Environ Geol (2006) 49: 458–466
DOI 10.1007/s00254-005-0103-2

ORIGINAL ARTICLE

J.M. Andujar
J. Aroba
M.L. la deTorre
J.A. Grande

Contrast of evolution models for agricultural contaminants in ground waters by means of fuzzy logic and data mining

Received: 20 May 2005
Accepted: 13 September 2005
Published online: 20 October 2005

M.L. de la Torre and J.A. Grande belongs to Water Resources and Quality Research Group. J.M. Andujar and J. Aroba belongs to Control and Robotics Research Group

J.M. Andujar · J. Aroba
M.L. la deTorre (✉) · J.A. Grande
Escuela Politécnica Superior, Universidad de Huelva, Ctra Palos de la Frontera, s/n., 21819 Palos de la Frontera, Huelva, Spain
E-mail: mltorre@uhu.es
Tel.: +34-959-217345
Fax: +34-959-217304

Abstract This work aims at contrasting, by means of a set of fuzzy logic- and data mining-based algorithms, the functioning model of a detritic aquifer undergoing overexploitation and nitrate excess input coming from strawberry and citrus intensive crops in its recharge zone. To provide researchers unskilled in data mining techniques with an easy and intuitive interpretation, the authors have developed a computer tool based on fuzzy logic that allows immediate qualitative analysis of the data contained in a data mass from the water chemical analyses, and serves as a contrast to functioning models previously proposed with classical statistics.

Keywords Detritic aquifer system · Pollution · Nitrate · Aquifer · Fuzzy logic · Data mining

Introduction

In this article a new computer tool is applied: Predictive Fuzzy Rules Generator (PreFuRGe) (Aroba 2003) that allows qualitative interpretation of data recorded in a database relative to the chemistry of water.

Specifically, the authors intend to contrast the model proposed by Grande et al. (1996), which applies classical statistic techniques, such as factorial analysis, to a data mass resulting from the sampling and analysis of a network of 54 wells distributed across the system's recharge zone. By studying this model, the existence of a close dependency relationship between nitrate ion concentrations in the saturated zone of the study area and the presence of strawberry crops in the medium was established.

General setting: location, characteristics and present problems of the aquifer system of Ayamonte-Huelva

The study area is on the southwest border of the depression of the Guadalquivir in the Huelva province (Spain) between the Guadiana and Piedras Rivers (Fig. 1). Aquifer system 25 (Ayamonte-Huelva) provides groundwater to a population of 150,000 in various municipalities, and an irrigated land area of approximately 9,000 ha (ITGE 1989).

The structure and hydrogeological functioning of the system are typical of a multilayer aquifer in a semiconfined complex system (Grande 1993). It is characterized by alternating gravels and sands, semipermeable sandy marls, and with ages between the Andalusian period and Holocene period, limited by an impermeable substratum formed by blue marls (Andalusian) and shales (Grande 1993).

The permeable materials present a monoclinal geometry from north to south (Grande et al. 1991) with transgressive overlap and thickness increasing from 8 m in the northern sector to 40 m in the southern most part. The depth from groundwater surface ranges between 10 and 28 m.

Data (Grande et al. 1996) establish the natural average recharge to the aquifer to be about 100 hm^3/year. However, this estimate is probably too high considering the prolonged drought that has affected the region in the last few years, and the continuous increase in agriculture. These facts have led to an increase in evapotranspiration and consequently a considerable piezometric level drop (4 m) found in the whole system due to pumpage (Grande et al. 1996).

From a global point of view, the area of study shows a distribution of crops grown on irrigated land, of which strawberries, oranges and other fruit-bearing trees have much relevance, occupying more than 50% of the total area.

State of the art

Even in the case of this tool of new application to the research field of water quality, the background review of the described sector leads, first of all, to the work by Grande et al. (2005), where interdependency relationships are established between different contaminants from the mining activity in SW Spain, rainfall and its position in water courses by application of the PREFURGE tool (Aroba 2003).

González et al. (2005a) study the relationships between crops and nitrate concentrations in an aquifer by applying multivariate analysis techniques. Assimakopoulos et al. (2003) use a fuzzy classification methodology for agricultural soils according to the kind and rate of application of N fertilizers used in W Greece. Ferraro et al. (2003) develop fuzzy logic based and field scale indicators to assess the effect of pesticides on cropping systems in the Inland Pampa (Argentina). Mertens and Huwe (2002) present a model for nitrogen balance by means of fuzzy logic to determine percolated nitrate concentrations in agrarian fields. Cornelissen et al. (2001) introduce a fuzzy-set theory and develop fuzzy mathematical models to establish guidelines for sustainable agricultural development based on economic, ecologic and social indicators. Wang and Tang (1999) propose an example of application on agriculture of a fuzzy adoptive control technique for time-varying intelligent systems. Center and Verna (1998) present an overview of fuzzy logic modelling techniques and their application to biological and agricultural systems. Grande et al. (1995a) study the transfer of nitrates across the unsaturated zone in a strawberry experimental plot by using porous ceramic cups. Grande et al. (1995b) propose a model for predicting the spatial–temporal evolution of nitrogen contaminants in a detritic aquifer.

