b, AG85f, GasG<sub>6</sub>, GasG<sub>8</sub>, GasG<sub>7</sub>, GasG<sub>5</sub>, and GasG<sub>4</sub>, respectively, whereas for the MAE and MAPE metrics, the five positions corresponded to the new gas geothermometers: GasG<sub>3</sub>, GasG<sub>1</sub>, GasG<sub>2</sub>, GasG<sub>6</sub>, and GasG<sub>5</sub>.

#### **Analysis of the statistical Theil's U test (VAPDR)**

After applying the Theil's U results (Table 10), two interpretations are inferred: (1) it was systematically observed that the GasG<sub>3</sub> predictor model (Theil's U values lower than 1) is the most reliable and accurate gas geothermometer over the rest of the new geothermometers (GasG<sub>1</sub>–GasG<sub>2</sub> and GasG<sub>4</sub>–GasG<sub>8</sub>); and (2) it was found that two (GasG<sub>1</sub> and GasG<sub>3</sub>) out of the eight new geothermometers exhibited lower errors in estimating reservoir temperatures in comparison with those estimates predicted by the existing geothermometers.

A full description of the results obtained from the use of the thirty-three gas geothermometers in VAPDR systems is reported in Additional file 1: Tables S11 to S13.

As a final remark of this research work, it was demonstrated, for the first time, the effectiveness of the MCDA optimization method for ranking and selecting the most reliable ANN architectures to predict a dependent variable (BHT<sub>m</sub>) as a function of multiple independent variables (gas-phase compositions: CO<sub>2</sub>, H<sub>2</sub>S, CH<sub>4</sub>, and H<sub>2</sub>). We consider that this new evaluation proposal is actually innovating the evaluation methods commonly used in ANNs for evaluating their training, validation, and test stages, which typically rely on the simple use of the linear correlation coefficients ( $r$ ) obtained between measured and simulated data.

With the optimal selection of the best predictor models used to correlate BHT<sub>m</sub> and gas-phase compositions, eight new improved gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) were successfully developed. Most of these new geothermometers provide reliable estimations of geothermal reservoir temperatures. We have also demonstrated that the prediction efficiency of these new geothermometric tools (mainly GasG<sub>1</sub> and GasG<sub>3</sub> geothermometers) exceeds the efficiency of those existing gas geothermometers available in the literature.

#### **Conclusions**

Eight new improved gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) based on an optimized selection of artificial neural networks by using the MCDA method were successfully developed for the reliable prediction of the geothermal reservoir temperatures. For an effective and practical use of these geothermometers, a new computer program GaS\_GeoT was successfully developed. The evaluation of the efficiency of the new improved gas geothermometers in predicting the reservoir temperatures was successfully demonstrated for geothermal wells of LIQDR and VAPDR systems.

The new gas geothermometers (GasG<sub>1</sub> to GasG<sub>8</sub>) provided the best prediction efficiencies for geothermal wells from LIQDR, whereas two out the eight (GasG<sub>1</sub> and GasG<sub>3</sub>) demonstrated their best efficiency in predicting reservoir temperatures for VAPDR. Among the new gas geothermometers, the best geothermometric tool for predicting reservoir temperatures in LIQDR and VAPDR systems is the GasG<sub>3</sub>

geothermometer which uses the gas ratios:  $\ln(\text{H}_2\text{S}/\text{CO}_2)$ ,  $\ln(\text{CH}_4/\text{CO}_2)$ , and  $\ln(\text{H}_2/\text{CO}_2)$ .

Out of the total twenty-five existing gas geothermometers, the most consistent geothermometers to predict reservoir temperatures comparable to those values inferred from the new geothermometers were as follows: (1) the ND84c, AG85b, AG85d, and AG85f geothermometers for LIQDR (which use  $\text{CO}_2\text{-H}_2\text{S}$ ,  $\text{CO}_2\text{-H}_2$ ,  $\text{H}_2$ , and  $\text{H}_2\text{S}$  concentrations); (2) the ND84b, ND84c, AG85f, and A98c geothermometers for VAPDR (which use  $\text{CO}_2\text{-H}_2$ ,  $\text{CO}_2\text{-H}_2\text{S}$ ,  $\text{H}_2\text{S}$ , and  $\text{H}_2\text{S}$  concentrations).

Taking in consideration the higher prediction efficiencies observed in predicting reservoir temperatures, the new gas geothermometers and the GaS\_GeoT program may have the potential to become one of the most preferred geothermometric tools for the reliable estimation of the reservoir temperatures in geothermal prospection and exploitation.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40517-020-00182-9>.

**Additional file 1.** The four statistical evaluation scenarios created with the MCDA methodology for ranking and selection of thirty-nine ANNs. Details of the NWGDB compiled from world-wide geothermal fields, and used for the development of new gas geothermometers are also included, together with deep temperatures determined by using gas geothermometers (new and existing). **Table S1.** MCDA Scenario 1 and ranking scores estimated for thirty-nine ANNs using statistical evaluation metrics. **Table S2.** MCDA Scenario 2 and ranking scores estimated for thirty-nine ANNs using the statistical evaluation metrics. **Table S3.** MCDA Scenario 3 and ranking scores estimated for thirty-nine ANNs using the statistical evaluation metrics. **Table S4.** MCDA Scenario 4 and ranking scores estimated for thirty-nine ANNs using the statistical evaluation metrics. **Table S5.** Chemical composition of gases from thirteen geothermal fields compiled from the literature: NWGDB ( $n_w = 265$ ). **Table S6.** Temperature estimates predicted by the new gas geothermometers (GasG<sub>1</sub> to GasG<sub>6</sub>) using the NWGDB ( $n_w = 265$ ). **Table S7.** Temperature estimates predicted by twenty-five existing gas geothermometers using the gas-phase compositions compiled in NWGDB ( $n_w = 265$ ). **Table S8.** Comparison among reservoir temperatures estimated by the new gas geothermometers and BHT<sub>m</sub> of LIQDR systems ( $g_1 = 178$ ). **Table S9.** Residual errors calculated for the thirty-three gas geothermometers using the gas compositions of LIQDR systems ( $g_1 = 178$ ). **Table S10.** Statistical Theil's U results for evaluating the prediction efficiency of the gas geothermometers using LIQDR systems. **Table S11.** Comparison among reservoir temperatures estimated by the new gas geothermometers and BHT<sub>m</sub> of VAPDR systems ( $g_2 = 87$ ). **Table S12.** Residual errors calculated for the thirty-three gas geothermometers using the gas compositions of VAPDR systems ( $g_2 = 87$ ). **Table S13.** Statistical Theil's U results for evaluating the prediction efficiency of the gas geothermometers using VAPDR systems.

