Environ Earth Sci
DOI 10.1007/s12665-011-1387-z

ORIGINAL ARTICLE

An interval fuzzy model for magnetic monitoring: estimation of a pollution index

Mauro A. E. Chaparro · Marcos A. E. Chaparro · Ana M. Sinito

Received: 15 December 2010 / Accepted: 30 September 2011

Abstract In this contribution, a methodology is reported in order to build an interval fuzzy model for the pollution index PLI (a composite index using relevant heavy metal concentration) with magnetic parameters as input variables. In general, modelling based on fuzzy set theory is designed to mimic how the human brain tends to classify imprecise information or data. The "interval fuzzy model" reported here, based on fuzzy logic and arithmetic of fuzzy numbers, calculates an "estimation interval" and seems to be an adequate mathematical tool for this nonlinear problem. For this model, fuzzy *c*-means clustering is used to partition data, hence the membership functions and rules are built. In addition, interval arithmetic is used to obtain the fuzzy intervals. The studied sets are different examples of pollution by different anthropogenic sources, in two different study areas: (a) soil samples collected in Antarctica and (b) road-deposited sediments collected in Argentina. The datasets comprise magnetic and chemical variables, and for both cases, relevant variables were selected: magnetic concentration-dependent variables, magnetic features-dependent variables and one chemical variable. The model output gives an estimation interval; its width depends on the data density, for the measured values. The results show not only satisfactory agreement between the estimation interval and data, but also provide valued information from the rules analysis that allows understanding the magnetic behaviour of the studied variables under different conditions.

Keywords Fuzzy *c*-means clustering · Interval fuzzy model · Magnetic monitoring · Magnetic parameters · Pollution

M. A. E. Chaparro
Instituto Multidisciplinario de Ecosistemas y Desarrollo Sustentable (UNCPBA)-CONICET, Pinto 399, B7000GHG Tandil, Argentina

M. A. E. Chaparro · A. M. Sinito
Instituto de Física Arroyo Seco (UNCPBA)-CONICET, Pinto 399, B7000GHG Tandil, Argentina

M. A. E. Chaparro (✉)
Instituto de Física Arroyo Seco (UNCPBA), Pinto 399, B7000GHG Tandil, Argentina
e-mail: chapator@exa.unicen.edu.ar

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 analyse the distribution and the reach around the contamination sources. Although the anthropogenic 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 Elwood 1999). In particular, in previous studies (Chaparro et al. 2006, 2007, 2008; Marié et al. 2010), multivariate statistical analyses were investigated for magnetic monitoring in soils, stream sediments and road-deposited sediments, revealing a link between magnetic and chemical variables.

Although a large number of authors found linear relationships between magnetic parameters and chemical variables for particular environments, non-significant correlations were reported as well (Georgeaud et al. 1997; Petrovský et al. 1998; Chaparro et al. 2007; Ng et al. 2003; Desenfant et al. 2004; Spiteri et al. 2005; Lu et al. 2005; Magiera et al. 2006). The latter, as well as non-reported studies, reveals that 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 (Bityukova et al. 1999; Petrovský et al. 2001; Knab et al. 2001; Wang and Qin 2006; Chaparro et al. 2008). Recently, Chaparro et al. (2010) 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.

The defuzzification comprises the transformation from a fuzzy set to a crisp number. Some useful information is lost in classical defuzzification process, for this reason, and to achieve better results, here a fuzzy interval instead of a crisp output is calculated.

In this contribution, a methodology is reported in order to build a mathematical model for calculating an "estimation interval" of the index PLI. The model is based on fuzzy logic and arithmetic of fuzzy numbers, this type of model is called "interval fuzzy model".

Methodology

Model

Basically, a fuzzy model or fuzzy inference system (FIS, Klir and Yuan 1995) is formed by four parts (Fig. 1): (a) the *input processor*, which translates (non-) quantifiable inputs into fuzzy sets of their respective universes; (b) the *fuzzy rule base*, consisting of a collection of fuzzy IF-THEN rules aggregated by the disjunction or the conjunction, which is a key knowledge-encoding component of fuzzy rule-based systems; (c) the *fuzzy inference engine*, performing approximate reasoning by using the compositional rule of inference, hence a fuzzy set answer or global conclusion will be calculated by aggregation of the partial solutions contributed by each rule; (d) the *defuzzifier*, which assigns a real (or crisp) number that is representative of the corresponding fuzzy set answer. The last process is called *defuzzification*.

First, it is necessary to partition the space into possible clusters to build the model. The knowledge and data are used to select variables and to partition the space using fuzzy *c*-means clustering. This fuzzy partition, instead of an exact partition, and the fuzzification comprise the process of transforming the data into degrees of membership for fuzzy sets.

As different numbers of clusters can be assumed, the optimum number is evaluated from the analysis of the modified partition coefficient (V_{MPC}, Dave 1996). These processes allow defining the number of membership functions and therefore the fuzzy rules. The rules comprise different situations and they are defined as conditional statements involving linguistic variables, and values determined by fuzzy sets. Some useful information about the rule inferences is lost in classical defuzzification process, for this reason, fuzzy arithmetic was used to achieve better results.