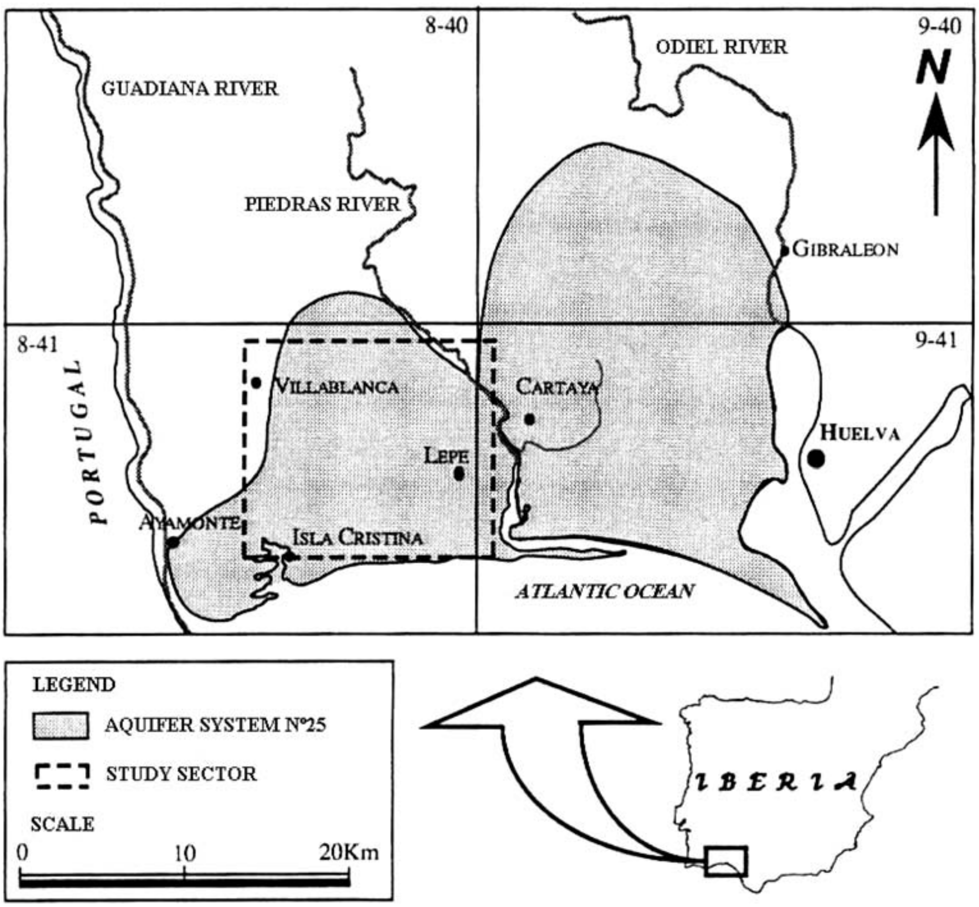

Fig. 1 Location map

Objectives and method

The main objective of the present study is the contrast of the functioning model proposed by Grande et al. (1996) regarding the behaviour of nitrate and other contaminant concentrations in a detritic aquifer undergoing overexploitation and intensive cropping of strawberries and citrus trees. The model to be contrasted was developed by factorial and correlation analysis techniques, while the model proposed here is developed by applying fuzzy logic and data mining techniques to the same data mass already used for the first model.

The samples were taken every 3 months during the period October 1991–September 1992, each sample being analysed by the parameters: sulphates, nitrates, chlorine, conductivity, sodium, potassium, calcium, bicarbonate and magnesium.

First, the existing cartography was updated to a scale of 1:10,000 (COPUT 1991) to determine the values corresponding to the cultivated area. This was necessary because the original map did not coincide with the observable reality at the time the experiments began, because of crop rotation and periods of rest in the agriculture of the sector.

Next, the authors proceeded with the quantification of the above-mentioned areas, grouping types depending on the nature of the crops, and with fertilizer use (Halliday and Wolfe 1991). Thus, strawberry crops and other crops grown on irrigated land have been grouped together. Oranges and other fruit growing trees have been grouped in another column. In both cases, the first component (whether strawberries or oranges) represents 51% of the reflected total.

The quantification of crops has been made using three areas. The first one is a circle with a radius of 400 m with a centre at the point where the sample was taken. The second one is a circle with a radius of 200 m with a centre at the same point. The third one is an ellipse with an area equal to that of the smaller circle and oriented in a way in which its longest axis coincides with the dominant direction of the ground water flow in each sector (Fig. 2). Inside the area of study, the areas occupied by each crop were measured and grouped, according to the above-mentioned criteria, for the statistic study.