### Abbreviations

LIQDR: Liquid-dominated reservoir; VAPDR: Vapour-dominated reservoir; ANN: Artificial neural networks; MCDA: Multi-criteria decision analysis; NWGDB: New Worldwide Geochemical Database; BHT<sub>m</sub>: Bottom-hole temperatures (measured); BHT<sub>ANN</sub>: Simulated ANN temperatures; WG\_SubDB<sub>1</sub>: Worldwide geochemical sub-database ( $q_1 = 527$ ); WG\_SubDB<sub>2</sub>: Worldwide geochemical sub-databases ( $q_2 = 498$ ); WG\_SubDB<sub>3</sub>: Worldwide geochemical sub-databases ( $q_3 = 97$ ); LM: Levenberg–Marquardt algorithm; AC<sub>1</sub>: Analysis case 1; AC<sub>2</sub>: Analysis case 2; AC<sub>3</sub>: Analysis case 3; MAVT: Multi-attribute value theory; S-1: MCDA scenario 1; S-2: MCDA scenario 2; S-3: MCDA scenario 3; S-4: MCDA scenario 4; BGF: Berlin geothermal field; ZGF: Zunil geothermal field; KGF: Krafla geothermal field; KaGF: Kamojang geothermal field; SGF: Sibayak geothermal field; AGF: Amiata geothermal field; LGF: Larderello geothermal field; OGF: Olkaria geothermal field; CPGF: Cerro Prieto geothermal field; LTVGF: Las Tres Virgenes geothermal field; LAGF: Los Azufres geothermal field; LHGF: Los Humeros geothermal field; PGF: Palinpinon geothermal field; DIFF%: Percent difference; RMSE: Root mean square error; MAE: Mean absolute error; MAPE: Mean absolute percentage error.

### Acknowledgements

The authors also express their gratitude to E.O. García-Mandujano of IER-UNAM for her valuable comments and assistance in the literature compilation. The first author also wants to thank CONACYT for the scholarship awarded to carry out this research work as part of his PhD studies. The second author also wants to thank to the Physics and Mathematics Department of the Universidad Iberoamericana (Mexico) for the Professorship support in a sabbatical leave program carried out on 2018–19, time period where the present research work was started.

### Authors' contributions

AAA contributed to design of ANNs and computer code programming; ES is the Leader of the research, and involved in design of computational methodology, fluid geochemistry, and geochemometrics; DPZ contributed to gas geothermometry analysis and unit conversions; KP is involved in gas geothermometry and mineral–gas assemblages;

MG created the geochemical databases; LDG calculated statistical metrics. All authors read and approved the final manuscript.

#### Funding

This work was partially funded by the P09 CeMIE-Geo research project (207032 CONACyT-SENER).

#### Availability of data and materials

The data used in this work are included in both the article and additional material. GaS\_GeoT computer program was written in Java programming language, and is available for downloading from the following public server or repository ([https://github.com/ANNGroup/GaS\\_GeoT.git](https://github.com/ANNGroup/GaS_GeoT.git)), which may be directly accessed by the users. A User's Manual for running GaS\_GeoT is also included, together with a README file, and the open-source code license (MIT). Hardware specifications to install and to run effectively the GaS\_GeoT program are also described in the User's Manual.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup> Doctorado en Ciencias, Centro de Investigación en Ciencias-ICBA, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, 62209 Morelos, México. <sup>2</sup> Instituto de Energías Renovables, Universidad Nacional Autónoma de México, Priv. Xochicalco S/N, Temixco, 62580 Morelos, México. <sup>3</sup> CONACYT-Instituto de Geofísica, Universidad Nacional Autónoma de México, Circuito interior s/n, Coyoacán, 04510 Ciudad de México, México. <sup>4</sup> Centro de Investigación en Ciencias-ICBA, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, 62209 Morelos, México.

Received: 15 September 2020 Accepted: 26 November 2020

Published online: 18 January 2021

#### References

- Abidin Z, Alip D, Nenneng L, Ristin PI, Fauzi A. Environmental isotopes of geothermal fluids in Sibayak geothermal field. In: Use of Isotope Techniques to Trace the Origin of Acidic Fluids in geothermal systems, IAEA Press, Austria. 2005;37–60. [https://inis.iaea.org/search/search.aspx?orig\\_q=RN:36065710](https://inis.iaea.org/search/search.aspx?orig_q=RN:36065710). Accessed 19 Nov 2020.
- Adem EB, Geneletti D. Multi-criteria decision analysis for nature conservation: a review of 20 years of applications. *Methods Ecol Evol.* 2018;9:42–53.
- ÁlvarezdelCastillo A, Santoyo E, García-Valladares O. A new empirical void fraction correlation inferred from artificial neural networks for modeling two phase flow in geothermal wells. *Comput Geosci.* 2012;41:25–39.
- Arellano VM, García A, Barragán RM, Izquierdo G, Aragón A, Nieva D. An updated conceptual model of the Los Humeros geothermal reservoir (Mexico). *J Volcanol Geotherm Res.* 2003;124:67–88.
- Arellano VM, Torres MA, Barragán RM. Thermodynamic evolution of the Los Azufres, Mexico, geothermal reservoir from 1982 to 2002. *Geothermics.* 2005;34:592–616.
- Arellano VM, Barragán RM, Ramírez M, López S, Paredes A, Aragón A, Tovar R. The response to exploitation of the Los Humeros (México) geothermal reservoir. In: Proceedings of the World Geothermal Congress 2015; Melbourne, Australia. 2015. <https://www.geothermal-energy.org/pdf/IGStandard/WGC/2015/14029.pdf>. Accessed 19 Nov 2020
- Arnórsson S, Gunnlaugsson E. New gas geothermometers for geothermal exploration—calibration and application. *Geochim Cosmochim Acta.* 1985;49:1307–25.
- Arnórsson S. Gas chemistry of the Krísvík geothermal field, Iceland, with special reference to evaluation of steam condensation in upflow zones. *Jökull.* 1987;37:31–48. <https://jokulljournal.is/21-39/1987/031.pdf>. Accessed 19 Nov 2020
- Arnórsson S, Fridriksson T, Gunnarsson I. Gas chemistry of the Krafla geothermal field, Iceland. In: Intl Symp Water-Rock Interaction, Auckland, New Zealand. 1998;613–616. <https://www.tib.eu/en/search/id/BLCP:CN023965846/Gas-chemistry-of-the-Krafla-Geothermal-Field-Iceland?cHash=ea876fde0d9fec307ca89a09bd0ce5ca>. Accessed 20 Nov 2020
- Arnórsson S, Bjarnason JÖ, Giroud N, Gunnarsson I, Stefánsson A. Sampling and analysis of geothermal fluids. *Geofluids.* 2006;6:203–16. <https://doi.org/10.1111/j.1468-8123.2006.00147.x>.
- Arslan O, Yetik O. ANN based optimization of supercritical ORC-Binary geothermal power plant: Simav case study. *Appl Therm Eng.* 2011;31:3922–8.
- Barragán RM, Arellano VM, Nieva D, Portugal E, García A, Aragón A, Tovar R, Torres-Alvarado I. Gas geochemistry of the Los Humeros geothermal field, México. In: Proceedings World Geothermal Congress 2000; Kyushu-Tohoku, Japan. 2000. <https://www.geothermal-energy.org/pdf/IGStandard/WGC/2000/R0130.PDF>. Accessed 19 Nov 2020
- Barragán RM, Gómez VA, Portugal E, Sandoval F, Segovia N. Gas geochemistry for the Los Azufres (Michoacán) geothermal reservoir, México. *Ann Geophys.* 2005;48:145–157. <https://www.annalsofgeophysics.eu/index.php/annals/article/viewFile/3189/3234>. Accessed 19 Nov 2020
- Barragán RM, Arellano VM, Armenta M, Aguado R. Cambios químicos en fluidos de pozos del campo geotérmico de Los Humeros: Evidencia de recarga profunda. *Geotermia.* 2008;21:11–20. <http://pubs.geothermal-library.org/lib/journals/Geotermia-Vol21-2.pdf>. Accessed 20 Nov 2020
- Barragán RM, Arellano VM, Mendoza A, Reyes L. Variación de la composición del vapor en pozos del campo geotérmico de Los Azufres, México, por efecto de la reinyección. *Geotermia.* 2012;25(1), 3–9. <https://biblat.unam.mx/hevila/Geotermia/2012/vol25/no1/1.pdf>. Accessed 20 Nov 2020
- Barragán RM, Núñez J, Arellano VM, Nieva D. EQUILGAS: Program to estimate temperatures and in situ two-phase conditions in geothermal reservoirs using three combined FT-HSH gas equilibria models. *Comput Geosci.* 2016;88:1–8.