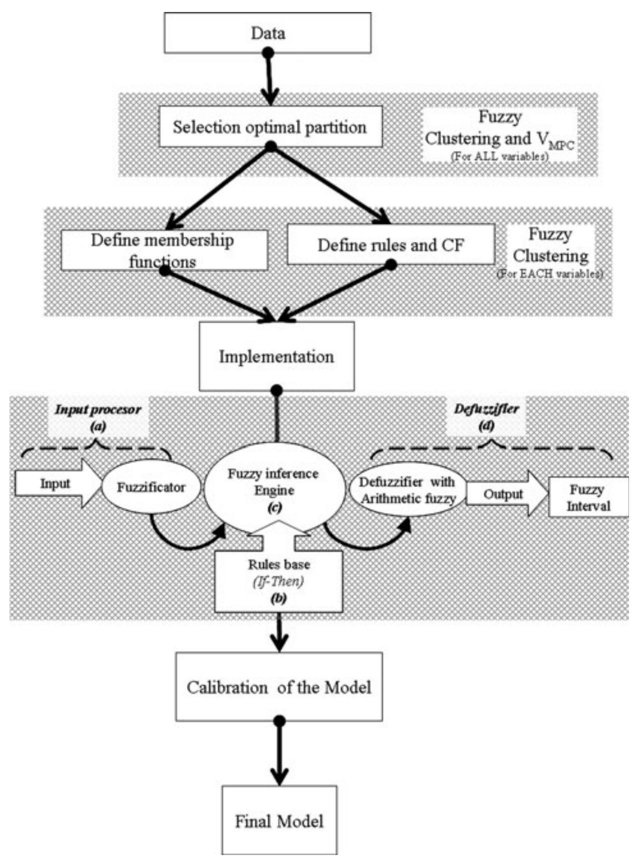

Fig. 1 Schematic diagram of the fuzzy inference system

Membership functions (MF)

The membership functions are defined in two steps using all data (n) and fuzzy c-means clustering analysis. First, the FCM is applied to all (input and output) variables. The FCM is an unsupervised clustering algorithm that allows finding several partitions ambiguously from 2 to $n-1$ for all variables. As it is necessary to preassume the number of clusters (c) for FCM and this number c is unknown, a method have to be performed to find an optimal c and solve the cluster validity (Bezdek 1974).

The optimal number of clusters is usually determined from popular validity indexes, e.g. the partition coefficient (PC, Bezdek 1981) and the entropy index (PE, Bezdek 1974). Hence, for a good partition, minimum partition entropy or maximum partition coefficient is studied (Höppner et al. 1999). Both PC and PE possess monotonic evolution tendency with c. Modification of the V_{MPC} index proposed by Dave (1996) can reduce the monotonic tendency and is defined as:

$$V_{\mathrm{MPC}} = 1 - \frac{c}{c-1}\left(1 - \frac{1}{n}\sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^{2}\right) = 1 - \frac{c}{c-1}(1 - \mathrm{PC}) \quad (1)$$

where u_{ji} is the membership grade of the jth cluster and the ith data. The index values from Eq. 1 range in [0; 1] and the optimal cluster number ($c*$) is found by solving $\max_{2<j<c-1} V_{\mathrm{MPC}}$ to produce the best clustering performance for the studied data set.

Secondly, once the number $c*$ is defined, the FCM is applied again but to each variable. From this fuzzy partition for each variable, a parametrization is carried out to define parameters for each trapezoidal MF. The trapezoidal MF is known in fuzzy arithmetical terms like as the flat fuzzy number. This relationship between fuzzy set and fuzzy number allows applying fuzzy arithmetic based on operations on closed intervals. This selection is appropriated and necessary for the next step in the construction of the model.

Rules

The inference rules are built using piece of information from the fuzzy partition. In this work, the 70% of the data were used, which were selected by a random sampling with replacement.

From the selected 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.

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. For instance, if for a sample, with four input variables (*Inp#*) and one output variable (*Outp*), is observed that:

(a) For *Inp1* the value belongs to fuzzy set $Inp1_1$ (label *1*) with membership value *0.80*.
(b) For *Inp2* the value belongs to fuzzy set $Inp2_5$ (label *5*) with membership value *0.70*.
(c) For *Inp3* the value belongs to fuzzy set $Inp3_2$ (label *2*) with membership value *0.66*.
(d) For *Inp4* the value belongs to fuzzy set $Inp4_3$ (label *3*) with membership value *0.86*.
(e) For *Outp* the value belongs to fuzzy set $Outp_2$ (label *2*) with membership value *0.90*.

Therefore, for this example the IF-THEN rule will be *15232*. In Table 2, another example of IF-THEN rule (rule *13432*) is detailed.

In addition, a "grade of confidence" or weight is determined for each rule, which is observed from the rule weight (CF, Eq. 2). This CF is defined by the following expression,

$$\mathrm{CF}(x) = \prod_{k=1}^{m} \max\left(f_{k,i}(x_i)\right) \quad (2)$$

where $f_{k,i}(x_i)$ is the membership degree of the kth variable for the ith datum. Higher values imply more confident rules. For the previous example, the CF will be *0.29*. Hence, the rule will be defined as *15232* (*0.29*).