Fuzzy logic and data mining

Fuzzy logic (Zadeh 1965) works with reasoning rules very close to the human way of thinking, which is approximate and intuitive. The main characteristic of fuzzy logic is that it allows the definition of values without specifying a precise value, something which is not possible with classical logic, upon which computer development has been based so far. In classical logic, the membership to one class or set is binary, i.e. one is either member or not, so that only two precise values are coordinated (1 and 0, yes or no). Thus, if "very low nitrate concentration" is defined for some samples, it is evident that a sample with $NO_3 = 2$ ppm belongs to the cluster and another one with $NO_3 = 40$ ppm does not,

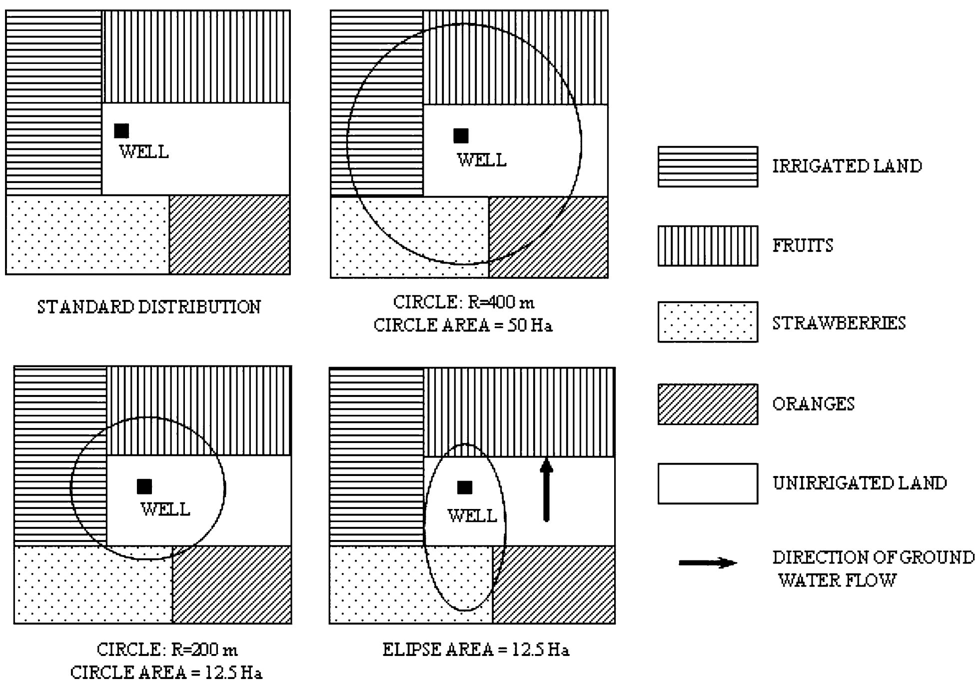

Fig. 2 Quantification of crops

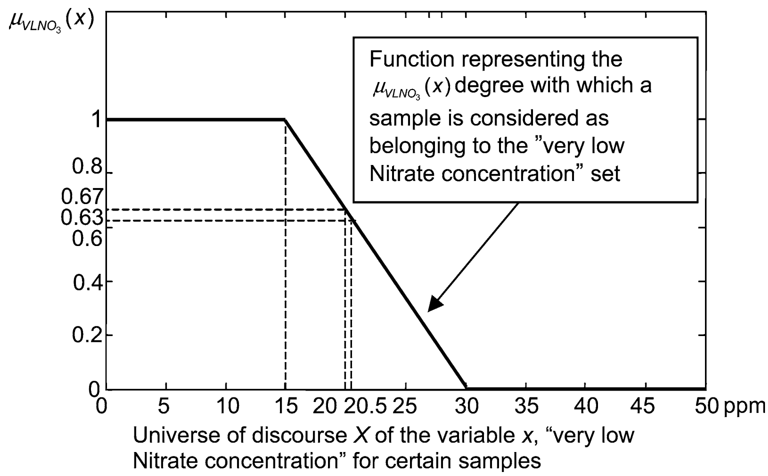

Fig. 3 Example of membership function for the fuzzy set "very low nitrate concentration"

but how is it classified as a sample with $NO_3 = 20$ ppm? It is precisely in the answer to this kind of question where classical logic shows its limitations.

Fuzzy logic allows the association of each sample with a certain degree of fulfilment of the "very low nitrate concentration" prototype. This grade is called "membership grade" $\mu_{VLNO_3}(x)$ of the element $x \in X$ to the set "very low nitrate concentration". The set X is called universe of discourse—range of values—of the variable x. The range of μ_{VLNO_3} ranges from 0 to 1, each value representing the absolute non-membership or membership to the set, respectively.