- Bertrami R, Cioni R, Corazza E, D'Amore F, Marini L. Carbon monoxide in geothermal gases. Reservoir temperature calculations at Larderello (Italy). *Geother Res Council Trans*. 1985;9:299–303. <http://pubs.geothermal-library.org/lib/grc/1001282.pdf>. Accessed 20 Nov 2020
- Blamey NJ. H<sub>2</sub>S concentrations in geothermal and hydrothermal fluids—a new gas geothermometer. In: Proceedings thirty-first workshop on geothermal reservoir engineering, Stanford University, Stanford, California, USA. 2006;403–407. <https://pangea.stanford.edu/ERE/pdf/IGASstandard/SGW/2006/blamey.pdf>. Accessed 20 Nov 2020
- Chamorro CR, Mondéjar ME, Ramos R, Segovia JJ, Martín MC, Villamañán MA. World geothermal power production status: Energy, environmental and economic study of high enthalpy technologies. *Energy*. 2012;42:10–8.
- Chiodini G, Marini L. Hydrothermal gas equilibria: the H<sub>2</sub>O–H<sub>2</sub>–CO<sub>2</sub>–CO–CH<sub>4</sub> system. *Geochim Cosmochim Acta*. 1998;62:2673–87.
- D'Amore F, Panichi C. Evaluation of deep temperatures of hydrothermal systems by a new gas geothermometer. *Geochim Cosmochim Acta*. 1980;44:549–56.
- D'Amore F, Ramos-Candelaria, MN, Seastres JrJ, Ruaya JR, Nuti S. Applications of gas chemistry in evaluating physical processes in the Southern Negros (Palinpinon) geothermal field, Philippines. *Geothermics*. 1993;22:535–553.
- Díaz-González L, Santoyo E, Reyes J. Tres nuevos geotermómetros mejorados de Na/K usando herramientas computacionales y geoquímicas: aplicación a la predicción de temperaturas de sistemas geotérmicos. *Rev Mex Cienc Geol*. 2008;25:465–482. <http://www.scielo.org.mx/pdf/rmcg/v25n3/v25n3a7.pdf>. Accessed 20 Nov 2020
- Dincer I, Acar C. A review on clean energy solutions for better sustainability. *Int J Energy Res*. 2015;39:585–606.
- Ellis AJ. Chemical equilibrium in magmatic gases. *Am J Sci*. 1957;255:416–31.
- Estévez RA, Alamos FH, Walshe T, Gelcich S. Accounting for uncertainty in value judgements when applying multi-attribute value theory. *Environ Model Assess*. 2018;23:87–97.
- García-López CG, Pandarinath K, Santoyo E. Solute and gas geothermometry of geothermal wells: a geochemometrics study for evaluating the effectiveness of geothermometers to predict deep reservoir temperatures. *Int Geol Rev*. 2014;56:2015–49.
- García-Mandujano EO. SYS\_GASCHEM: Information system based on web technologies for the processing of geochemical databases and the determination of the temperature of geothermal systems. Universidad Politécnica del Estado de Morelos. 2019;137. [http://www.cie.unam.mx/cgi-bin/CemieGeo/sys\\_gaschem.pl](http://www.cie.unam.mx/cgi-bin/CemieGeo/sys_gaschem.pl). Accessed 20 Nov 2020
- Garson DG. Interpreting neural network connection weights. *AI Expert*. 1991;6:46–51.
- Giggenbach WF. Geothermal gas equilibria. *Geochim Cosmochim Acta*. 1980;44:2021–32.
- Giggenbach WF. Chemical techniques in geothermal exploration. In: D'Amore F. (Ed) Applications of geochemistry in geothermal reservoir development: series of technical guides on the use of geothermal energy, UNITAR/UNDP Centre Press, Rome. 1991;119–142. <https://www.studocu.com/cl/document/universidad-catolica-del-norte/introduccion-a-la-geoquimica/otros/giggenbach-1991/5193236/view>. Accessed 20 Nov 2020
- Giggenbach WF, Glover RB. Tectonic regime and major processes governing the chemistry of water and gas discharges from the Rotorua geothermal field. *New Zealand Geothermics*. 1992;21:121–40.
- Guo Q, Pang Z, Wang Y, Tian J. Fluid geochemistry and geothermometry applications of the Kangding high-temperature geothermal system in eastern Himalayas. *Appl Geochem*. 2017;81:63–75.
- Gutiérrez-Negrín LC. Current status of geothermal-electric production in Mexico. *Environ Earth Sci*. 2019;249:1–11. <https://doi.org/10.1088/1755-1315/249/1/012017/pdf>.
- Gutiérrez-Negrín LC, Canchola Félix I, Romo-Jones JM, Quijano-León JL. Geothermal energy in Mexico: update and perspectives. In Proceedings, Proceedings World Geothermal Congress 2020; Reykjavik, Iceland. 2020. [https://www.researchgate.net/profile/Luis\\_Gutierrez-Negrin/publication/343111483\\_Geothermal\\_energy\\_in\\_Mexico\\_update\\_and\\_perspectives/links/5f17406292851cd5fa3a0275/Geothermal-energy-in-Mexico-update-and-perspectives.pdf](https://www.researchgate.net/profile/Luis_Gutierrez-Negrin/publication/343111483_Geothermal_energy_in_Mexico_update_and_perspectives/links/5f17406292851cd5fa3a0275/Geothermal-energy-in-Mexico-update-and-perspectives.pdf). Accessed 20 Nov 2020
- Henley RW, Truesdell AH, Barton PB, Whitney JA. Fluid-mineral equilibria in hydrothermal systems. *Reviews in economic geology*, vol 1, Society of Economic Geologists Inc. 1985;1:258. [https://www.segweb.org/store\\_info/REV/REV-01-Additional-Product-Info.pdf](https://www.segweb.org/store_info/REV/REV-01-Additional-Product-Info.pdf). Accessed 20 Nov 2020
- Kacandes GH, Grandstaff DE. Differences between geothermal and experimentally derived fluids: how well do hydrothermal experiments model the composition of geothermal reservoir fluids? *Geochim Cosmochim Acta*. 1989;53:343–58.
- Karingithi CW, Arnórsson S, Grönvold K. Processes controlling aquifer fluid compositions in the Olkaria geothermal system. *Kenya J Volcanol Geotherm Res*. 2010;196:57–76.
- Koga A, Kita I, Hikino E, Nitta K, Taguchi S. New gas geothermometers using CO<sub>2</sub>/H<sub>2</sub> and CH<sub>4</sub>/H<sub>2</sub> ratios. *J Geother Res Soc of Japan*. 1995;17:201–211. [https://www.jstage.jst.go.jp/article/grsj/1979/17/3/17\\_3\\_201\\_.pdf](https://www.jstage.jst.go.jp/article/grsj/1979/17/3/17_3_201_.pdf). Accessed 20 Nov 2020
- Laksminingpuri N, Martinus A. Studi Kandungan Dan Temperatur Gas Panas Bumi Kamojang Dengan Diagram Grid. *Beta Gamma Tahun*. 2013;4:69–79. <http://jurnal.batan.go.id/index.php/BetaGamma/article/download/1502/1431>. Accessed 20 Nov 2020
- Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. *Appl Energy*. 2010;87:2313–20.
- López-Mendiola JM, Munguía F. Evidencias geoquímicas del fenómeno de ebullición en el campo de Los Humeros. *Geotermia*. 1989;5:89–106. <https://colecciondigital.cemiegeo.org/xmlui/handle/123456789/2062>. Accessed 20 Nov 2020
- Minissale A, Evans WC, Magro G, Vaselli O. Multiple source components in gas manifestations from north-central Italy. *Chem Geol*. 1997;142:175–92.
- Moya D, Aldás C, Kaparaju P. Geothermal energy: Power plant technology and direct heat applications. *Renew Sustain Energy Rev*. 2018;94:889–901.
- Nehring N, D'Amore F. Gas chemistry and thermometry of the Cerro Prieto, Mexico, geothermal field. *Geothermics*. 1984;13:75–89.

