TECHNICAL COMMUNICATION

Fuzzy Intelligence Approach for Modeling the Migration of Contaminants in a Reservoir Affected by AMD Pollution

María Santisteban · José Antonio Grande · María Luisa de la Torre ·
Teresa Valente · Eva Perez-Ostalé · Juan Carlos Cerón · Javier Aroba

Abstract The Sancho Reservoir, located in the Huelva province (SW Spain), is supplied by the Meca River, which receives water contaminated by mining activities in Tharsis. This study focused on determining the relationship that temperature, pH, and electrical conductivity (EC) had with rainfall. The temperature, pH, and EC were simultaneously measured every 30 min by two probes suspended in the Sancho Reservoir. It was anticipated that the use of fuzzy logic and data mining would lead to a model that would show how the contaminant load evolved over space and time. Similar results were obtained for the two locations, except that the parameters had more outliers near the dam due to the greater distance from the contamination source. As expected, higher pH corresponded with lower EC, since, in the absence of chloride, sulphate was the principal anion. The dependency relationship of the variables as well as the cause–effect relationship with the rate of rainfall was more evident in the up-gradient sampling location than near the dam due to the different residence time and the transit time between the two points.

M. Santisteban · J. A. Grande (✉) · M. L. de la Torre ·
T. Valente · E. Perez-Ostalé · J. C. Cerón
Centro de Investigación para la Ingeniería en Minería Sostenible,
Escuela Técnica Superior de Ingeniería, Univ de Huelva,
Ctra Palos de la Frontera s/n, 21819 Huelva, Spain
e-mail: grangil@uhu.es

T. Valente
Dept de Ciências da Terra, Centro de Investigação Geológica,
Ordenamento e Valorização de Recursos, Univ do Minho,
Campus de Gualtar, 4710-057 Braga, Portugal

J. Aroba
Dept de Tecnologías de la Información, Escuela Técnica
Superior de Ingeniería, Univ de Huelva, Huelva, Spain

Introduction

Acid mine drainage (AMD) associated with surface and underground mining in the Iberian Pyrite Belt (IPB) has contaminated surface water in southwestern Europe (Valente and Leal Gomes 2007; Younger 2001; Younger et al. 2002). There is a legacy of about 90 abandoned mines (Grande et al. 2013). The wells, open mines, open pits, kilometres of tunnels, and 350 million m^3 of dispersed waste are all potential contamination sources that affect water, sediment, and the riverbanks. The contaminated water, with its low pH, and elevated metals and sulphate, flows downstream and enters water reservoirs (Santisteban et al. 2013), which must meet water quality standards for human consumption and intensive agricultural activity.

