Evaluation of Artificial Neural Networks and Eddy Covariance Measurements for Modelling the CO_2 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_2 measurements of a short-term period carried out in the geothermal zone of Acoculco, Puebla (Mexico). CO_2 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_2 measurements, and to fill the gap issues during the monitoring period. This survey was also performed for the identification of CO_2 anomalies in the zone, and also to determine the natural emission baseline of CO_2 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_2) 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_2 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_2 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 geochemical tool for monitoring the net exchange rate of CO_2 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_2 flux measurements without ground surface interferences, which are averaged with time, and recorded from a larger spatial scale (m^2 to km^2) 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_2 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_2 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_2 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_2 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_2 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 km2 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

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).

doi: 10.18178/ijesd.2018.9.10.1117

~2 km, where a hot-dry rock temperature of 300^{o}C was measured without a fluid reservoir evidence (i.e., with a very poor rock-permeability) [7].

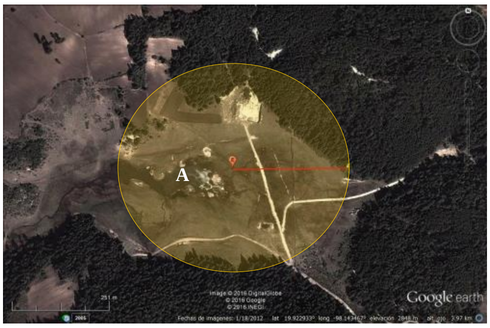

Fig. 1. A Google Earth photography of Los Azufres area in the geothermal Caldera of Acoculco, Puebla.

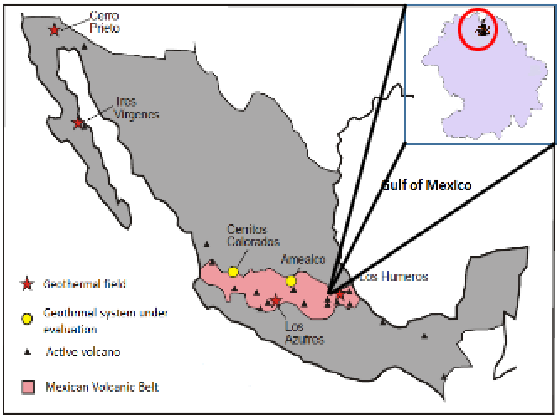

Fig. 2. Geographical Location of the Geothermal Zone of Acoculco, Puebla (México).

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_2 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_2 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_2 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_2 fluxes (F_{CO2}), 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_{CO2} are usually determined [9].

IV. Results and Discussion

Experimental measurements. A comprehensive CO_2 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_{CO2} were recorded in the field by using an EC station for approximately 3 months. Positive F_{CO2} measured with the EC micrometeorological station ranged from ~0.001 to 200 μmol $m^{-2}d^{-1}$ with a mean and standard deviation of 5 and 11 μmol $m^{-2}d^{-1}$, respectively.

Measurement gaps in the time series of F_{CO2} 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_2) concentrations, surface CO_2 fluxes, and total CO_2 discharges seem to be associated with local meteorological and hydrological processes.

Maximum F_{CO2} up to 200 μmol $m^{-2}d^{-1}$ were mainly registered in March, whereas for the April and May, the recorded fluxes varied up to 111 μmol $m^{-2}d^{-1}$ and 140 μmol $m^{-2}d^{-1}$, respectively (Fig. 4).

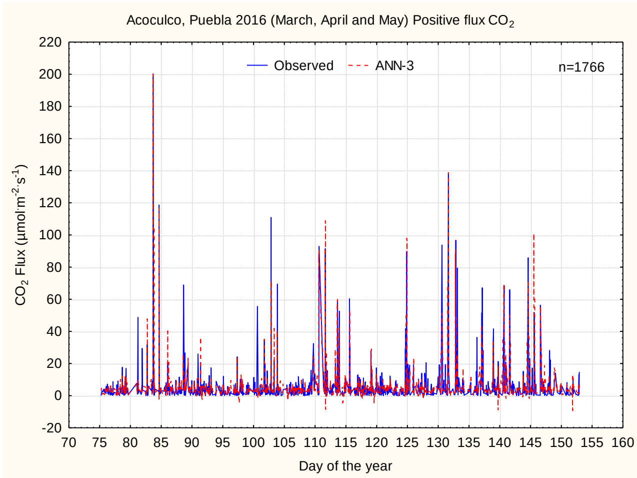

Fig. 4. Time series plot of CO2 fluxes measured in the geothermal Caldera of Acoculco, Puebla (Mexico): ANN-3.

A comprehensive CO_2 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_{CO2} were recorded in the field by using an EC station for approximately 3 months. Positive F_{CO2} measured with the EC micrometeorological station ranged from ~0.001 to 200 μmol $m^{-2}d^{-1}$ with a mean and standard deviation of 5 and 11 μmol $m^{-2}d^{-1}$, respectively.

Measurement gaps in the time series of F_{CO2} 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 (CO_2) concentrations, surface CO_2 fluxes, and total CO_2 discharges seem to be associated with local meteorological and hydrological processes. Maximum F_{CO2} up to 200 μmol $m^{-2}d^{-1}$ were mainly registered in March, whereas for the April and May, the recorded fluxes varied up to 111 μmol $m^{-2}d^{-1}$ and 140 μmol $m^{-2}d^{-1}$, respectively (Fig. 4).

