Latinmag Letters, Volume 1, Special Issue (2011), D34, 1-6. Proceedings Tandil, Argentina

A FUZZY MODEL FOR ESTIMATING HEAVY METALS IN AN ANTARTIC SOIL

Chaparro Mauro[1*], Chaparro Marcos[2], Sinito Ana[2]

[1] Instituto Multidisciplinario de Ecosistemas y Desarrollo Sustentable (UNCPBA)-CONICET, Pinto 399, B7000GHG Tandil, Argentina (mchaparr@exa.unicen.edu.ar)
[2] Instituto de Física Arroyo Seco (UNCPBA)-CONICET, Pinto 399, B7000GHG Tandil, Argentina (chapator@exa.unicen.edu.ar)

Abstract

Magnetic monitoring is well-known tool and constitutes an alternative for assessing pollutants in different environments. In this contribution, a methodology is developed for building a model of Takagi-Sugeno-Kant type and their confidence intervals. The model yields values and confidence intervals of heavy metals (Pb, Cr, Cu and Zn) from soil samples collected in Base Marambio (Antarctica)
In this work, the input variables are: specific magnetic susceptibility (χ), anhysteretic remanent magnetisation (ARM), anhysteretic susceptibility to volumetric susceptibility - ratio (κ_{ARM}/κ), and remanent coercivity (Hcr). Fuzzy c-means clustering (FCM) was used to partition data, hence the membership functions and rules are built. In addition, a model of fuzzy lineal regression was used to construct the (output) consequent functions.
The results show not only satisfactory agreement between the model and data, but also provide valued information from the rules analysis that allows us to understand the behaviour of the magnetic and heavy metal variables.

Resumen

Los estudios magnéticos constituyen una herramienta alternativa para la evaluación de contaminantes en distintos medios. En la presente contribución proponemos una metodología para construir un modelo del tipo Takagi-Sugeno-Kant y sus intervalos de confianza para la estimación de metales pesados (Pb, Cr, Ni, Cu y Zn) de muestras recolectadas en alrededores de la base Marambio en Antártida. Este tipo de modelos son utilizados cuando es posible realizar una descripción local de dinámicas complejas, en las que se involucran gran cantidad de variables e información que no es fácilmente cuantificable.
Para este estudio, las variables de entrada consideradas son: susceptibilidad magnética específica (χ), magnetización remanente anhistérica (ARM), susceptibilidad anhistérica/susceptibilidad magnética (κ_{ARM}/κ-ratio), coercitividad de remanencia (Hcr). Para la determinación de los parámetros y tipos de funciones de membresía se utilizó un cluster Fuzzy C-means (FCM) y se utilizaron modelos de regresión lineal difusa para la construcción de las funciones consecuencia del modelo.
Los resultados no solo muestran una buena estimación entre los intervalos y los datos, sino que además de las reglas obtenidas se puede extraer información del comportamiento de las variables magnéticas frente a cada uno de los metales pesados.

Introduction

Pollution is a subject of current interest and there is a need for monitoring techniques developed by several fields of research, in order to analyze the distribution and the reach around the contamination sources. Although the man-made contribution of heavy metals and other pollutants can be studied by careful chemical methods (time-consuming, laborious and costly), magnetic monitoring constitutes an alternative tool for pollution studies (Petrovský and Ellwood 1999). The relationship between both kinds of variables constitutes complex cases of non-linear mathematics. In consequence, multivariate techniques have become necessary and used to investigate the problem (Petrovský et al. 2001; Knab et al. 2001; Chaparro et al. 2008). In particular, in previous studies (Chaparro et al. 2006; 2007; 2008, 2011a), multivariate statistical analyses were investigated for magnetic monitoring in soils, stream and river sediments, revealing a link between magnetic and chemical variables. Recently, Chaparro et al. (2011a) studied river sediments from India; they used successfully principal coordinate analysis (PCoordA) and fuzzy c-means clustering analysis (FCM) to make a classification and to perform a magnetic-chemical characterization of data into four groups (from less to most impacted samples).