In semi-arid climates, dams are one of the most common ways to meet the essential water requirements for socio-economic development. The study area has a semi-arid Mediterranean climate with an annual rainfall of ≈630 mm/year, mild temperatures (averaging 17.1 °C and a range of 50 °C). Rainfall occurs mainly in the autumn and winter, with the summer and part of the spring being dry (Santisteban et al. 2013).
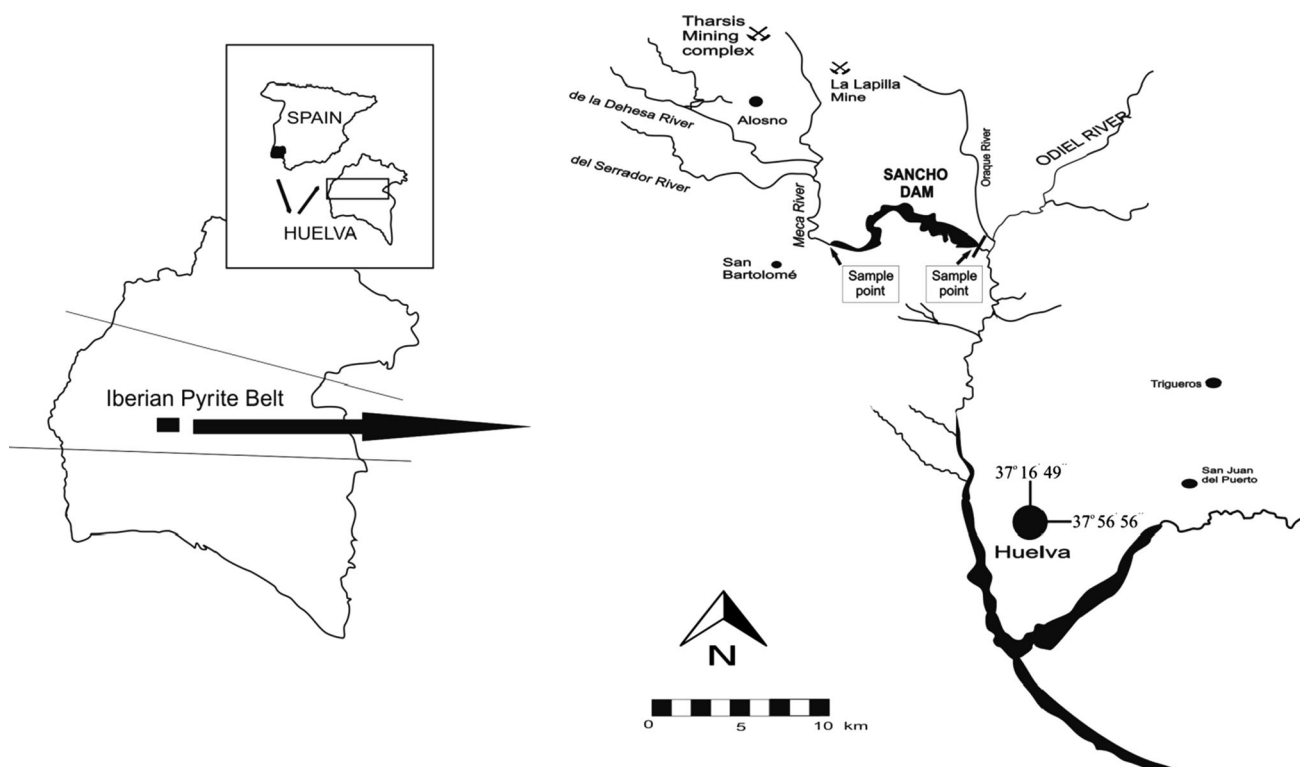
The Sancho Reservoir, located in the Huelva province (SW Spain), has a capacity of 58 MCM (Fig. 1). It is supplied by the waters of the Meca River, which receives contaminated AMD waters from Tharsis mining activities via its tributaries. The dammed water is used to supply a pulp and paper mill factory located 15 km downstream of the reservoir in San Juan del Puerto (SW Spain).

Fig. 1 Location map

Table 1 Statistical data of the variables of the 28 samples obtained at the Rivera Meca-Sancho AMD zone (Grande et al. 2010)

	pH	EC (µS/cm)	Eh (mV)	As (µS/cm)	Cd (µS/cm)	Cu (mg/L)	Zn (mg/L)	Fe (mg/L)	Mn (mg/L)	SO$_4^{2-}$ (mg/L)
Minimum	2.15	2,560	425	0.01	20.11	340.7	190.1	25.7	55.9	5,317
Maximum	2.72	15,810	623	2.11	45.26	607.3	556.5	261	183.6	11,700
Average	2.43	6,035	520.7	0.89	32.36	443.3	440.9	91.4	147.5	7,937
Variance	0.05	2.06E07	4,325.07	0.84	104.52	10,949.8	12,256.8	5,257.17	1,518.21	4.12E06

Table 2 Hydrochemical data of the Sancho Reservoir for October, 2011 (Ceron et al. 2013)

	Minimum	Maximum	Average	Variance
pH	3.5	4.67	3.93	0.23
T (°C)	14.6	25.8	22.4	20.1
CE (µS/cm)	401	472	442	279
TDS (mg/L)	257	302	283	116
DO (mg/L)	0.16	6.16	3.4	5.51
Fe (mg/L)	0.54	1.16	0.755	0.02508
Cu (mg/L)	0.418	0.807	0.6362	0.00763
Zn (mg/L)	1.493	2.723	22.717	0.12881
Mn (mg/L)	1.079	1.677	1.484	0.01409
Cd (mg/L)	0.108	0.147	0.132	9.2E−05
Ni (mg/L)	0.00947	0.01049	0.01028	6.61E−08
Sb (mg/L)	0.0012	0.0034	0.0023	5.2E−07
Pb (mg/L)	0.275	1.912	10.268	0.12669
Al (mg/L)	0.0278	0.2018	0.1029	0.00267
SO$_4$ (mg/L)	170	199	182	91.2