The membership grade may be represented by a function (von Altrock 1995). Figure 3 shows an example of membership function. Note, for example, that a nitrate concentration of 20 ppm and another one of 20.5 ppm are evaluated differently, but only by means of a slight change, and not by means of a threshold.

Fuzzy sets are a generalization of traditional sets. The $\mu_{VLNO_3}(x) = 0$ and $\mu_{VLNO_3}(x) = 1$ cases, which would correspond to conventional sets, are just special cases of fuzzy sets. The use of fuzzy sets defined by means of membership functions in logic expressions is called *fuzzy logic*. In these expressions, the membership grade of a set is the degree of certainty of the sentence. For example, in Fig. 3, the expression "the nitrate concentration of the sample is very low" would be true in a grade of 0.67 for a sample with $NO_3 = 20$ ppm and 0.63 for a sample with $NO_3 = 20.5$ ppm.

The geometric form of membership functions is totally arbitrary, but in general, simple geometry and known equation functions, such as trapeziums, triangles or sigmoids, are used.

Once all variables involved in the problem are coded to the qualitative domain by means of membership functions, it is possible to write a set of rules representing the relation between input and output variables. These rules present the format *if-then*, and are made up of an antecedent and a consequent; the fulfilment of the antecedent implies the conclusion. From the standpoint of knowledge representation, a fuzzy rule *if-then* is a structure for representing imprecise knowledge. The main characteristic implied by the reasoning based on this type of rule is its ability to represent partial coincidence, which allows a fuzzy rule to provide inference even when the condition is satisfied only partially (Yen and Langari 1999). The following implications allow the brief illustration of these logic inferences:

$$\text{IF } x \text{ is A THEN } y \text{ is C} \quad (1)$$

$$\text{IF } x \text{ is A and } z \text{ is B THEN } y \text{ is C} \quad (2)$$

The first rule has a single antecedent, i.e. of the type "if the variable x is a member of class A". However, the second rule has a compound antecedent (compound antecedents are logical combinations of single antecedents).

The process of extracting knowledge from a database is called knowledge discovery in databases (KDD). This process is made up of several stages ranging from data preparation to achievement of results (Fallad and Uthurusamy 1996; Zaïane 1999). One of these stages is called *data mining* and can be defined as the non-trivial process of extracting implicit, a priori unknown useful information from the stored data (Holsheimer and Siebes 1994).

The computer tool: Predictive Fuzzy Rules Generator (PreFuRGe)

Classical clustering algorithms generate a partition of the population in a way that each case is assigned to a cluster. These algorithms use the so-called "rigid partition" derived from the classical sets theory: the elements of the partition matrix obtained from the data matrix can only contain values 0 or 1; with zero indicating null membership and one indicating whole membership. That is, the elements must fulfill:

$$\begin{aligned}&\text{(a) } \mu_{ik} \in \{0,1\}, \quad 1 \le i \le c, \ 1 \le k \le n\\ &\text{(b) } \sum_{i=1}^{c} \mu_{ik} = 1, \quad 1 \le k \le n\\ &\text{(c) } 0<\sum_{i=1}^{c} \mu_{ik}<n, \quad 1 \le i \le c\end{aligned} \tag{3}$$

Fuzzy partition is a generalization of the previous one, so that it holds the same conditions and restraints for its elements, except that in this case real values between zero and one are allowed (partial membership grade). Therefore, samples may belong to more than one group, so that the selecting and clustering capacity of the samples increases. From this it can be deduced that the elements of a fuzzy partition fulfill the conditions given in Eq. 3, except that now condition (a) will be written as

$$\mu_{ik} \in [0,1], \quad 1 \le i \le c, \quad 1 \le k \le n. \tag{4}$$

The best known general-purpose fuzzy clustering algorithm is the so-called Fuzzy C-Means (FCM) (Bezdek 1981). It is based on the minimization of distances between two points (data) and the prototypes of cluster centres (*c-means*). For this purpose, the following cost function is used:

$$J(X;U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \|x_k - v_i\|_A^2 \tag{5}$$

where $\mathbf{U}$ is a fuzzy partition matrix of $\mathbf{X}$, $V = V = [v_1, v_2, \ldots, v_c]$ is a vector of cluster centre prototypes which must be determined and $m \in [1, \infty]$ is a weighting exponent which determines the degree of fuzziness of the resulting clusters. Finally,

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^{\mathrm{T}} A (x_k - v_i) \tag{6}$$

is the norm used for measuring distances (matrix $\mathbf{A}$ induces the rule to be used—provided that it is the unit matrix, which is very frequent—that is, the Euclidean norm). The described algorithm was used (Sugeno and Yasukawa 1993) to build a fuzzy model based on rules of the form:

$$R^l: \text{IF } x \in A^l \text{ THEN } y \in B^l \tag{7}$$

where $x = (x_1, x_2, \ldots, x_n) \in \Re^n$ are input variables, $A = (A_1, A_2, \ldots, A_n)$ are n fuzzy sets, $y \in \Re$ is the output variable and B is the fuzzy set for this variable.