Mathematical models are not simple descriptive statistics for particular datasets, but they allow having a wider and global knowledge of the case study. The building technique for a model is based on quantitative (measurements) and qualitative (gained experience) knowledge; this weighted combination enriches the quality outcome, giving a better fitting between data and modelled results. The qualitative knowledge may be useful, but sometimes it is not easily quantifiably and therefore cannot be available for classical mathematical models. The fuzzy tools may usually be appropriated to model uncertainties that are inherent in colloquial language, as well as to emulate some logic mechanisms and to mimic how the human brain tends to classify imprecise information or data.

In this contribution, continuing and improving previous studies (Chaparro et al. 2011b), a methodology is reported in order to build a mathematical model for calculating a "confidence interval" of the heavy metals.

Methodology

The model

A fuzzy logic model is also known as a fuzzy inference system or a fuzzy rule based system. The essence of fuzzy logic rests on the truism that all things admit degrees of vagueness. Black and white cases are the exception in a world of grey (Mackinson et al. 1999). In very formal terms a fuzzy set A defined in a discursive universe X is a set of pairs $(x, \mu_A(x))$ where x belongs to X and $\mu_A(x)$ is a number in the interval $[0, 1]$ representing the degree of membership of x in A. Expert knowledge is represented by a set of fuzzy rules, they are of the form "*IF this THEN that*". Rules made associations between input and output fuzzy sets.

Basically, any fuzzy logical model is formed by four parts: (a) the *input processor*; (b) the *fuzzy rule base*; (c) the *fuzzy inference engine*; (d) the *defuzzifier*. This last process is called *defuzzification*. For detailed information, the reader is referred to Chaparro et al. (2011b).

In this work, the fuzzy logic models are of the Takagi-Sugeno-Kant type (Nguyen and Walker, 1997) (FTKS). This approach is essentially based on the possibility of describing the local dynamics of a problem in approximate terms. This is the case, for example, when for each member of a fuzzy partition of the input space of X, the difference equation of the problem is linear to some degree. This suggests forming rules as follows,

$$R_j : \text{"}IF\ \ x_1\ is\ A_j^1\ \ and\ \ x_2\ is\ A_j^2\ ...and\ \ x_N\ is\ A_j^N\ \ THEN\ \ y{=}f_j(x_1{,}x_2{,}...{,}x_N)\text{"},\quad j = 1, 2, ..., r$$

where: x_i are the actual observed values of input variables, and $f_j\ (.)$ is some specific linear function, such as

$$f_j(x_1, x_2, ..., x_N) = \sum_{i=1}^{N} \alpha_{i,j} \cdot x_i$$

For FTSK systems, the consequent in each R_j is expressed by a constant value. The rule R_j will produce a *crisp output* given by:

$$y_j = \tau_j\ f_j(x_1, x_2, ..., x_N)$$

where τ_j is the degree of applicability or weight of the rule R_j. Then the overall output value is taken to be a weighted average:

$$y(x_1, x_2, ..., x_N) = \frac{\sum_{j=1}^{r} \tau_j \; f_j(x_1, x_2, ..., x_N)}{\sum_{j=1}^{r} \tau_j}$$

The membership functions are defined using all data (n) and fuzzy c-means clustering analysis (FCM). The FCM is an unsupervised clustering algorithm that allows finding several partitions ambiguously from 2 to $n-1$ for all variables. The number of cluster (c) was fixed by the expert. Once the number c is defined, the FCM is applied to each variable. From this fuzzy partition for each variable, a parametrization is carried out to define type and parameters for each membership function.

The inference rules are built using piece of information from the fuzzy partition. From the samples, the maximum membership degree (if above 0.60) of each input and output variables is considered to build a rule. If the membership value is below 0.60, this datum is not used for the rules. It is necessary for all the variables of a sample to have membership grades above 0.60; otherwise the sample will not be used in the model. According to Chaparro et al. (2011b), each “useful” sample is identified and labelled using its corresponding fuzzy set with maximum membership. Thus, the rule is established by the label of each set for each variable. In addition, a “grade of confidence” or weight is determined for each rule, which is observed from the rule weight (*CF*). This CF is defined by the following expression,

$$CF\ (x) = \prod_{k=1}^{m} \max\ \left(u_{k,i}(x_i)\right)$$

where $u_{k,i}(x_i)$ is the membership degree of the k-th variable for the i-th datum. Higher values imply more confident rules.