- Nicholson K. Geothermal fluids: Chemistry and Exploration Techniques. Berlin, Alemania, Springer-Verlag. 1993;263. <https://www.springer.com/gp/book/9783642778469>. Accessed 20 Nov 2020
- Nieva D, Fausto J, González J, Garibaldi F. Flow of vapor into the production zone of Cerro Prieto I wells. In: Proceedings Fourth Symposium on the Cerro Prieto Geothermal Field, Baja California, México. 1982;2:455–461. <https://www.osti.gov/servlets/purl/7369515>. Accessed 20 Nov 2020
- Nieva D, Gonzales J, Garfias A. Evidence of two extreme flow regimes operating in the production zone of different wells from Los Azufres. In: Proceedings Tenth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA. 1985;233–240. <https://www.osti.gov/servlets/purl/892539>. Accessed 20 Nov 2020
- Nieva D, Verma M, Santoyo E, Barragan RM, Portugal E. Chemical and isotopic evidence of steam upflow and partial condensation in Los Azufres reservoir. In: Proceedings Twelfth Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA. 1987;253–260. <https://www.osti.gov/servlets/purl/888553/>. Accessed 20 Nov 2020
- Nieva D, Barragán RM, Arellano V. Geochemistry of Hydrothermal Systems. In: Bronicki L, editor. Power stations using locally available energy sources. Encyclopedia of sustainability science and technology series. New York: Springer; 2018.
- Pandarinath K, Pérez-Barrera J, Pérez-Orozco JP. GasGeo—software to estimate the reservoir temperatures of geothermal systems using gas geothermometers. In: Proceedings XXI National Geochemistry Congress 2011, Actas INAGEQ 17. 2011.
- Pang Z. Isotope and chemical geothermometry and its applications. *Sci China Technol Sci*. 2001;44:6–20. <https://doi.org/10.1007/BF02916784.pdf>.
- Pérez-Zárate D, Santoyo E, Acevedo-Anicasio A, Díaz-González L, García-López C. Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. *Comput Geosci*. 2019;129:49–68.
- Porkhial S, Salehpour M, Ashraf H, Jamali A. Modeling and prediction of geothermal reservoir temperature behavior using evolutionary design of neural networks. *Geothermics*. 2015;53:320–7.
- Poulton MM. Computational Neural Networks for Geophysical Data Processing. Pergamon Press, Amsterdam. 2001;30:335. <https://www.elsevier.com/books/computational-neural-networks-for-geophysical-data-processing/poulton/978-0-08-043986-0>. Accessed 20 Nov 2020
- Powell T. A review of exploration gas geothermometry. In: Proceedings 25th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA. 2000;9. <https://pangea.stanford.edu/ERE/pdf/IGAstANDARD/SGW/2000/Powell.pdf>. Accessed 20 Nov 2020
- Powell T, Cumming W. Spreadsheets for geothermal water and gas geochemistry. In: Proceedings 35th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA. 2010;10. <https://pangea.stanford.edu/ERE/pdf/IGAstANDARD/SGW/2010/powell.pdf>. Accessed 20 Nov 2020
- Renderos RE. Chemical characterization of the thermal fluid discharge from well production tests in the Berlin geothermal field, El Salvador: geothermal training program. The United Nations University. 2002;2:205–231. <https://orkustofnun.is/gogn/unu-gtp-report/UNU-GTP-2002-12.pdf>. Accessed 20 Nov 2020
- Santoyo E, Verma SP, Nieva D, Portugal E. Variability in the gas phase composition of fluids discharged from Los Azufres geothermal field. *Mexico J Volcanol Geotherm Res*. 1991;47:161–81.
- Santoyo-Castelazo E. Sustainability assessment of electricity options for Mexico: current situation and future scenarios. Ph. D. Dissertation, The University of Manchester, United Kingdom, 2011:286. [https://www.research.manchester.ac.uk/portal/files/54515414/FULL\\_TEXT.PDF](https://www.research.manchester.ac.uk/portal/files/54515414/FULL_TEXT.PDF). Accessed 20 Nov 2020
- Santoyo-Castelazo E, Gujba H, Azapagic A. Life cycle assessment of electricity generation in Mexico. *Energy*. 2011;36:1488–99.
- Santoyo-Castelazo E, Azapagic A. Sustainability assessment of energy systems: integrating environmental, economic and social aspects. *J Clean Prod*. 2014;80:119–38.
- Saracco L, D'Amore F. CO2B: a computer program for applying a gas geothermometer to geothermal systems. *Comput Geosci*. 1989;15:1053–65.
- Serpen G, Palabiyik Y, Serpen U. An artificial neural network model for Na/K geothermometer. In: Proceedings 34th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA. 2009;12. <https://pangea.stanford.edu/ERE/pdf/IGAstANDARD/SGW/2009/serpen.pdf>. Accessed 20 Nov 2020
- Spycher N, Peiffer L, Finsterle S, Sonnenthal E. GeoT User's Guide, A Computer Program for Multicomponent Geothermometry and Geochemical Speciation, Version 2.1. 2016. <https://escholarship.org/content/qt8hs3b99h/qt8hs3b99h.pdf>. Accessed 20 Nov 2020
- Stefánsson A. Gas chemistry of Icelandic thermal fluids. *J Volcanol Geotherm Res*. 2017;346:81–94.
- Supranto S, Budianto T, Djoko W, Idrus A. Proposed empirical gas geothermometer using multidimensional approach. In: proceedings twenty first workshop on geothermal reservoir engineering, Stanford University, Stanford, California, USA. 1996;195–199. <https://pangea.stanford.edu/ERE/pdf/IGAstANDARD/SGW/1996/Supranto.pdf>. Accessed 20 Nov 2020
- Tello E, Verma MP, Tovar R. Origin of acidity in the Los Humeros, Mexico, geothermal reservoir. In: Proceedings World Geothermal Congress 2000; Japan. 2000. <https://www.geothermal-energy.org/pdf/IGAstANDARD/WGC/2000/R0081.PDF>. Accessed 20 Nov 2020
- Theil H. Economic Forecasts and Policy. North-Holland Press. 1961;567.
- Thien BM, Kosakowski G, Kulik DA. Differential alteration of basaltic lava flows and hyaloclastites in Icelandic hydrothermal systems. *Geotherm Energy*. 2015;3:1–32.
- Tonani F. Equilibria that control the hydrogen content of geothermal gases. *Bull Volcanol*. 1973;44:547–64.
- Verma SP, Santoyo E. New improved equations for Na/K, Na/Li and SiO<sub>2</sub> geothermometers by outlier detection and rejection. *J Volcanol Geotherm Res*. 1997;70:9–23.
- Verma SP, Pandarinath K, Santoyo E, González-Partida E, Torres-Alvarado IS, Tello-Hinojosa E. Fluid chemistry and temperatures prior to exploitation at the Las Tres Vírgenes geothermal field. *Mexico Geothermics*. 2006;35:156–80.