The developed computer tool, PreFuRGe (Aroba 2003), is based on the previously described methodology (Sugeno and Yasukawa 1993) and represented by Eq. 7. This initial methodology has been adapted and improved in the following aspects:

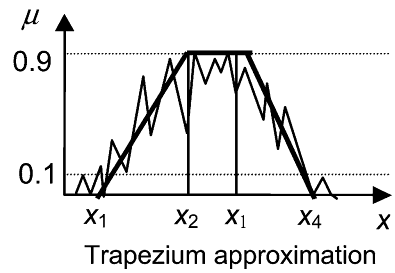

Fig. 4 Trapezium approximation of a fuzzy set

1. It allows working with quantitative databases, with n input and m output parameters.
2. The different variables object of study can be weighted by assigning them weights for the calculation of distances between points of the space being partitioned.
3. The achieved fuzzy clusters are processed by another algorithm to obtain graphic rules trapeziums (Fig. 4).
4. An algorithm processes and solves cases of multiple projections in the input space (mounds).
5. The output provided in the original method has been improved with a graphic interface showing the graphic of the achieved rules.
6. An algorithm provides automatically the interpretation of the fuzzy graphic rules in natural language.Figures 5 and 6 show two examples of rules generated by means of the PreFuRGe tool.

In the rule of Fig. 5, the fuzzy set assigned to each parameter is represented by a polyhedron. The parameter values are represented on the x-axis of each fuzzy set, and the value of membership to a cluster on the y-axis. This fuzzy rule would be interpreted as follows:

IF A1 is small and A2 is bigger or equal
to average THEN S is very small

When applying the fuzzy clustering algorithm (Aroba et al. 2001) to the generated databases, it is possible to obtain multiple projections in the input parameters (mountain). In the fuzzy rule of Fig. 6, a multiple projection (mountain) is represented in the input parameter $A1$. In this case it is observed how the parameter $A1$ can take different types of values for a certain kind of output. This fuzzy rule can be interpreted as follows:

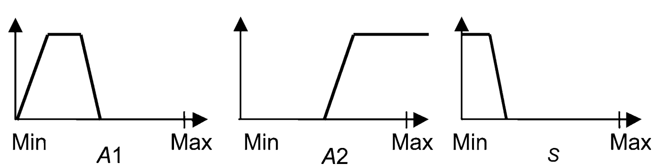

Fig. 5 Example of fuzzy rules generated with PreFuRGe

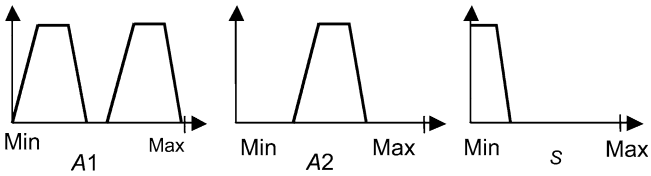

Fig. 6 Example of fuzzy rules generated with PreFuRGe

IF A1 is small or big and A2 is average
THEN S is very small

Recently, one of the authors of this article has investigated the stability of fuzzy logic control systems, as well as their advanced industrials applications (Andujar et al. 2004, 2005a, b).

Results

Figure 7 shows a clear dependency relationship between strawberries and irrigated crops, and nitrate concentration in the saturated zone for each studied case. At the same time, it can be observed that orange and other fruit trees hardly affect nitrate concentration, as suggested by Grande et al. (1996). This phenomenon is corroborated by observing Fig. 7a, e, where it can be noted that a total change in the surface of orange and other fruit trees hardly produce any change in nitrate concentration, as strawberries and irrigated crops are on similar surfaces. However, if Fig. 7d is compared it can observed that for similar orange tree surfaces, a sudden change in the surface of strawberries provokes an immediate response in nitrate concentrations.

Sulphate concentration is not affected by variations of orange tree surfaces. However, something curious occurs with strawberry surfaces: variations in the strawberry surface from low to maximum values stabilize sulphate concentrations at low values. Nonetheless, for very low strawberry surface (Fig. 7a), and sulphate concentration as well.