Fig. 2 Representative diagram of the sampling points
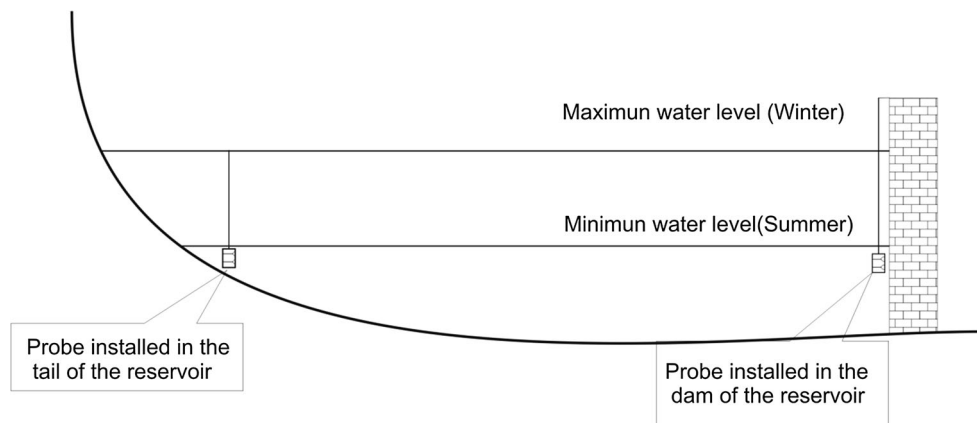
The Tharsis mining complex is responsible for the worst levels of contamination by AMD recorded in the Odiel River basin (Sarmiento et al. 2009). Tables 1 and 2 show the physicochemical parameters of the Meca-Sancho Rivera AMD zone and the hydrochemistry of Sancho Reservoir, respectively.

Ceron et al. (2013) and Grande et al. (2013) defined vertical hydrochemical variations in the Sancho water dam and the establishment of potential vertical and horizontal stratification patterns with respect to metals and sulphate. Sarmiento et al. (2009) conducted a study tracking the accumulation and release of metals associated with AMD in the Sancho Reservoir, analysing the impact of these cycles relative to water quality and sediment.

The application of fuzzy-logic tools to the characterization of pollution processes provides more consistent answers than classical statistics (Aroba et al. 2007; Grande et al. 2005). Grande et al. (2005) characterized the hydrochemistry at the Chorrito stream, located in the IPB, and describes the dynamic behavior of the riverbed. Jimenez et al. (2009) applied fuzzy logic and data mining to the acquired data from a multi-parameter probe located in the Chorrito stream, which allowed him to establish a qualitative pattern describing the relationship between pH and conductivity in an AMD-contaminated channel. Grande et al. (2005), using classical statistics, compared the geochemical behavior of two channels, the AMD-contaminated Chorrito stream and the uncontaminated Higuereta stream, both located in the IPB. This comparison focused on the relation between the variables studied, including pH and conductivity.

Objectives and Methods

This study focused on the relationship between three physical parameters, temperature (T), pH, and electrical conductivity (EC), with rainfall. The Meca River collects AMD from the Tharsis Mine and flows into the Sancho Reservoir. Daily precipitation data were taken from a State Meteorological Agency (AEMET) monitoring station located at San Bartolome de la Torre. The T, pH, and EC data was collected using two multi-parameter probes (Hydrolab Company, Minsonde and MS5 models) suspended from buoys at a depth of 0.5 m from the water surface. One probe was located near where the Meca River flows into the Sancho Reservoir and the other near the reservoir's dam (Fig. 2). They were programmed to simultaneously measure T, pH, and EC every 30 min at the two sampling points. The probes were maintained (cleaning of sensors, calibration, and battery replacement) weekly, and data were dumped into a laptop computer. The monitoring lasted from 1 October 2012 to 2 April 2013, which corresponds to the rainy season in this area, and therefore, with the period of the year when water flows through the channel that carries AMD. This makes a full series of 184 days during which a total of 8,825 data points on pH and EC were collected every 30 min. This large mass of data was processed using fuzzy logic and data mining using PreFuRGe (Aroba 2003; Aroba et al. 2007).