- Verma SP, Pandarinath K, Santoyo E. SolGeo: A new computer program for solute geothermometers and its application to Mexican geothermal fields. *Geothermics*. 2008;37:597–621.
- Wamalwa RN. Evaluation of factors controlling the concentration of non-condensable gases and their possible impact on the performance of wells in Olkaria, Kenya. Geothermal Training Programme, United Nations University, Reykjavik, Iceland. 2015;787–808. <https://orkustofnun.is/gogn/unu-gtp-report/UNU-GTP-2015-34.pdf>. Accessed 20 Nov 2020
- Wang JJ, Jing YY, Zhang CF, Zhao JH. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renew Sustain Energy Rev*. 2009;13:2263–78.
- Wang W, Lu Y. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *Mater Sci Eng*. 2018;324:1–10. <https://doi.org/10.1088/1757-899X/324/1/012049/pdf>.
- Willmott CJ, Matsuura K, Robeson SM. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos Environ*. 2009;43:749–52.
- Wu Y, Li P. The potential of coupled carbon storage and geothermal extraction in a CO<sub>2</sub> enhanced geothermal system: a review. *Geotherm Energy*. 2020;8:1–28.
- Yan-guang L, Bing L, Chuan L, Xi Z, Gui-ling W. Reconstruction of deep fluid chemical constituents for estimation of geothermal reservoir temperature using chemical geothermometers. *J Groundw Sci Eng*. 2017;5:173–181. <http://gwse.ihg.org.cn/article/id/271>. Accessed 20 Nov 2020

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---



# ANEXO 3

# Evaluation of Artificial Neural Networks and Eddy Covariance Measurements for Modelling the CO<sub>2</sub> Flux Dynamics in the Acoculco Geothermal Caldera (Mexico)

E. Santoyo, A. Acevedo-Anicasio, D. Pérez-Zarate, and M. Guevara

**Abstract**—The aim of this research work is to report the CO<sub>2</sub> measurements of a short-term period carried out in the geothermal zone of Acoculco, Puebla (Mexico). CO<sub>2</sub> measurements were logged using a micrometeorological station Eddy Covariance. According to the complex geochemical phenomena involved in these measurements, artificial neural networks have been used for reproducing the CO<sub>2</sub> measurements, and to fill the gap issues during the monitoring period. This survey was also performed for the identification of CO<sub>2</sub> anomalies in the zone, and also to determine the natural emission baseline of CO<sub>2</sub> at the early exploration stage of this promissory geothermal system. Details of this evaluation study are outlined.

**Index Terms**—Enhanced geothermal systems, environmental sustainability, geothermal energy, hot-dry rock, soil-gas emission.

## I. INTRODUCTION

Geochemical knowledge of the carbon dioxide (CO<sub>2</sub>) flux dynamics into and out of the atmosphere is essential for understanding carbon sink and sources of magmatic or geothermal systems, among other terrestrial ecosystems [1], [2].

Field monitoring programmes, mapping, and modelling of CO<sub>2</sub> fluxes in such systems are suggested as suitable geochemical tasks for the prospection of promissory geothermal zones, and for a better understanding of their natural contribution to the global carbon budget.

The determination of the background or natural baseline gas emissions existing in such systems has been identified as an important challenge to be achieved. This crucial information may be useful for geothermal industry decision makers to install future commercial projects for the electricity generation and other direct uses.

The modelling of CO<sub>2</sub> flux dynamics in large covering areas requires the use of suitable measuring techniques together with simulation models for addressing some technical and scientific issues.

Eddy covariance (EC) technique has been proposed as

Manuscript received June 10, 2018; revised July 14, 2018. This work was funded by the research project: P09 CeMIE-Geo (Project 207032 CONACyT-SENER, Mexico).

E. Santoyo and M. Guevara are with the Institute for Renewable Energy (UNAM), Temixco, Mexico (e-mails: esg@ier.unam.mx, mygg@ier.unam.mx).

A. Acevedo-Anicasio is with the PhD Programme of Computational Sciences, Faculty of Sciences (UAEMor), 62209 Mexico (e-mail: agustin.acevedoa@uaem.edu.mx).

D. Perez-Zarate is with the Institute of Geophysics (UNAM), CONACyT Catedra, 04510 Mexico (e-mail: depez@igeofisica.unam.mx).

geochemical tool for monitoring the net exchange rate of CO<sub>2</sub> through the interface between the atmosphere and the surface emission, e.g., biogenic (plant canopy) or soil-gas emitters. EC has the capability to provide automated CO<sub>2</sub> flux measurements without ground surface interferences, which are averaged with time, and recorded from a larger spatial scale (m<sup>2</sup> to km<sup>2</sup>) depending on the EC tower height. However, it is also recognized some technical limitations of the EC technique either to find out a good agreement with CO<sub>2</sub> flux simulations or to face out some instrumental problems related to the measuring gaps (i.e., loss of data during continuous monitoring programmes) [3].

The non-linearity among CO<sub>2</sub> fluxes and other meteorological flux parameters (e.g., energy fluxes, wind velocity and direction patterns, which are also measured by EC) have limited the applicability of complex theoretical (mechanistic approaches) and empirical (curve fitting or regression) CO<sub>2</sub> flux models to predict the flux dynamics with accuracy [3]. Artificial neural networks (ANN) have been proposed as a multivariate modelling tool for identifying complex non-linear relationships between input and output variables without a comprehensive explanation of the actual physical nature of the phenomena, which are difficult to study by conventional techniques [4].

In the present study, we have carried out a field measurement campaign of CO<sub>2</sub> fluxes in a new promissory geothermal zone of Mexico (known as Caldera of Acoculco, Puebla: see location in Fig. 1) by using an EC micrometeorology station. The applicability of ANN architectures for the prediction of CO<sub>2</sub> fluxes from other meteorological input variables was evaluated. The capability of the best ANN architectures was used as a gap filling tool to solve the loss of information caused by local climate issues occurred during the EC monitoring programme. Details of this evaluation study are outlined.