Chloride, sodium and conductivity values move following similar standards much influenced by both types of crop. For low cropping surfaces, both of strawberries and citrus trees, chloride values are not influenced. However, for very low strawberry surfaces, even with very high orange tree surfaces, chloride, conductivity and sodium values are the lowest ones. It cannot

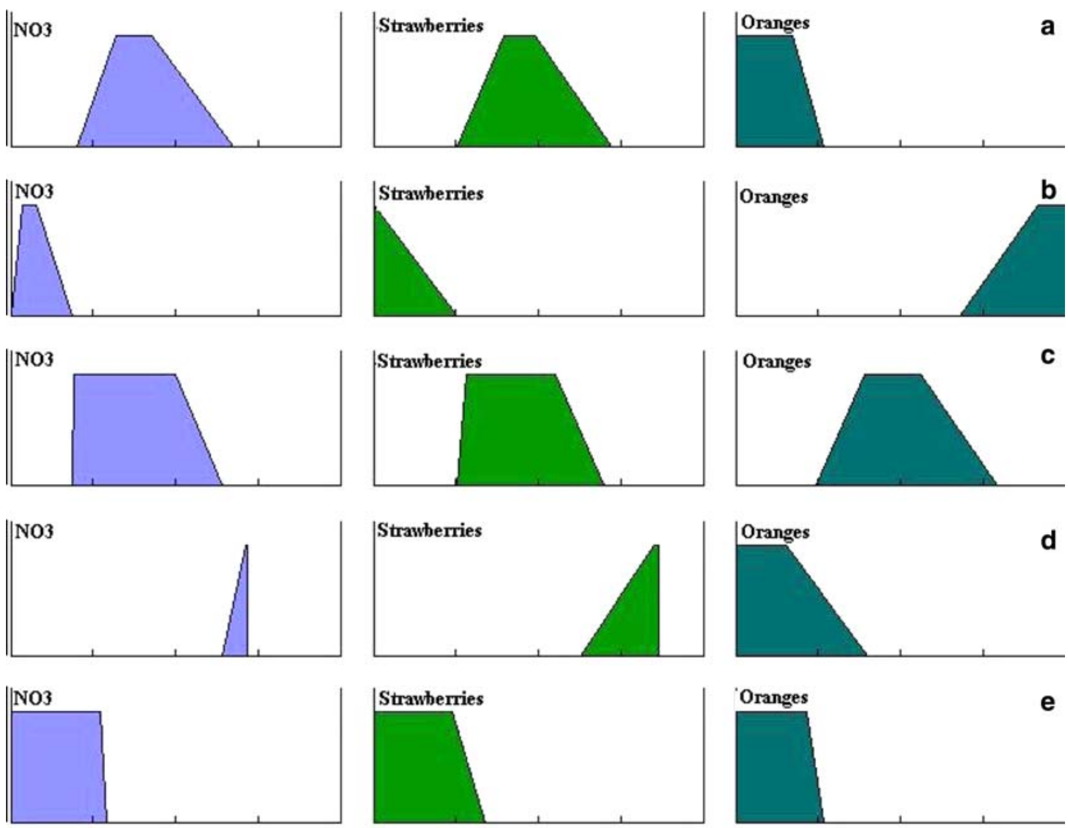

Fig. 7 Graphic fuzzy rules for response of nitrate concentration with respect to strawberries and oranges