Defuzzification—fuzzy arithmetic

The defuzzification comprises the transformation from a fuzzy set to a crisp number. As aforementioned, some useful information is lost in classical defuzzification process, for this reason and to achieve better results, a fuzzy interval was calculated instead of a crisp output. Often, results from each active rule are added using operations (defined by t-norms) on fuzzy sets. Although each inference output can be aggregated as a fuzzy set, these results were constrained to a more specific output, that is, a fuzzy interval. A fuzzy interval was defined by a fuzzy set A satisfying the following: (a) A is normal; (b) The support $\{x: A(x) > 0\}$ of A is bounded; (c) The α-cuts of A are closed intervals.

Fuzzy arithmetic is based on two properties of fuzzy numbers: each fuzzy number can fully and uniquely be represented by its α-cuts and the α-cuts of each fuzzy number are closed intervals of real numbers for all $\alpha \in (0, 1]$ (Klir and Yuan 1995). These properties enable to define arithmetic operations on fuzzy numbers, in terms of operations on closed intervals. The four operations on closed intervals are defined by:

$$[a,b]+[d,e]=[a+d,b+e]$$
$$[a,b]-[d,e]=[a-d,b-e]$$
$$[a,b].[d,e]=[\min(ad,ae,bd,be),\max(ad,ae,bd,be)]$$
$$[a,b]/[d,e]=[a,b].\left[\frac{1}{e},\frac{1}{d}\right] = \left[\min\left(\frac{a}{e},\frac{a}{d},\frac{b}{e},\frac{b}{d}\right),\max\left(\frac{a}{e},\frac{a}{d},\frac{b}{e},\frac{b}{d}\right)\right] \quad (3)$$

Arithmetic (Eq. 3) on closed interval satisfies some useful properties, i.e. associative, commutative and a very important property for this work is:

$$\text{If } b.c \geq 0 \text{ for every } b \in B \text{ and } c \in C, \text{ then } A.(B.C) = A.B + A.C$$
$$\text{Furthermore, if } A=[a,a], \text{ then } a.(B.C) = a.B + a.C \quad (4)$$

According to the defined operation for this model, a new "average" fuzzy set, that is a fuzzy interval, is obtained. This new fuzzy interval is calculated using the Eq. 5,

$$X_{\text{fuzzy}} = \frac{1}{n}\sum_{i=1}^{n} \alpha_i A_i = \left(\bar{x}_i^{l}, \bar{x}_i^{cl}; \bar{x}_i^{cs}, \bar{x}_i^{s}\right) \quad (5)$$

where each ith element, $A_i = (a_i^l, a_i^{cl}, a_i^{cs}, a_i^s)$, is a fuzzy interval that is constituted by four parameters and α_i is the maximum membership degree.

From the Eq. 5, the estimation interval (EI, Eq. 6) is defined from the maximum membership interval, that is for $i = 1, 2, \ldots, n$,

$$\text{EI}_i = \left(\bar{x}_i^{cl}; \bar{x}_i^{cs}\right) \quad (6)$$

where $\bar{x}_i^{cl}$ and $\bar{x}_i^{cs}$ are the central values obtained by Eq. 5. This interval EI constitutes the confidence indicator for this model, which width is related to the precision, narrower EI is indicative of better results. If a datum belongs to the interval EI_i, then it is a satisfactory approximation.

For example, in Fig. 2, three rules were activated, and therefore, three fuzzy intervals (each one with its corresponding α) obtained. These intervals are used to calculate a new fuzzy interval using interval arithmetic. After that, the model output for such a sample is obtained from four crisp numbers: a^l a^{cl} a^{cs} and a^s. The numbers a^{cl} and a^{cs} belong to an interval whose width or EI (Eq. 6) depends on the degree of membership of the X_{fuzzy}.

Data and methods

The studied sets are different examples of pollution by different anthropogenic sources, in two different study areas: (a) soil samples collected in Base Marambio from

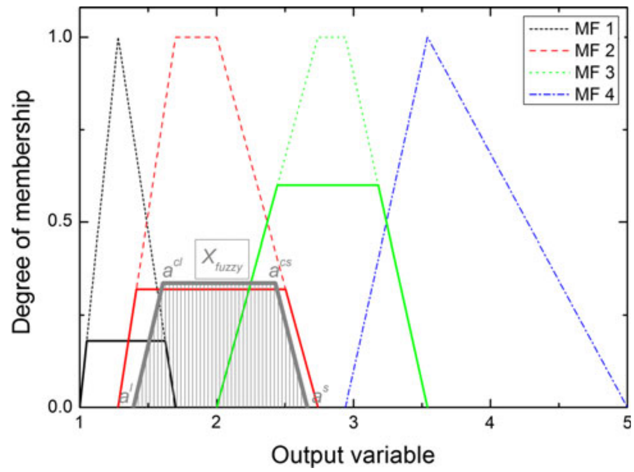

Fig. 2 Example of the defuzzification process. In this case, three rules were activated and a fuzzy number X_{fuzzy} (*shaded area*) is calculated by interval arithmetic. The output is obtained from the crisp numbers: a^l, a^{cl} a^{cs} and a^s, which determine the parameters for the fuzzy interval (a^l,a^{cl},a^{cs},a^s) for this model

Antarctica ($n = 20$); and (b) road-deposited sediments collected in the road Autovia 2 from Buenos Aires Province, Argentina ($n = 31$). The data under study were recently published, for detailed information the reader is referred to Chaparro et al. (2007) and Marié et al. (2010).