Statistical Methods

Fuzzy logic is a highly effective technique for diagnosing the extent that the environment is affected by agricultural, urban, and industrial activities. Using this technique allow one to obtain very valuable information that would be impossible using only classical statistics. Fuzzy logic (Zadeh 1965) operates using reasoning rules that closely approximates a human's intuitive way of thinking. The main characteristic of fuzzy logic is that it allows one to define values without being precise. This is not possible with classical logic, which is typically binary in nature: one is either a member or not. Fuzzy logic allows for relationships of samples with a certain degree of associated sets. This degree is called the *membership degrees* of an element of a set, S. The set X represents a *range* of values of the variable x, with the range spanning from 0 to 1,
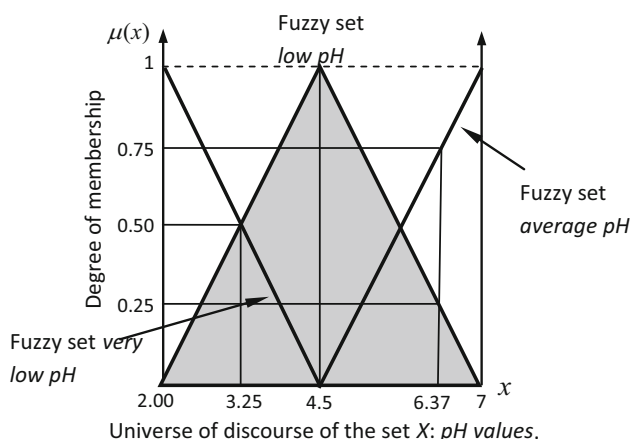
Fig. 3 Example of membership functions used to codify a set of pH values by means of fuzzy logic

where each extreme value represents an absolute, either membership or non-membership, to the set under consideration. The membership grade can be represented by shape functions, normally trapezia, triangles, or sigmoidal. For example, working with pH data that range from 2.00 to 7.00, the universe of discourse for pH variation could be described by the following fuzzy sets (Fig. 3): *very low pH, low pH,* and *moderate pH.* The fuzzy sets at the extremes are right-angled triangles, and the core set is an isosceles triangle (highlighted in gray for better understanding). Then, the expression, *the pH of the sample is average* is true, with a grade of 0.75 for a sample with pH 6.37. However, for the same samples, the expression *the pH of the sample is low* is true, with a grade of 0.25.

Once all of the variables involved in a problem are coded to the qualitative domain applying membership functions criteria, then it is possible to write a set of rules describing the relationships that exist between input and output variables. These rules are set up in an *if–then* format, consisting of an antecedent and a consequence; the fulfillment of an antecedent leads to a conclusion. The main characteristic of reasoning based on rules of this type is the ability to represent and consider partial coincidence, allowing fuzzy rules to provide interference, even when a condition or premise is only partially satisfied. For

example: *if x is A, then y is C* or *if x is A, and z is B, then y is C,* where *A, B,* and *C* are fuzzy sets, and *x, y,* and *z* are variables defined in their respective universes of discourse. The first rule has a single antecedent and the second has a compound antecedent.

Classical clustering algorithms generate a partition of the population where the assignation to each case is linked to only one cluster. These algorithms use a *rigid partition* derived from classical set theory: the elements of the partition matrix (obtained from the data matrix) can only contain values 0 or 1, where zero means null memberships and one means full membership.