## II. GEOLOGICAL SETTING

The Acoculco Caldera belongs to the Eastern part of the Mexican Volcanic Belt (MVB) (Fig. 2). The MVB is the largest Neogene volcanic arc in North America, covering ~160,000 km<sup>2</sup> and a length of ~1,000 km between 18°30' and 21°30' N in central Mexico. The Acoculco Caldera (18 km in diameter) is associated to a volcanism episode (1.7–0.24 Ma) that occurred within the older and larger Tulancingo Caldera [5], [6].

Two exploratory geothermal wells (EAC-1 and EAC-2) were drilled in Los Azufres zone (N Lat. 19° 55' 29.4"; W Long. 98° 08' 39.9"; and altitude 2,839 m.a.s.l.) at a depth of

~2 km, where a hot-dry rock temperature of 300°C was measured without a fluid reservoir evidence (i.e., with a very poor rock-permeability) [7].

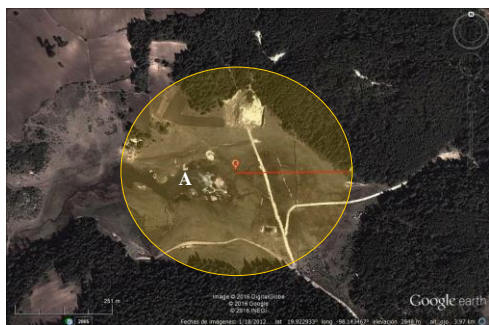


Fig. 1. A Google Earth photography of Los Azufres area in the geothermal Caldera of Acoculco, Puebla.

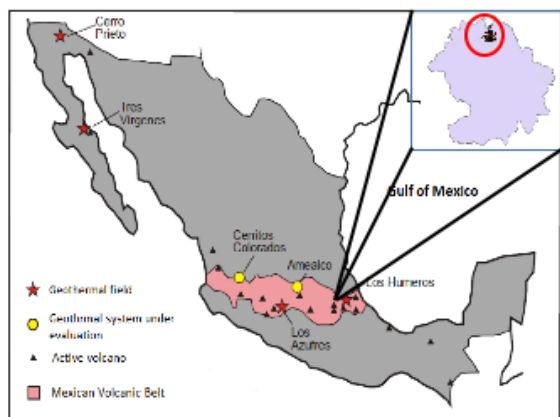


Fig. 2. Geographical Location of the Geothermal Zone of Acoculco, Puebla (M éxico).

### III. EXPERIMENTAL

A micrometeorological station of Eddy Covariance (EC) was successfully installed in the Los Azufres zone of the Caldera of Acoculco, Puebla (Fig. 3).



Fig. 3. A drone landscape photography of Los Azufres area in the geothermal Caldera of Acoculco, Puebla.

EC technique was used in this exploration study to determine the net exchange rate of CO<sub>2</sub> across the interface between the atmosphere and a geothermal ecosystem (plant canopy and natural soil-gas emissions) by measuring the covariance between fluctuations in vertical wind velocity and CO<sub>2</sub> mixing ratio.

This method is most reliable when the atmospheric conditions (wind, temperature, humidity, soil-gas emissions) are steady, the underlying plants is homogeneous, and installed on a flat topography for an extended distance

upwind [1], [8]. EC is an instrument whereby high frequency measurements of atmospheric CO<sub>2</sub> are recorded at a height of 3 m above ground by an infrared gas analyser (IRGA), accompanied by other detectors for sensing some other meteorological variables (e.g., wind velocity, air temperature, relative humidity, among others).

EC technique was used after assuming spatial homogeneity of surface CO<sub>2</sub> fluxes (F<sub>CO<sub>2</sub></sub>), a flat terrain, and temporal stationary conditions which were all fulfilled in the geothermal zone of Acoculco. A gross conservation of energy and mass over land area is generally provided by these measurements (known as EC footprint) from which net F<sub>CO<sub>2</sub></sub> are usually determined [9].

### IV. RESULTS AND DISCUSSION

**Experimental measurements.** A comprehensive CO<sub>2</sub> short-term monitoring programme based on the use of the EC technique was carried out during the dry season period: n=1,766 (15/03/2016 to 31/05/2016). Half-hourly measurements of F<sub>CO<sub>2</sub></sub> were recorded in the field by using an EC station for approximately 3 months. Positive F<sub>CO<sub>2</sub></sub> measured with the EC micrometeorological station ranged from ~0.001 to 200 μmol m<sup>-2</sup>d<sup>-1</sup> with a mean and standard deviation of 5 and 11 μmol m<sup>-2</sup>d<sup>-1</sup>, respectively.

Measurement gaps in the time series of F<sub>CO<sub>2</sub></sub> were detected due to some local climatological changes mostly occurring during the night time (e.g., fogs, water condensation). Large diurnal to seasonal variations of soil-gas (CO<sub>2</sub>) concentrations, surface CO<sub>2</sub> fluxes, and total CO<sub>2</sub> discharges seem to be associated with local meteorological and hydrological processes.

Maximum F<sub>CO<sub>2</sub></sub> up to 200 μmol m<sup>-2</sup>d<sup>-1</sup> were mainly registered in March, whereas for the April and May, the recorded fluxes varied up to 111 μmol m<sup>-2</sup>d<sup>-1</sup> and 140 μmol m<sup>-2</sup>d<sup>-1</sup>, respectively (Fig. 4).

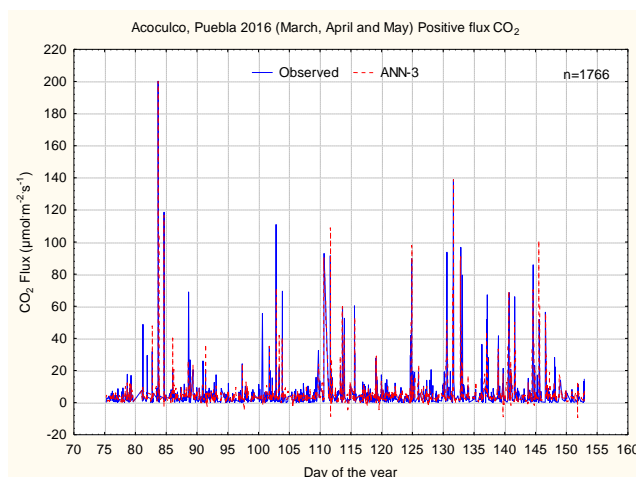


Fig. 4. Time series plot of CO<sub>2</sub> fluxes measured in the geothermal Caldera of Acoculco, Puebla (Mexico): ANN-3.

A comprehensive CO<sub>2</sub> short-term monitoring programme based on the use of the EC technique was carried out during the dry season period: n=1,766 (15/03/2016 to 31/05/2016). Half-hourly measurements of F<sub>CO<sub>2</sub></sub> were recorded in the field by using an EC station for approximately 3 months. Positive F<sub>CO<sub>2</sub></sub> measured with the EC micrometeorological station ranged from ~0.001 to 200 μmol m<sup>-2</sup>d<sup>-1</sup> with a mean and

standard deviation of 5 and 11  $\mu\text{mol m}^{-2}\text{d}^{-1}$ , respectively.