The model was implemented using the software SCILAB 5.0.2 (INRIA-ENPC, http://www.scilab.org). The program was written by the authors and runs the simulation with the output option as a fuzzy interval.

The datasets comprised magnetic and chemical variables, i.e. thirteen (13) variables; specifically, magnetic variables: mass-specific magnetic susceptibility (χ), anhysteretic remanent magnetisation (ARM), saturation of isothermal remanent magnetisation (SIRM), anhysteretic susceptibility to volumetric susceptibility-ratio ($\kappa_{\text{ARM}}/\kappa$), S-ratio ($\text{IRM}_{-300\text{mT}}$/SIRM), remanent coercivity (Hcr); and chemical variables: contents of chromium (Cr), nickel (Ni), cooper (Cu), zinc (Zn), lead (Pb), iron (Fe) and Tomlinson pollution load index (PLI) defined by Tomlinson et al. (1980).

$$\text{PLI} = \sqrt[q]{\prod_{p=1}^{q}\left(C_{\text{HM},p}/C_{\text{baseline},p}\right)} \quad (7)$$

where $C_{\text{HM},p}$ is the concentration of each heavy metal and $C_{\text{baseline},p}$ is the baseline value for each heavy metal.

However, for this work, the following four input variables: χ, ARM (concentration-dependent variables), $\kappa_{\text{ARM}}/\kappa$ and Hcr (magnetic features-dependent variables) were selected. The output variable is the PLI, which is a composite index of Cr, Ni, Cu, Zn and Pb ($q = 5$) calculated from the Eq. 7. This selection of variables was carried out according to the empirical knowledge and relevance of parameters in magnetic monitoring.

Results and discussion

Data interpretation

The datasets are different examples of pollution in contrasting areas, their descriptive statistics are summarised in Table 1. The magnetic features-dependent variables are related with the magnetic mineralogy of these soils and sediments. In Antarctica and Argentina, magnetite-like minerals were identified as the main carriers, as well as subordinate hard magnetic carriers (higher values of Hcr), by magnetic and scanning electron microscopy (SEM) studies (Chaparro et al. 2007, 2010, Marié et al. 2010). The presence of these subordinate carriers is evident from the interquartile interval [Q1; Q3], which is wider for Antarctica (from 29.2 to 63.8 mT) than for Argentina (from 33.6 to 37.0 mT). The κ_{ARM}/κ variable is especially sensitive to magnetic grain size; higher values may indicate finer magnetic grain sizes.

The magnetic concentration-dependent parameters show high median values that were interpreted as "magnetic enhancement" by Chaparro et al. (2007) and Marié et al. (2010). In addition, note the different impact between both areas from χ (e.g. 37.6 × 10^{-8} m^3 kg^{-1} for Antarctica and 338.8 × 10^{-8} m^3 kg^{-1} for Argentina) and ARM (e.g. 67.2 × 10^{-6} Am^2 kg^{-1} for Antarctica and 542.8 × 10^{-6} Am^2 kg^{-1} for Argentina).

On the other hand, the PLI variable gives an assessment of the overall toxicity status of a sample, indicating to what extent a sample exceeds the heavy metal content for natural or unpolluted environments (PLI = 1). Such variable has median values up to three times of the values of natural environments.

Membership functions

Although the number of data are different for both cases, i.e. Antarctica ($n = 20$) and Argentina ($n = 31$), the evaluation of Eq. 1 gives a common optimal number $c* = 4$ as can be appreciated in Fig. 3.

It is worth mentioning that the partition of the input and output variable space is entirely determined and constrained to the values of the used dataset. In Fig. 4, it is possible observe the fuzzy partition for each variable, as well as the four membership functions from both cases. The centres of MF are distinctively different between both study areas, which is connected to the chosen dataset used to build the corresponding MF. From this, the expert's empiric knowledge is important to find an adequate choice of samples, as noted in this work; it is possible to obtain different parameters for the MF using another dataset.

A possible way to solve or avoid this problem is adding information to the model from the optimization of parameters. This optimization may be obtained from the expert's knowledge and the increase of dataset with new samples. In this contribution, it is worth of mentioning that the MF's parameters were obtained using FCM.

Rules

In Table 2, the inference rules, CF results and the rule ranking are listed. The number of rules is 17 for Antarctica and 20 for Argentina; some of them are similar showing slight differences between them. This number may be reduced by analyzing such differences and their corresponding CF (Table 2), and therefore fusing repeated rules. Another possibility to optimize the relation data-rules could come from the increase of data; anyway, this topic will be studied in a next stage.