ANN simulation. A close correlation among F_{CO2} and energy fluxes (net radiation, RN; sensible, H and latent, LE heats; and evapotranspiration, ET), H_2O vapour fluxes (H_2O_Flux), and temperatures (air, TA; dew, T_{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 ($F_{CO2\text{-}Measured}$) and the predicted output ($F_{CO2\text{-}ANN}$). 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_{dew})	8	1 (F_{CO2})
2	9 (RN, LE, H, TA, TS, SWC, H2O_Flux, ET, T_{dew})	11	1 (F_{CO2})
3	8 (RN, LE, H, TA, SWC, H2O_Flux, ET, T_{dew})	11	1 (F_{CO2})
4	6 (RN, LE, H, TA, TS, T_{dew})	14	1 (F_{CO2})

RN: Net radiation; H: Sensible heat; LE: Latent heat; ET: Evapotranspiration; H_2O_Flux: H_2O vapour fluxes; TA: air temperature; T_{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 ($F_{CO2\text{-}Measured}$) and the predicted ($F_{CO2\text{-}ANN}$) 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:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}\left(FCO2_{measured} - FCO2_{predicted}\right)^2}{n}}$$

where n is the number of EC measurements; whereas the *MAE* parameter is calculated by means of the following expression:

$$MAE = \frac{\sum_{k=1}^{n}\left|FCO2_{predicted} - FCO2_{measured}\right|}{n}$$

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 ($F_{CO2\text{-}ANN}$) inferred from some ANN architectures showed an acceptable matching with $F_{CO2\text{-}Measured}$ 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).

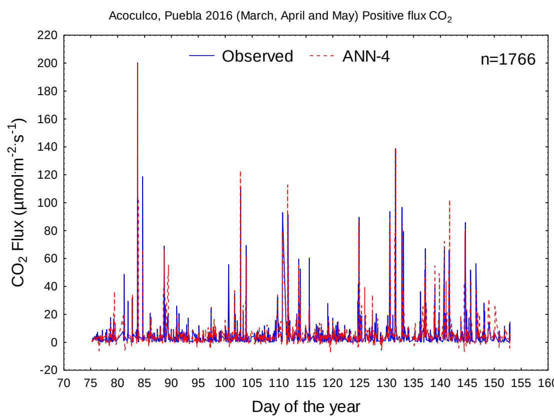

Fig. 5. Time series plot of CO_2 fluxes measured in the geothermal Caldera of Acoculco, Puebla (Mexico): ANN-4.

The correlation coefficient (*r*) of a linear regression between the target ($F_{CO2\text{-}Measured}$) and the predicted ($F_{CO2\text{-}ANN}$) variables was used for evaluating the prediction performance of the ANNs (Figs. 6 and 7).

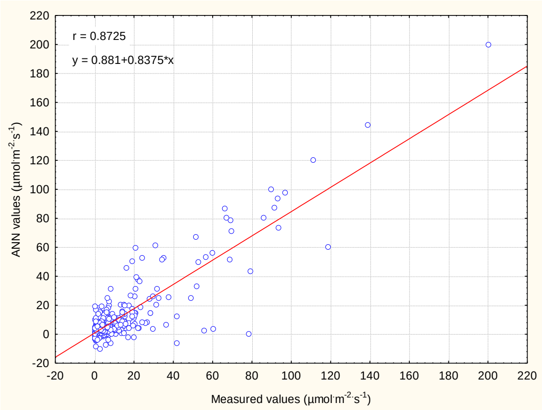

Fig. 6. Linear regression results obtained between $F_{CO2\text{-}Measured}$ and $F_{CO2\text{-}ANN}$ for the ANN-3 architecture.

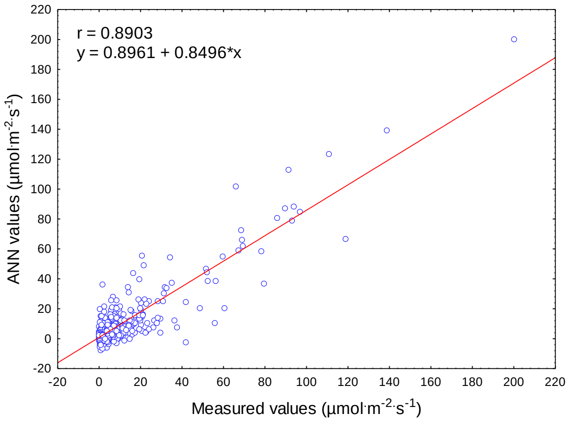

Fig. 7. Linear regression results obtained between $F_{CO2\text{-}Measured}$ and $F_{CO2\text{-}ANN}$ for the ANN-4 architecture.

A sensitivity analysis was finally performed by using the Garson method [11] to evaluate the relative importance of input variables on the output target ($F_{CO2\text{-}ANN}$).

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_{dew}	SWC	H2O_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_{dew}: 9.2%).

V. CONCLUSION

Artificial neural networks were successfully evaluated for the multivariate analysis among F_{CO2} 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. Macías, D. R. Avellán, N. Salazar-Hermenegildo, M. E. Boijseauneau-López, J. Pérez-Orozco, and J. Volcanol, *Geother. Res.*, 2018.
[6] A. García-Palomo, J. L. Macías, A. Jiménez, 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-Hernández, G. García-Estrada, G. Aguirre-Díaz, E. González-Partida, H. Palma-Guzmán, and J. L. Quijano-León, *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).

Agustin 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.