Measurement gaps in the time series of  $F_{\text{CO}_2}$  were detected due to some local climatological changes mostly occurring during the night time (e.g., fogs, water condensation). The large diurnal to seasonal variations of soil-gas ( $\text{CO}_2$ ) concentrations, surface  $\text{CO}_2$  fluxes, and total  $\text{CO}_2$  discharges seem to be associated with local meteorological and hydrological processes. Maximum  $F_{\text{CO}_2}$  up to 200  $\mu\text{mol m}^{-2}\text{d}^{-1}$  were mainly registered in March, whereas for the April and May, the recorded fluxes varied up to 111  $\mu\text{mol m}^{-2}\text{d}^{-1}$  and 140  $\mu\text{mol m}^{-2}\text{d}^{-1}$ , respectively (Fig. 4).

**ANN simulation.** A close correlation among  $F_{\text{CO}_2}$  and energy fluxes (net radiation, RN; sensible, H and latent, LE heats; and evapotranspiration, ET),  $\text{H}_2\text{O}$  vapour fluxes ( $\text{H}_2\text{O\_Flux}$ ), and temperatures (air, TA; dew,  $T_{\text{dew}}$ ) was mainly found. Using the commercial Matlab software a network was designed by using the experimental EC responses.

The EC measurements were used as input data for the evaluation of all the ANN architectures, which were randomly divided into training (50%), testing (25%), and validation (25%) data sets. The ANN architectures used feed-forward models [10], which were characterized by an input layer (with an adjustable number of input neurons), a hidden layer (with an adaptable number of hidden neurons), and an output layer with the target ( $F_{\text{CO}_2\text{-Measured}}$ ) and the predicted output ( $F_{\text{CO}_2\text{-ANN}}$ ). In this way, a total number of 3,240 ANN architectures were evaluated. Table I shows a summary of the best ANN architectures used in this study.

TABLE I: ANN INPUT-OUTPUT DATA ARCHITECTURES

ANN	Input Layer	Hidden Layer	Output
1	6 (RN, LE, H, TA, TS, $T_{\text{dew}}$ )	8	1 ( $F_{\text{CO}_2}$ )
2	9 (RN, LE, H, TA, TS, SWC, $\text{H}_2\text{O\_Flux}$ , ET, $T_{\text{dew}}$ )	11	1 ( $F_{\text{CO}_2}$ )
3	8 (RN, LE, H, TA, SWC, $\text{H}_2\text{O\_Flux}$ , ET, $T_{\text{dew}}$ )	11	1 ( $F_{\text{CO}_2}$ )
4	6 (RN, LE, H, TA, TS, $T_{\text{dew}}$ )	14	1 ( $F_{\text{CO}_2}$ )

RN: Net radiation; H: Sensible heat; LE: Latent heat; ET: Evapotranspiration;  $\text{H}_2\text{O\_Flux}$ :  $\text{H}_2\text{O}$  vapour fluxes; TA: air temperature;  $T_{\text{dew}}$ : Dew temperature; TS: Soil temperature; SWC: soil water content.

As the nature of the EC data patterns were so complex, the number of the hidden layers were monitored for avoiding the well-known over-fitting or over-learning problems of the ANN. For all the ANN architectures, the activation function used for the hidden layer was the hyperbolic tangent sigmoid transfer function (tansig), whereas for the output layer, the linear function (purelin) was used.

Multivariate statistical analysis and log-ratio transformations were used for normalizing the input variables. As a part of the methodology used, the input variable normalization, the use of learning rate factors, and the design of ANN architectures were together evaluated. The Levenberg-Marquardt algorithm, hyperbolic tangent sigmoid and linear transfer functions were used to optimize the ANN learning.

The network architectures were trained and their performance expressed, as the linear correlation coefficient ( $r$ )

between the target ( $F_{\text{CO}_2\text{-Measured}}$ ) and the predicted ( $F_{\text{CO}_2\text{-ANN}}$ ) variables were recursively evaluated, as well as for monitoring the over-fitting problems among the training, testing, and validation stages. The prediction capability of the ANN architectures were also evaluated through the calculation of classical statistical parameters: the root mean squares of errors ( $RMSE$ ) and the mean absolute error ( $MAE$ ).  $RMSE$  is a quadratic scoring rule that measures the average magnitude of the errors; whereas  $MAE$  measures the average magnitude of the errors in a set of estimations. Ideal values of  $RMSE$  and  $MAE$  should indicate a perfect prediction from the ANN model.

The  $RMSE$  is given by the following equation:

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (FCO2_{measured} - FCO2_{predicted})^2}{n}}$$

where  $n$  is the number of EC measurements; whereas the  $MAE$  parameter is calculated by means of the following expression:

$$MAE = \frac{\sum_{k=1}^n |FCO2_{predicted} - FCO2_{measured}|}{n}$$

TABLE II: PREDICTION RESULTS ACHIEVED BY THE BEST ANN ARCHITECTURES EVALUATED IN THIS STUDY

ANN	$r$				RMSE	MAE
	Global	Training	Validation	Testing		
1	0.8730	0.941	0.682	0.821	5.453	2.733
2	0.8760	0.951	0.706	0.778	5.343	2.701
3	0.8725	0.949	0.722	0.742	5.321	2.583
4*	0.8903	0.910	0.890	0.860	5.050	2.810

\*These results were obtained when the ANN was slightly modified with a different input data array: 60% for training, 20% for validation, and 20% for testing.

As can be observed in Table II, the prediction results ( $F_{\text{CO}_2\text{-ANN}}$ ) inferred from some ANN architectures showed an acceptable matching with  $F_{\text{CO}_2\text{-Measured}}$  values, having the best results for the ANN-3 (which had 8 input neurons and 11 neurons for the hidden layer) when the design was characterized by 50% for training, 25% for validation, and 25% for testing (Fig. 4); whereas for a slightly different input data array, the ANN-4 provided the best prediction results with consistently  $r$  values for the training, validation and testing stages, which avoid the ANN over-fitting problem (see Fig. 5).

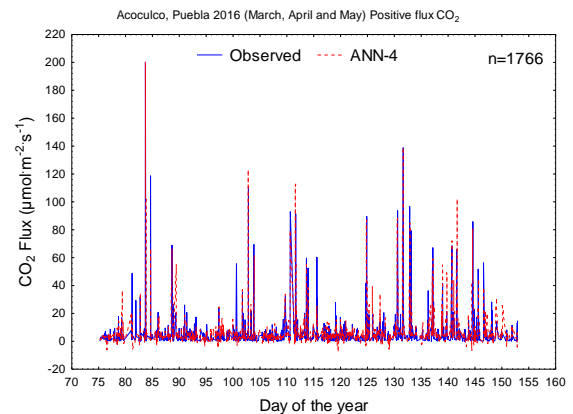


Fig. 5. Time series plot of  $\text{CO}_2$  fluxes measured in the geothermal Caldera of Acoculco, Puebla (Mexico): ANN-4.

The correlation coefficient ( $r$ ) of a linear regression between the target ( $F_{CO_2-Measured}$ ) and the predicted ( $F_{CO_2-ANN}$ ) variables was used for evaluating the prediction performance of the ANNs (Figs. 6 and 7).