In addition, the rules are not only useful to the defuzzification process and therefore to obtain a representative model, but also they have an added-value because of the parameter interpretation. In general, taking into account the

Table 1 Descriptive statistics of the data: Antarctica and Argentina

	Mean	SD	Min	Q1	Median	Q3	Max
Antarctica ($n = 20$)							
χ [10^{-8} m^3 kg^{-1}]	84.5	99.8	11.9	25.4	37.6	96.7	339.6
ARM [10^{-6} A m^2 kg^{-1}]	97.8	90.6	29.3	53.7	67.2	97.1	425.7
κ_{ARM}/κ [dimensionless]	2.6	1.2	0.9	1.4	2.6	3.5	4.7
Hcr [mT]	49.2	27.9	14.8	29.2	39.8	63.8	142.3
PLI [dimensionless]	2.2	0.8	1.0	1.6	1.9	2.8	4.1
Argentina ($n = 31$)							
χ [10^{-8} m^3 kg^{-1}]	338.5	137.2	33.6	230.1	338.8	422.2	579.4
ARM [10^{-6} A m^2 kg^{-1}]	537.1	136.0	117.4	452.3	542.8	667.6	745.7
κ_{ARM}/κ [dimensionless]	2.5	1.5	1.2	1.6	1.9	3.1	8.3
Hcr [mT]	36.1	6.6	31.7	33.6	34.4	37.0	70.0
PLI [dimensionless]	2.8	0.6	1.4	2.4	2.8	3.1	4.3

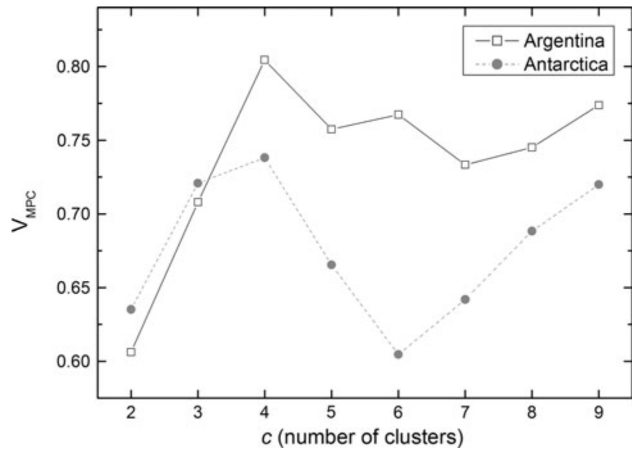

Fig. 3 The validity index V_{MPC} (the modified partition coefficient). The optimal number of clusters is 4 for both cases of study

values of CF (see Table 2), it is concluded from the construction of these rules for both studied examples that,

- In Antarctica, IF low values of χ and ARM, and moderate-high values of κ_{ARM}/κ and low values of Hcr THEN very low values of PLI (CF = 0.903, ranking 4);
- In Argentina, IF very high values of χ and ARM, and low values of κ_{ARM}/κ and Hcr THEN very high values of PLI (CF = 0.902, ranking 2);
- In both sites: (1) IF high-moderate values of χ and ARM, and low values of κ_{ARM}/κ and Hcr THEN moderate values of PLI [CF = 0.980, ranking 1 (Antarctica); CF = 0.903, ranking 1 (Argentina)]; (2) IF high-moderate values of χ and ARM, and moderate values of κ_{ARM}/κ and low values of Hcr THEN high values of PLI [CF = 0.931, ranking 3 (Antarctica); CF = 0.730, ranking 7 (Argentina)].

The rules allow extracting information from both study areas, like as shared characteristics as well as differences between them. For example, (a) in Antarctica, *Rule 11211: IF low values of* χ *and ARM, and moderate-high values of* κ_{ARM}/κ *and low values of Hcr THEN very low values of PLI* (CF = *0.903, ranking 4*), while in Argentina, the equivalent rule according to the classification in Table 2 is, (b) *Rule 22321: IF moderate values of* χ *and ARM, and moderate-high values of* κ_{ARM}/κ *and moderate values of Hcr THEN very low values of PLI* (CF = *0.525, ranking 16*). Such difference between rules from each study area is expected according differences between the environments under study, their pollution sources and the load of pollutants input. As reported by Chaparro et al. (2007) and Marié et al. (2010), human activities and pollution sources are quite different in the Antarctic area and in the Continental (Argentina) area. The pollution influence is local,