In order to validate the model, two indicators were used:

i) Mean estimation error (MEE) between the observed T_i and the estimated $\hat{T}_i$ times which is defined by $MEE = \frac{1}{n}\sum_{i=1}^{n}\left(T_i - \hat{T}_i\right)$.

ii) Coefficient of determination R^2, which is defined by $R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{T}_i - T_i\right)^2}{\sum_{i=1}^{n}\left(T_i - \overline{T}\right)^2}$,

where $\overline{T}$ is the average observed development time. As the value of R^2 is closer to 1, it indicates a better fit.

Dataset and methods

The studied set is example of pollution by different anthropogenic sources in soil samples collected in Marambio station from Antarctica (64º14’S; 56º37’W, $n= 20$). The data under study were recently published, for detailed information the reader is referred to Chaparro et al. (2007).

The datasets comprised magnetic and chemical variables, i.e. twelve (12) variables; specifically, magnetic variables: χ, ARM, saturation of isothermal remanent magnetisation (SIRM), κ_{ARM}/κ, S-ratio (IRM_{-300mT}/SIRM), Hcr; and chemical variables: contents of Cr, Ni, Cu, Zn, Pb, Fe. However, for this work, the following four input variables: χ, ARM (concentration-dependent variables), κ_{ARM}/κ-ratio and Hcr (magnetic features-dependent variables) were selected. The output variables were Cr, Cu, Zn and Pb ($m= 4$). This selection of variables was carried out according to the empirical knowledge and relevance of parameters in magnetic monitoring.

The model was implemented using the software MATLAB R2009b. The program was written by the authors and runs the simulation with the output option as a confidence interval.

Results

The input membership functions are shown in Fig. 1, and the parameters of output consequent functions are detailed in Table 1.
The outcome of applying the model is shown in Fig. 2. In this figure, the data, modeled values and confidence intervals of Zn, Pb, Cu and Cr are displayed for comparison purpose.

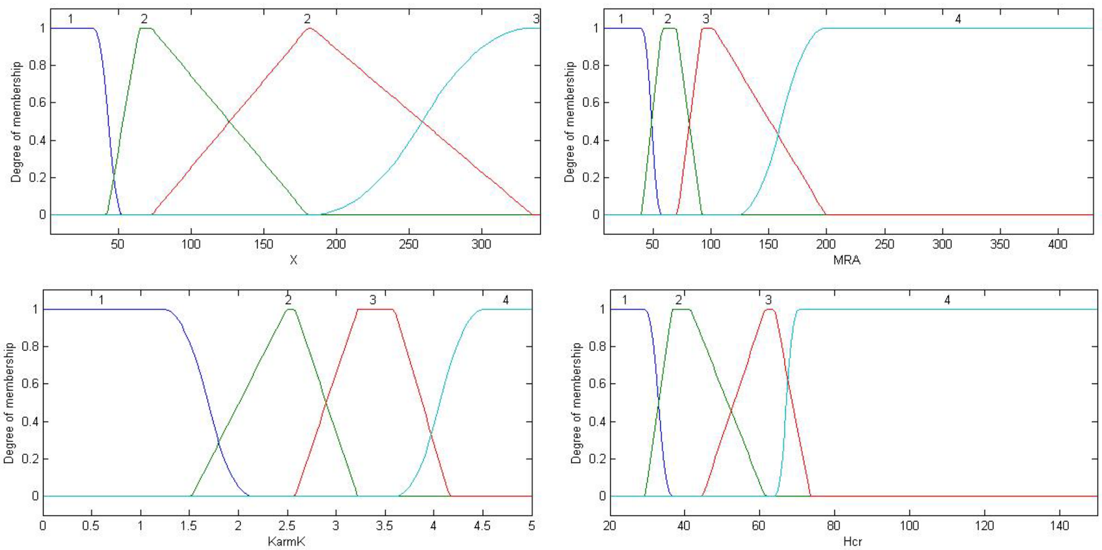

Figure 1. Input memberships functions

Table 1. Parameters of consequent functions (output)
(Metal= $a\,\chi + b\,ARM + c\,\kappa_{ARM}/\kappa + d\,Hcr + e$)