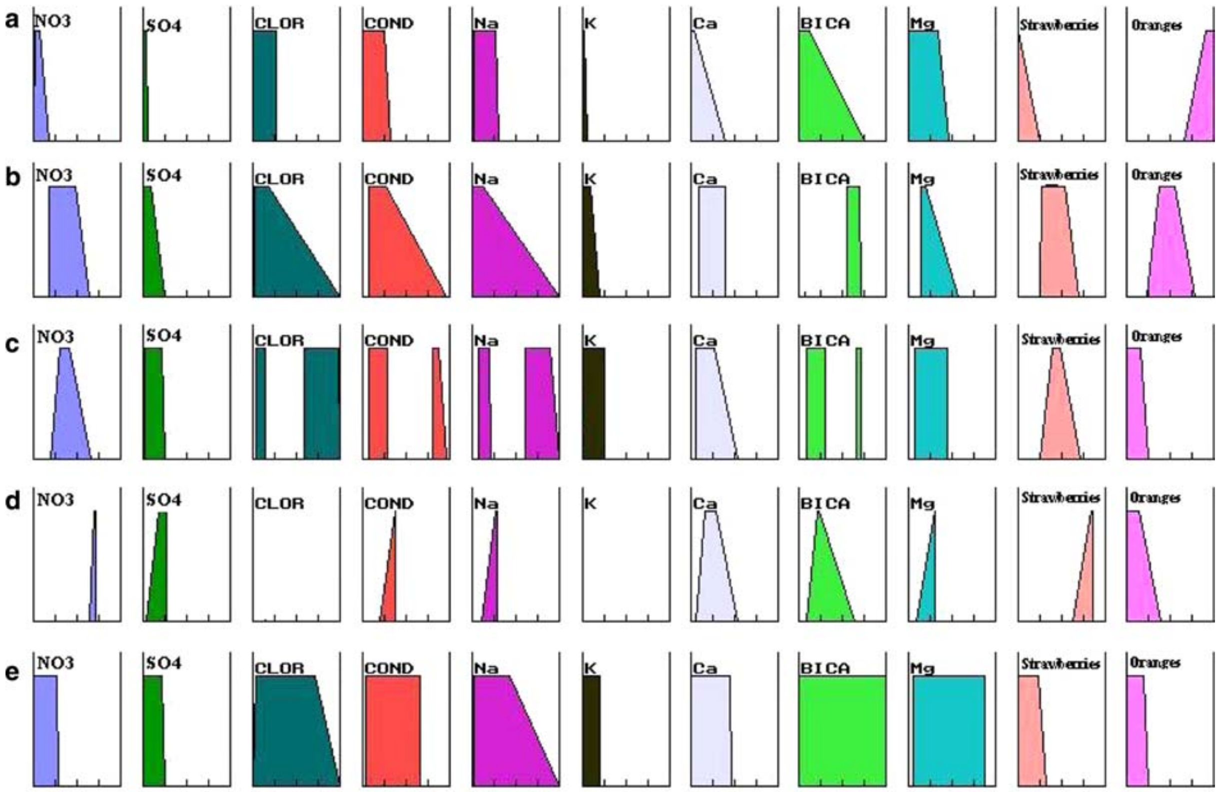

Fig. 8 Graphic fuzzy rules for response of all sampled variables with respect to strawberries and oranges

be actually inferred that the strawberry crop surface gradient moves in the same direction as chloride, sodium and conductivity values, as the increase in cultivated surface does not imply the linear increase of these variables (Fig. 7d). Nonetheless, the influence of the strawberry surface is decisive: if Fig. 7d, e is observed, it can be noted that with a similar orange tree crop surface, the strawberry crop surface moves chloride, sodium and conductivity contents of any possible value (Fig. 7e) to very concentrated low values (Fig. 7d).

As usual in this kind of studies, potassium presents a curious behaviour: when the strawberry crop surface is clearly larger than the other one (Fig. 7a, d), there is practically no potassium. However, when both crops coexist in similar values, then the values for potassium remain practically constant in low to very low values (Fig. 7b, c, e).

Bicarbonates and magnesium do not show any influence when strawberry and orange tree crop surfaces are low (Fig. 7e). For the remainder of the combinations, no cause–effect relationships can be established. Regarding calcium, it shows no affection for any cultivated surface.

In Fig. 8, it can be observed that strawberries and orange tree behaviour, both acting as consequents as opposed to nitrate concentration as the only antecedent.

Figure 8a, d, e shows how for practically similar orange tree surfaces in low to very low values, nitrate concentration goes from low to high (transit of Fig. 8e–a–d). It can be observed how the nitrate concentration evolution follows exactly the same model as the strawberry crop surface.

This fact can be corroborated by analysing Fig. 8b, c. In Fig. 8b it can be observed that nitrate concentration is practically governed by variations in strawberry crop surface, as values of scarce strawberry crop (Fig. 8b, e), nitrates remain at low values, even when in Fig. 8b, e it can be seen that orange trees present a sudden change from low to very high. Finally, Fig. 8c shows how the transit from the cultivated surface to medium values also moves nitrate concentrations towards the same section.

Discussion /conclusions

The aim of this work has been to present qualitative models which allow, in an easy, intuitive and at-a-glance way, and without the need of calculations or data processing, a clear idea of the physical processes that generate the data clusters shown by this computer tool.

The developed methodology allows the establishment of cause–effect relationships, as the cause (fuzzy partition of cultivated surface) originates the effect (NO_3 concentration), represented by the fuzzy clusters at the income. Of course, interpretation must be qualitative, i.e. as a human being would reason, so that no numeric values but predicates are used: high, low, medium, very high, very low, etc.

The application of fuzzy logic and data mining for characterizing hydrochemical processes in the same sector and from the same mass of data confirm and enrich operational models previously proposed for this sector by means of multivariate analysis.

The model proposed by Grande et al. (1996) is thus contrasted by using the techniques described in the present work. As a result, it can be concluded that for the studied sector, the process of nitrate contamination is almost a direct consequence of the development of strawberry crops in the medium, while orange trees hardly contribute to the increase in nitrate concentration in the saturated zone, as already proposed in the model to be contrasted.

Justification of this phenomenon must be found in the kind of crop associated to each species: while citrus trees are grown on large surfaces and by large companies well equipped with important human, technical and economic resources, strawberry crops are the way of life to numerous small family companies where, in many cases and until recently, not even fertilizer dispensers were available and fertilizers were directly added to the well water for its subsequent dosage on plants. In sum, citrus trees do not contaminate because they are grown under a severely controlled irrigation and fertilization regime supported by technical and human means, while strawberries are grown under a much more anarchic system which is not subjected to any kind of water or fertilizer consumption control.

In the described context, the PreFuRGe computer tool used in this work gains a dimension of remarkable efficiency for the qualitative overall diagnosis of the situation, and can also be applied for establishing cause–effect relationships that, in contrast with classical statistical treatments, improve work considerably and make the understanding of the involved processes easier.