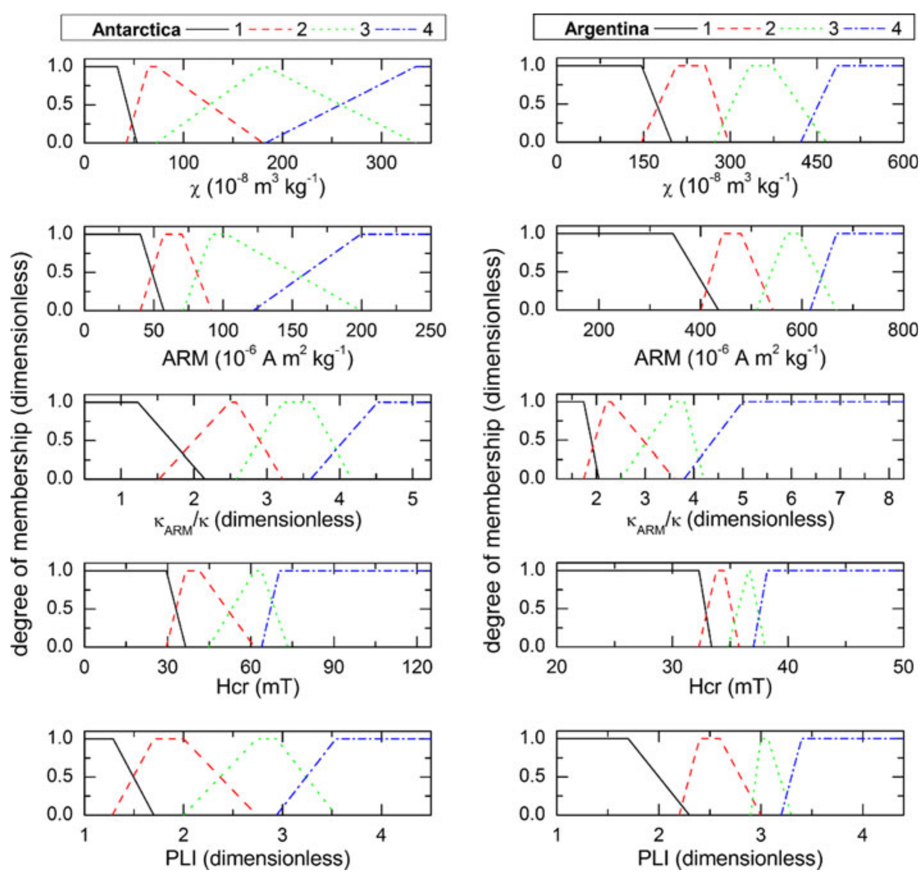

Fig. 4 Trapezoidal membership functions obtained from fuzzy *c*-means clustering analysis

Table 2 Collection of fuzzy IF-THEN rules and their corresponding confidence grade CF

Rule*	CF	Ranking	Classification (from output var.)
Antarctica			
11211	0.903	4	The lowest values
11331	0.894	6	
11341	0.764	10	
44112	0.980	1	Moderate values
11342	0.941	2	
12232	0.858	7	
23212	0.828	8	
12322	0.803	9	
13422	0.733	12	
13432	0.694	13	
12442	0.649	14	
34313	0.931	3	High values
44113	0.902	5	
23223	0.737	11	
12333	0.447	17	
22124	0.571	15	The highest values
33114	0.542	16	
Argentina			
22321	0.525	16	The lowest values
11441	0.325	20	
32112	0.903	1	Moderate values
23342	0.894	3	
43122	0.847	4	
11342	0.759	5	
24412	0.735	6	
33142	0.726	8	
22222	0.710	9	
12432	0.645	11	
31112	0.574	14	
44142	0.496	17	
33242	0.450	18	
22332	0.412	19	
33223	0.730	7	High values
44133	0.698	10	
34213	0.631	12	
12443	0.582	13	
44124	0.902	2	The highest values
34224	0.536	15	

The rule ranking is calculated from the CF values (the highest CF value corresponds to ranking 1)

* IF-THEN rule; e.g. rule 13432 means IF x1 is MF1 and x2 is MF 3 and x3 is MF4 and x4 is MF3 THEN y1 is MF2. Where x1, x2, x3, x4 are the input variables and y1 is the output variable

on nearby soils of Base Marambio, and the sources involve a small power plant, reduced vehicle traffic on summer period and different residues. This human settlement was established about 40 years ago and receives up to 100 researchers in summer. Such activities and sources are really contrasting with the ones from Argentina case. The latter only involves traffic-derived pollutants of a large area (tollbooth areas and roadside soils along a length of 120 km) from a national road (Autovia 2). As mentioned in Marié et al. (2010), this area has a considerable traffic density of about 5,500 vehicles per day, reaching about 8,000 vehicles per day on the weekends.

On the other hand, from equivalent rules of both study cases (e.g. *Rule 44112* (CF = *0.980, ranking 1, Antarctica*) and *Rule 32112* (CF = *0.903, ranking 1, Argentina*)) showing coincidences, it is possible note that there is a relationship between variables beyond the aforementioned differences. Moreover, this fact validates the use of multivariate techniques to study the association between heavy metals and magnetic variables.

Defuzzification—fuzzy arithmetic

Results showed in Fig. 5 belong to the Antarctic dataset. As can be appreciated in this figure, the model calculates the values of the variable PLI successfully. Most of data (90%) are modelled by the intervals, in particular the 55% of data belong to the central fuzzy interval EI (the most confident interval); only two data (10%) were not properly predicted by the model. These data correspond to the "Pristine area" or Control site (sample M44) and the Incinerator area from Base Marambio (sample M12). Control samples of the Antarctica's dataset are M44 and M47, and the model is able to classify them as unpolluted samples; this can be observed from the activated rules for such samples. In spite of the fact that the PLI value of M44 (PLI = 1.01) does not fit to the fuzzy interval (EI) calculated by the model, this sample is adequately classified as unpolluted one. It is necessary to point out that the corresponding fuzzy number is $X_{Fuzzy,M44} = (1.09, 1.35; 1.58, 1.77)$ (see Fig. 5).