Fuzzy partition is a generalization of the previous case, so it holds the same conditions and restraints for its elements, except in this case, real values between zero and one are allowed and reveal partial membership grade. Therefore, samples may belong to more than one group, increasing their selecting and clustering capacity. The best known general-purpose fuzzy clustering algorithm is the so-called fuzzy C-means (FCM) (Bezdek et al. 1984), which can be used to build fuzzy models (Sugeno and Yasukawa 1993). We have adapted and improved the following aspects of these familiar methodologies:

1. Quantitative databases can be worked with n input and m output parameters,
2. The different variables, which are the object of the study, could be assigned with different weights in order to calculate distances between points in a partitioned space.
3. The fuzzy clusters obtained are processed by another algorithm to acquire graphic trapezia rules.
4. An algorithm processes and solves cases of multiple projections in the input space (mounds of data).
5. The output provided in the original method has been improved with a graphic interface showing graphics of the rules obtained.
6. An algorithm automatically provides the graphic interpretation of the fuzzy rules in a natural language.

The result is PreFuRGe (Aroba 2003), a computer tool that provides graphical fuzzy rules of the type shown in Fig. 4, which illustrates the graphical fuzzy

Fig. 4 Example of graphical fuzzy rule
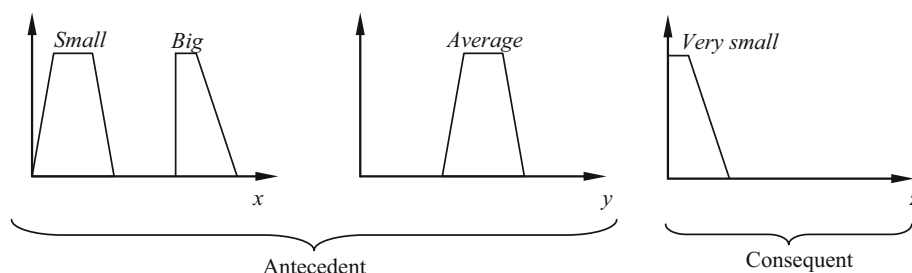
rule: *if x is small or big and y is average, then z is very small.*

Results

Figures 5, 6, 7, and 8 show the fuzzy rules obtained from the data collected by the multi-parameter probes installed at the Sancho Reservoir describing the behaviour of the variables under study: rainfall, pH, EC, temperature (T), and sampling period expressed in months (Month). All information collected for the different variables is represented in 6 rows and 5 columns. The furthest right column represents the consequence; this graphic represents how the considered parameters (antecedents) behave relative to the

consequent value. Notice how, for each fuzzy rule, the affected variables shown at the bottom of the figure represents its degree or range of dispersion. There are several numerical values that correspond with the universe of discourse of the variable in question, which, in classical statistics, is known as the range or amplitude of the variable.