Zn	*a*	*b*	*c*	*d*	*e*
1	-4.023	2.431	-32.032	0.247	108.089
2	-0.287	0.289	-11.115	-0.009	81.084
3	-0.379	0.724	-8.907	-0.150	64.580
4	-0.151	0.331	-3.636	0.266	56.104
Cu	*a*	*b*	*c*	*d*	*e*
1	-0.362	0.360	-8.598	0.216	16.261
2	0.047	-0.006	0.904	0.011	7.384
3	-0.004	0.002	-0.620	0.066	16.628
4	0.214	-0.383	13.627	-0.608	33.559
Pb	*a*	*b*	*c*	*d*	*e*
1	0.050	-0.017	1.535	-0.004	7.125
2	-0.019	0.021	3.247	-0.357	28.496
3	0.791	-0.244	7.578	3.866	-127.409
4	1.952	-8.863	158.698	-10.938	781.139
Cr	*a*	*b*	*c*	*d*	*e*
1	1.896	-0.486	4.128	0.269	-19.297
2	-0.023	0.036	-1.547	0.007	31.802
3	0.059	-0.040	6.418	-0.012	22.015
4	-0.095	0.034	-12.018	-0.761	97.267

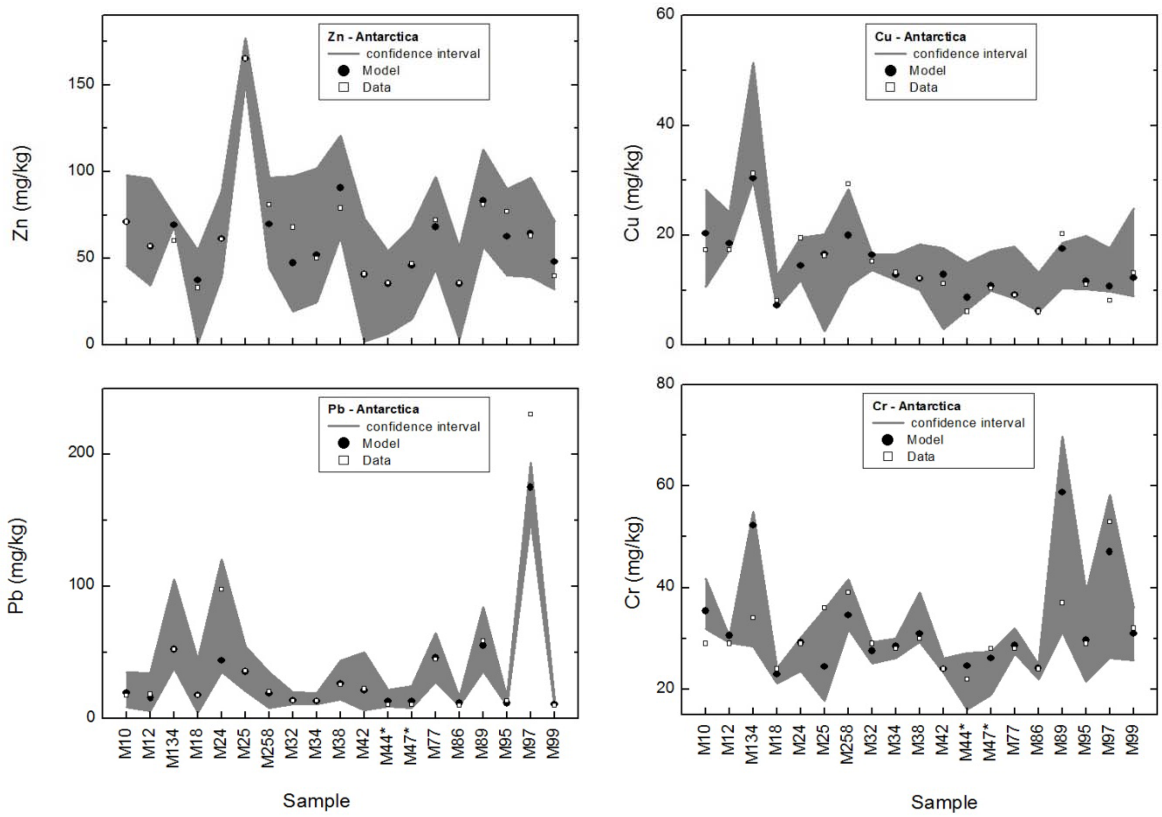

Figure 2. Data and results from the model.