On the other hand, the sample M12 shares (from the point of view of fuzzy partitions) similar magnetic properties with samples M89, M97 and M259 (classified as high values of contamination) and with samples M32, M34, M99 and M95 (classified as moderate values of contamination). For both mentioned subset of samples, the χ values of M12 is twice and three times higher. The κ_{ARM}/κ value is up to four/three times lower than subsets of high/moderate values of contamination. Although the sample M10 belongs to same area (Incinerator area) and has similar PLI values, it shows higher magnetic values (except the variable Hcr) than M12. It is concluded from this analysis that sample M12 evidences an anomalous behaviour regarding the Antarctic dataset. It may be a consequence of an erroneous classification or particular characteristics of such

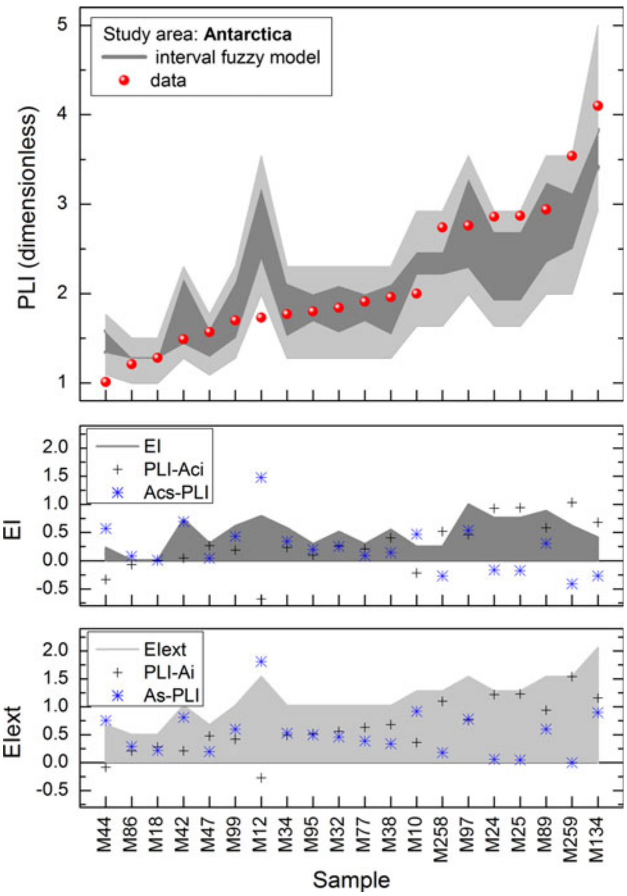

Fig. 5 The interval model and data from Antarctica. The estimation intervals EI (the most confident interval, *dark grey zone*) and *EIext* (*light grey zone*) are shown as well as the differences between the data and the upper/lower (*/+) interval limits

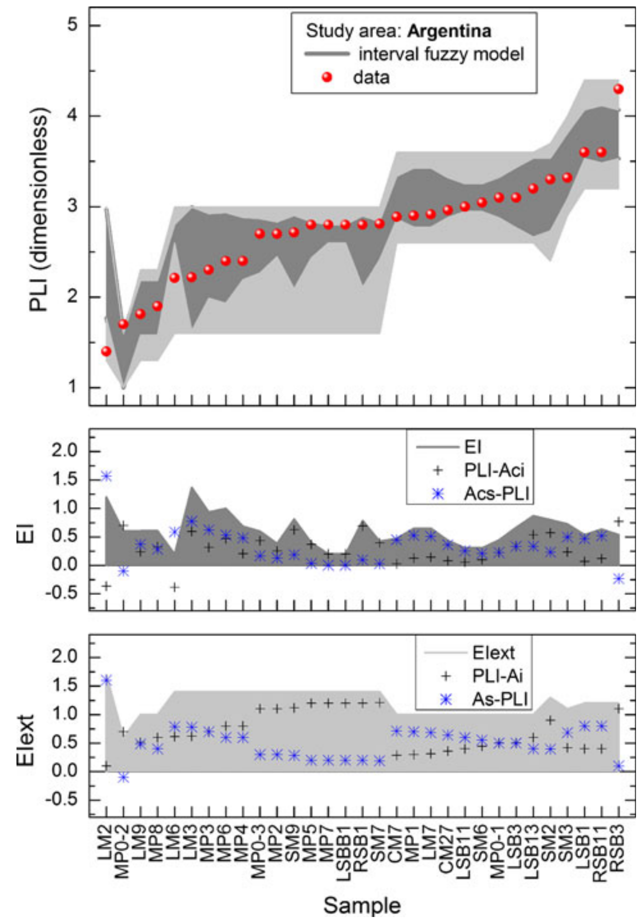

Fig. 6 The interval model and data from Argentina. The estimation intervals EI (*dark grey zone*) and *EIext* (*light grey zone*) are shown as well as the differences between the data and the upper/lower (*/+) interval limits

sampling site that deserves more detailed studies in the future.

On the other hand, results displayed in Fig. 6 belong to the Argentinean dataset. Although these data belong to a quite different environmental problem, the model also calculates the values of the variable PLI successfully. In this case of study, all data are fitted to the model and predicted by the non-central interval *EIext* (Fig. 6). In particular, the 87% of data (27 out of 31) are calculated by the most confident interval EI (central fuzzy interval).

The model shows satisfactory results for these cases, but it is noted that the calculated output for the second study case (Argentina) is better than the first one (Antarctica). As first remark, one explanation is related to difference in the number of data between datasets (i.e. $n = 31$ for Argentina and $n = 20$ for Antarctica), which could be tested in future studies. Other reasons, as aforementioned, differences between cases can come from the environmental characteristics of sites, their pollution sources and the load of pollutants input. This fact may have a direct influence on the ranges and distribution of magnetic and chemical variables, which can be appreciated in Table 1. Note from descriptive statistics wider interquartile intervals and higher SD values for Antarctica than for Argentina, especially for magnetic variables: χ, ARM and Hcr. For example, the coefficient of variation (CV) ranges from 57 to 118% for Antarctica and from 18 to 40% for Argentina.