References

Andújar JM, Bravo JM, Peregrín A (2004). Stability analysis and synthesis of multivariable fuzzy systems using interval arithmetic. Fuzzy Sets Syst 148(3):337–353

Andújar JM, Bravo JM (2005a) Multivariable fuzzy control applied to the physical–chemical treatment facility of a cellulose factory. Fuzzy Sets Syst 150(3):475–492

Andújar JM, Barragán AJ (2005b) A methodology to design nonlinear fuzzy control systems. Fuzzy Sets Syst 154(2):157–181

Aroba J (2003) Avances en la toma de decisiones en proyectos de desarrollo de software. PhD Thesis, University of Sevilla, Spain

Aroba J, Ramos I, Riquelme JC (2001) Application of machine learning techniques to software project management. In: ICEIS 2001 (III international conference on enterprise information systems), Setubal, Portugal, p 433

Assimakopoulos JH, Kalidas DP, Kollias UJ (2003) A GIS-based fuzzy classification for mapping the agricultural soils for N-fertilizers use. Sci Total Environ 3009:19–23

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

Center B, Verna BP (1998) Fuzzy logic for biological and agricultural systems. Artif Intell Rev 12:213–225

Coput (1991) Abastecimiento y saneamiento de la costa de Huelva. Serv Publ Junta Andalucía, Sevilla

Cornelissen AMG, van der Berg J, Koops WJ, Grossman M, Udo HMJ (2001) Assessment of the contribution of sustainability indicators to sustainable development: a novel approach using fuzzy set theory. Agric Ecosyst Environ 86:173–185

Fallad UM, Uthurusamy R (1996) Data mining y KDD. Com. of the ACM, vol 39, no 11

Ferraro DO, Ghersa CM, Sznaider GA (2003) Evaluation of environmental impacts indicators using fuzzy logic to assess the mixed cropping systems of the Inland Pampa, Argentina. Agric Ecosyst Environ 96:1–18

González JC, Grande JA, Barragán FJ, Ocaña JA, de la Torre ML (2005a) Nitrate accumulation and other components of the groundwater in relation to cropping system in an aquifer in Southwestern Spain. Water Resour Manage 19:1–22

Grande JA (1993) Problemática medioambiental de la contaminación por nitratos en las aguas subterráneas del sistema acuífero n° 25 entre los ríos Guadiana y Piedras. Tesis Doctoral, Universidad de Sevilla

Grande JA, Romero MJ, Muñoz F, González A (1991) Caracterización litoestratigráfica de los materiales postmesozoicos comprendidos entre los ríos Guadiana y Piedras. Proc. Cong. Nac. Ing. Tec. Min. León

Grande JA, González A, Orihuela D, Walter FL (1995a) Transfer of nitrogen contaminnts across the unsaturated zone in an experimental site (Western region Almonte-Marshes, Province of Huelva, Spain). In: Krasny J, Mls J (eds), Acta Univ Carol Geol 39(1):49–63

Grande JA, González A, Sánchez-Rodas D, de la Torre ML (1995b) Influence of the type of cultivation in the nitrate contamination in a detritic aquifer. Application of stepwise method to this phenomenon. Solutions ’95, June 4–10, Edmonton, Canada

Grande JA, González A, Beltrán R, Sánchez-Rodas D (1996) Application of factor analysis to the study of contamination in the aquifer system of Ayamonte-Huelva (Spain). Ground Water 34(1):155–161

Grande JA, Andujar JM, Aroba J, de la Torre ML, Beltrán R (2005) Precipitation, pH and metal load in AMD river basins: an application of fuzzy clustering algorithms to the process characterization. J Environ Monit 7:325–334

Halliday SL, Wolfe ML (1991) Assessing groundwater pollution potential from nitrogen fertilizer using a geographic information system. Water Resour Bull 27:2

Holsheimer M, Siebes A (1994) Data mining: the search for knowledge in databases. Report CS-R9406, CWI, Amsterdam

ITGE (1989) Manuales de utilización de acuíferos (Ayamonte-Huelva). Serv Publ Min Ind Energ, Madrid

Mertens M, Huwe B (2002) FuN-Balance: a fuzzy balance approach for the calculation of nitrate leaching with incorporation of data imprecision. Geoderma 109:269–287

Sugeno M, Yasukawa A (1993) A fuzzy-logic based approach to qualitative modelling. IEEE Trans Fuzzy Syst 1:7–31

Von Altrock C (1995) Fuzzy Logic & neurofuzzy applications explained. Prentice-Hall, Englewood Cliffs

Wang WJ, Tang BY (1999) A fuzzy adaptative for intelligent control. Expert Syst Appl 16:43–48

Yen J, Langari R (1999) Fuzzy logic: intelligence, control and information. Prentice-Hall, Englewood Cliffs

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Zaïane OR (1999) Principles of knowledge discovery in databases. CMPUT690. Department of Computing Science, University of Alberta, Canada