The model yields a satisfactory approximation of (measured) data for each heavy metal. This fact is reflected in results of both indicators used, mean estimation error and the coefficient of determination R^2 (Table 2).

Table 2. Results of the indicators used for the model validation.

	R^2	*MEE*
Zn	0.450	8.762
Cu	0.765	-1.293
Pb	0.478	5.440
Cr	0.662	-0.194

Discussion

A previous fuzzy model of Mandani type for calculating a pollution index (Chaparro et al., 2011b) was improved. The methodology of this previous study was used in order to determine parameters and membership functions for this work.

The new results using a model of FTSK seem to be more precise than the model based on Mandami type. This is a consequence of the problem description, which is almost local in FTSK model and the Mandami model gives a fuzzy set as outcome. The problem or disadvantage of FTSK model may come from the input values, especially if they differ substantially from the data used to build the model. This disadvantage lies in the model architecture that is based on regression models and the high dependence of data used for the construction. Such a fact should be taking into account when a new model is developed.

Although the approximation of models are satisfactory and useful results to explain the problem in multivariate terms, the information obtained from the rules are the most important result from the methodology. The latter are important because they determine the distinctive characteristics to focus attention on the samples under study.
The consequent functions vary and depend on the dataset used to build the model, however; if the dataset change, the interesting fact is the repetition or similarities of rules.
The proposed models, besides of calculating heavy metals values, they are useful tools to explain the relationships between magnetic and chemical variables. Additional information helps to know better the knowledge about the problem, which allows determining the rules to understand the relationships.

Conclusions

This methodology is easy to implement and provides to the user with a simple and effective modelling tool. A large amount of data is not necessary; however, the increase of data allows improving the model.
The model yields a satisfactory approximation for each heavy metal. This fact is reflected in results of the mean estimation error and the coefficient of determination R^2.
The rules constitute a useful tool to analyse the problem, allowing obtain information from the area. Similarities between equivalent rules validate the use of multivariate techniques to study the association between heavy metals and magnetic variables.

References

Chaparro MAE, Gogorza CSG, Chaparro MAE, Irurzun MA, Sinito AM, 2006. Review of magnetism and heavy metal pollution studies of various environments in Argentina. Earth Planets Space 58(10):1411–1422.

Chaparro M.A.E., Nuñez H., Lirio J.M., Gogorza C.S.G. and Sinito A.M., 2007. Magnetic screening and heavy metal pollution studies in soils from Marambio station, Antarctica. Antarct. Sci., 19, 379−393.

Chaparro Marcos A.E., Chaparro M.A.E., Marinelli C. and Sinito A.M., 2008. Multivariate techniques as alternative statistical tools applied to magnetic proxies for pollution: cases of study from Argentina and Antarctica. Environ. Geol., 54, 365–371.

Chaparro Marcos A.E., Chaparro Mauro A.E., Rajkumar P., Ramasamy V., Sinito Ana M. 2011a. Magnetic parameters, trace elements and multivariate statistical studies of river sediments from the south eastern India: a case study from Vellar River. Environ. Earth Scie., 63(2):297-310.

Chaparro Mauro A.E., Chaparro Marcos A.E., Sinito Ana M., 2011b. An interval fuzzy model for magnetic monitoring: estimation of a pollution index. Environ. Earth Scie. (in reviewing process)

Knab M, Appel E, Hoffmann V (2001) Separation of the anthropogenic portion of heavy metal contents along a highway by means of magnetic susceptibility and fuzzy c-means cluster analysis. Eur J Environ Eng Geophys, 6:125–140

Mackinson, S., Vasconcellos, M., Newlands, N., 1999. A new approach to the analysis of stock-recruitment relationships: ‘‘Model-free estimation’’ using fuzzy logic. Canadian Journal of Fisheries and Aquatic Sciences, 56: 686-699.

Nguyen T. H., Walker E. A., 1997. *A First Course in Fuzzy Logic*. New York: CRC Press

Petrovský E., Elwood B., 1999. Magnetic monitoring of air, land and water pollution. In: Maher B.A. and Thompson R. (Eds.), *Quaternary Climates, Environment and Magnetism*. Cambridge University Press, 279−322.

Petrovský E, Kapicka A, Jordanova N, Boruka L, 2001. Magnetic properties of alluvial soils contaminated with lead, zinc and cadmium. J App Geophys, 48:127–136.