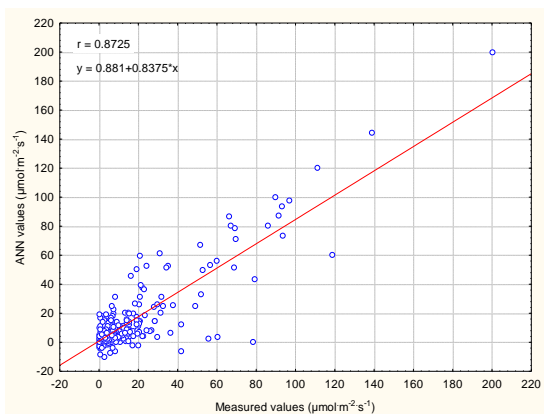


Fig. 6. Linear regression results obtained between  $F_{CO_2-Measured}$  and  $F_{CO_2-ANN}$  for the ANN-3 architecture.

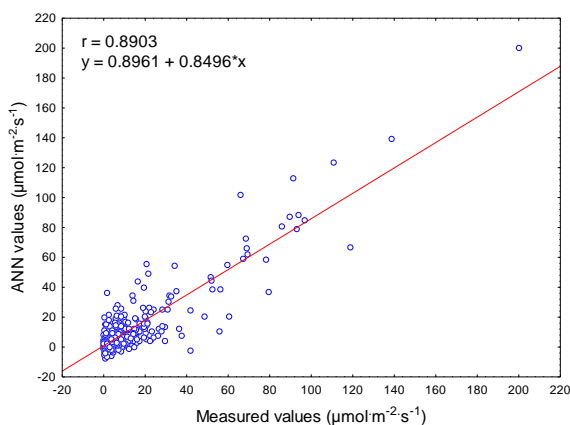


Fig. 7. Linear regression results obtained between  $F_{CO_2-Measured}$  and  $F_{CO_2-ANN}$  for the ANN-4 architecture.

A sensitivity analysis was finally performed by using the Garson method [11] to evaluate the relative importance of input variables on the output target ( $F_{CO_2-ANN}$ ).

Table III summarises the sensitivity results obtained for the best ANN architectures.

TABLE III: SENSITIVITY RESULTS OBTAINED BY THE BEST ANN ARCHITECTURES EVALUATED IN THIS STUDY

ANN	Input variables (%)								
	RN	LE	H	TA	TS	T <sub>de</sub> <sub>w</sub>	SW C	H <sub>2</sub> O F	ET
1	5.3	31.6	9.9	24.1	18.5	10.6	----	----	----
2	14.2	14.9	11.0	8.7	10.8	6.6	9.0	14.4	10.4
3	5.3	12.2	11.1	13.6	----	17.5	5.8	16.3	18.2
4	16.9	34.6	14.6	12.6	12.1	9.2	----	----	----

For the ANN-3, the sensitivity results showed that the evapotranspiration (ET) was the input variable with the larger weight percentage (18.2%); whereas the less relative importance variable was attributed to the net radiation (RN: 5.3%). On the other hand for the ANN-4, the most important input variable was the latent heat (LE: 34.6%), whereas the less important variable was the dew temperature (T<sub>dew</sub>:

9.2%).

## V. CONCLUSION

Artificial neural networks were successfully evaluated for the multivariate analysis among  $F_{CO_2}$  and micrometeorological variables (energy fluxes, and  $H_2O$  vapour fluxes, and temperatures) measurements carried out in the promissory geothermal caldera of Acoculco, Puebla (Mexico). The non-linearity of the complex relationships among  $CO_2$  fluxes and the micrometeorological variables were positively reproduced. The best ANN-4 simulation results of the study show that predicted  $CO_2$  fluxes closely matches the field measurements with a global linear correlation coefficient of  $r=0.8903$ . These results enable the measurement gap issues to be filled with reasonable accuracy by using the input variable weighting coefficients of each ANN equation. The present study demonstrates the efficient capability of applying ANN tools for modelling the complex  $CO_2$  flux dynamics in geothermal systems.

## ACKNOWLEDGEMENT

This work was funded by the research project: P09 CeMIE-Geo (Project 207032 CONACyT-SENER, Mexico).

## REFERENCES

- [1] A. M. Melesse and R. S. Henley, *Ecolog. Modelling*, vol. 189, 2005, p. 305.
- [2] J. L. Lewicki, M. L. Fischer, and G. E. Hilley, *J. Volcanol. Geother. Res.*, vol. 171, 2008, p. 178.
- [3] M. Ryan, C. Muller, H. J. Di, and K. C. Cameron, *Ecolog. Modelling*, vol. 175, 2004, p. 189.
- [4] A. Akkala, V. Devabhaktuni, A. Kumar, and D. Bhatt, *The Open Environ. & Biol. Monitoring J.*, vol. 4, 2011, p. 21.
- [5] G. Sosa-Ceballos, J. L. Mac ás, D. R. Avellán, N. Salazar-Hermenegildo, M. E. Boijseaneau-López, J. Pérez-Orozco, and J. Volcanol, *Geother. Res.*, 2018.
- [6] A. Garc í-Palomo, J. L. Mac ás, A. Jim énez, G. Tolson, M. Mena, J. M. Sánchez-Núñez, J. L. Arce, P. W. Layer, M. A. Santoyo, and J. Lermo-Samaniego, *J. Volcanol. Geother. Res.*, vol. 349, 2018, p. 218.
- [7] A. López-Hern ández, G. Garc í-Estrada, G. Aguirre-D íz, E. González-Partida, H. Palma-Guzmán, and J. L. Quijano-Le án, *Geothermics*, 38, 2009, p. 279.
- [8] D. D. Baldocchi, B. B. Hicks, T. P. Meyers, *Ecolog.*, vol. 69, 1988, p. 1331.
- [9] M. Ooba, T. Hirano, J. I. Mogami, R. Hirata, Y. Fujinuma, *Ecolog. Modelling*, 198 (2006) p. 473.
- [10] M. M. Poulton, *Computational Neural Networks for Geophysical Data*, Elsevier, Amsterdam, 2001.
- [11] G. D. Garson, *Artif. Intel. Expert*, vol. 6, 1991, p. 47.



**Edgar R. Santoyo Gutierrez** is a PhD in thermal energy, University of Salford (UK). He is the leader of the Geoenergy Research Group at the Institute for Renewable Energy, Universidad Nacional Autónoma de México (UNAM). His major research lines geothermal energy, geochemistry, water-rock interaction, statistics, geochemometrics and environmental sustainability (LCA).



**Agustín Acevedo-Anicacio** is a Ph.D student in computational sciences, Faculty of Sciences, Universidad Autonoma del Estado de Morelos. His research lines geothermal energy, artificial neural networks, artificial intelligence.



**Daniel Perez-Zarate** is CONACyT Chair Fellowship, researcher at the Institute of Geophysics, Universidad Nacional Autonoma de México (UNAM). His research lines geothermal energy, water-rock interaction, geochemistry and geochemometrics.



**Mirna Guevara Garcia** is academic of the Institute for Renewable Energy, Universidad Nacional Autónoma de México (UNAM). Her research lines geothermal energy, geochemistry and geochemometrics.