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 latter is observed from a better prediction for the larger dataset (Argentina, $n = 31$).

The model output shows satisfactory agreement with two quite different datasets (Antarctica and Argentina). However, it is noted better results for Argentina than for Antarctica; this difference may be related to the environmental characteristics and pollution sources, being the worst case (Antarctica) influenced by multiple pollution sources. This fact may have a direct influence on the ranges and distribution of magnetic and chemical variables, and therefore in the construction and prediction of the model.

The rules constitute a useful tool to analyse the problem, allowing obtain information from both study areas. Differences between rules from each area are expected according differences between study cases. Furthermore, similarities between equivalent rules validate the use of multivariate techniques to study the association between heavy metals and magnetic variables. The model may be used in large magnetic dataset to model chemical variables, such as PLI, Cr, Pb and Zn contents among them.

Acknowledgments The authors thank the UNCPBA, CONICET, Agencia Nacional de Promoción Científica y Tecnológica (ANCyT) for their financial support, and the Dirección Nacional del Antártico (DNA, Argentina) for their logistic support in Antarctica. The authors thank both reviewers for their constructive comments and suggestions.

References

Bezdek JC (1974) Cluster validity with fuzzy sets. J Cybernet 3:58–73

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York

Bityukova L, Scholger R, Birke M (1999) Magnetic susceptibility as indicator of environmental pollution of soils in Tallin. Phys Chem Earth (A) 24(9):829–835

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 MAE, Nuñez H, Lirio JM, Gogorza CGS, Sinito AM (2007) Magnetic screening and heavy metal pollution studies in soils from Marambio station, Antarctica. Antarct Sci 19(3):379–393

Chaparro MAE, Chaparro MAE, Marinelli C, Sinito AM (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 MAE, Marié DC, Gogorza CSG, Navas A, Sinito AM (2010) Magnetic studies and scanning electron microscopy—X-ray energy dispersive spectroscopy analyses of road sediments, soils, and vehicle-derived emissions. Stud Geophys Geodaet 54(4):633–650

Chaparro MAE, Chaparro MAE, Rajkumar P, Ramasamy V, Sinito AM (2011) Magnetic parameters, trace elements and multivariate statistical studies of river sediments from south eastern India: a case study from Vellar River. Environ Earth Sci 63(2):297–310

Dave RN (1996) Validating fuzzy partitions obtained through *c*-shells clustering. Pattern Recognit Lett 17:613–623

Desenfant F, Petrovský E, Rochette P (2004) Magnetic signature of industrial pollution of stream sediments and correlation with heavy metals: case study from South France. Water Air Soil Pollut 152(1):297–312

Georgeaud VM, Rochette P, Ambrosi JP, Vandamme D, Williamson D (1997) Relationship between heavy metals and magnetic properties in a large polluted catchment: the Etang de Berre (South of France). Phys Chem Earth 22(1–2):211–214

Höppner F, Klawonn F, Kruse R, Runkler T (1999) Fuzzy cluster analysis: methods for classification data analysis and image recognition. Wiley, New York

Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall, Upper Saddle River

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

Lu S-G, Bai S-Q, Cai J-B, Xu C (2005) Magnetic properties and heavy metal contents of automobile emission particulates. J Zhejiang Univ Sci 6B(8):731–735

Magiera T, Strzyszcz Z, Kapicka A, Petrovský E, MAGPROX Team (2006) Discrimination of lithogenic and anthropogenic influences on topsoil magnetic susceptibility in Central Europe. Geoderma 130:299–311

Marié DC, Chaparro MAE, Gogorza CSG, Navas A, Sinito AM (2010) Vehicle-derived emissions and pollution on the road Autovia 2 investigated by rock-magnetic parameters: a case of study from Argentina. Stud Geophys Geodaet 54:135–152

Ng SL, Chan LS, Lam KC, Chan WK (2003) Heavy metal contents and magnetic properties of playground dust in Hong Kong. Environ Monit Assess 89:221–232

Petrovský E, Kapicka A, Zapletal K, Sebestová E, Spanilá T, Dekkers MJ, Rochette P (1998) Correlation between magnetic properties and chemical composition of lake sediments from northern Bohemia. Preliminary study. Phys Chem Earth (A) 23(9–10):1123–1126

Petrovský E, Elwood B (1999) Magnetic monitoring of air, land and water pollution. In: Maher BA, Thompson R (eds) Quaternary climates, environment and magnetism. Cambridge University Press, Cambridge, pp 279–322

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

Spiteri C, Kalinski V, Rösler W, Hoffmann V, Appel E, MAGPROX Team (2005) Magnetic screening of a pollution hotspot in the Lausitz area, Eastern Germany: correlation analysis between magnetic proxies and heavy metal contamination in soils. Environ Geol 49:1–9

Tomlinson DL, Wilson JG, Harris CR, Jeffrey DW (1980) Problems in the assessment of heavy metals levels in estuaries and the formation of a pollution index. Helgol Meeresunters 33:566–575

Wang XS, Qin Y (2006) Use of multivariate statistical analysis to determine the relationship between the magnetic properties of urban topsoil and its metal, S, and Br content. Environ Geol 51:509–516