The semi-quantitative nature of this tool, applied to the scenarios described, allows them to be sorted into five classes, depending on the values taken by each variable within its own universe of discourse. So, the existence of extremely low, low, medium, high, and extremely high values is feasible, along with all possible intermediate combinations, e.g. medium–low and medium–high. In the same context, Fig. 5 shows the rules obtained for the up-
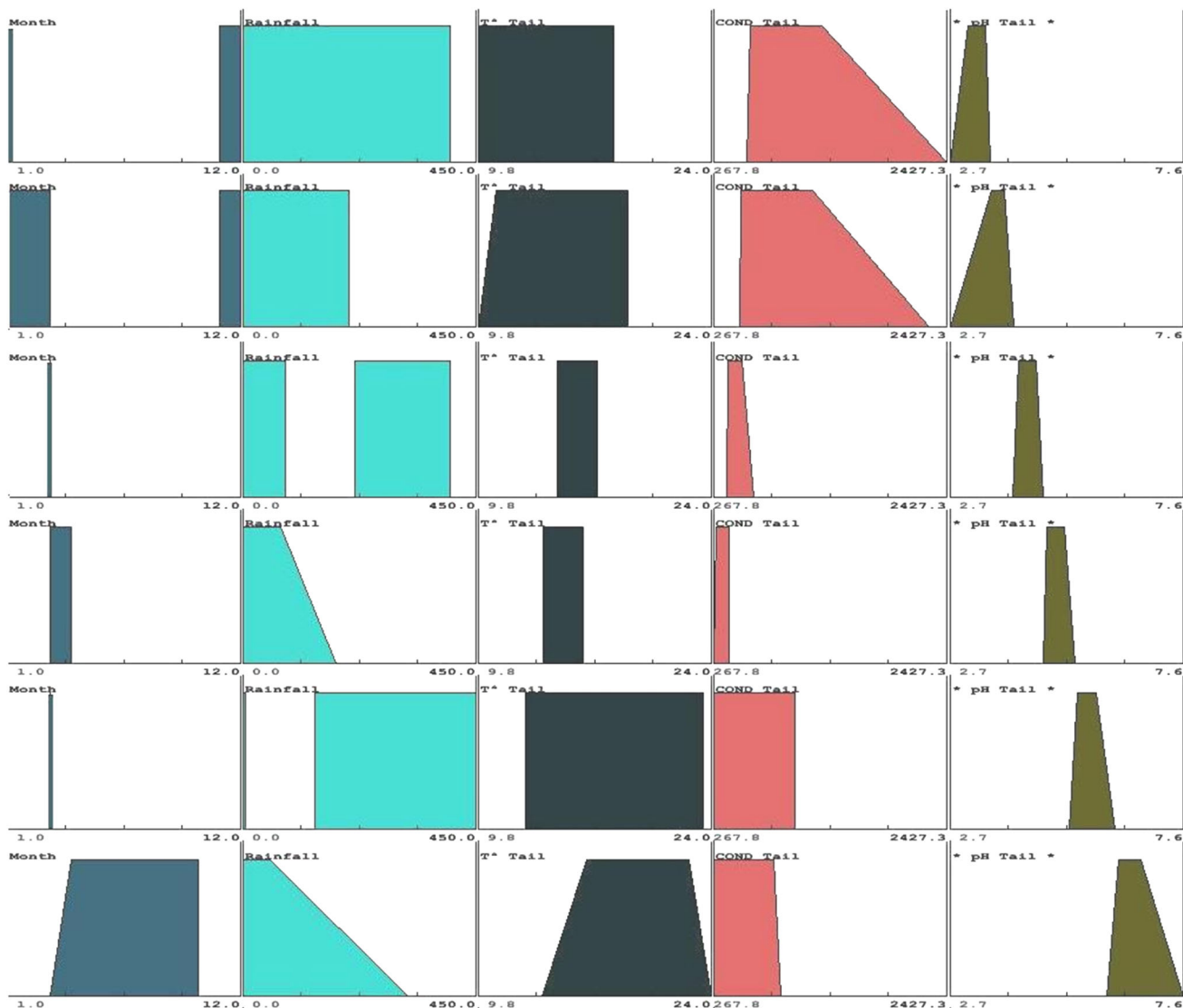

Fig. 5 Fuzzy rules taking pH as the consequent and the rest of the variables at the tail as the antecedent

gradient reservoir probe (which is subsequently referred to, per Spanish terminology, as the "tail" of the reservoir, which unfortunately happens to be the opposite of standard English terminology) using pH as the consequence. The pH values increase along with the months and temperature, while the EC presents the opposite behaviour, decreasing as pH increases.

It should be highlighted that at average pH, both EC and T show very close values, from low to extremely low, and from medium to low, respectively. This occurs during the months of March and April when rainfall varied from extremely low to high.

Figure 6 shows the rules obtained at the down-gradient sampling location, near the dam, with pH as the consequence. In this case, the behaviour during the months and

the temperatures versus pH is totally opposite what is observed at the tail of the reservoir. The pH has an inverse relationship with months and T; as the months and the temperature increase, the pH value decreases. The relationship between pH and EC is also different from what was observed at the tail. The EC does decrease as the pH increases from extremely low to medium–high values; however, for high and extremely high pH values, there is no relationship between pH and EC. As in the case at the tail, when the pH in the reservoir takes average values, the EC has very well defined values that go from low to extremely low during the months of March and April, when rainfall ranges from extremely low to medium.

Figures 7 and 8 show the rules obtained for the two monitoring points, with rainfall as the consequence. In
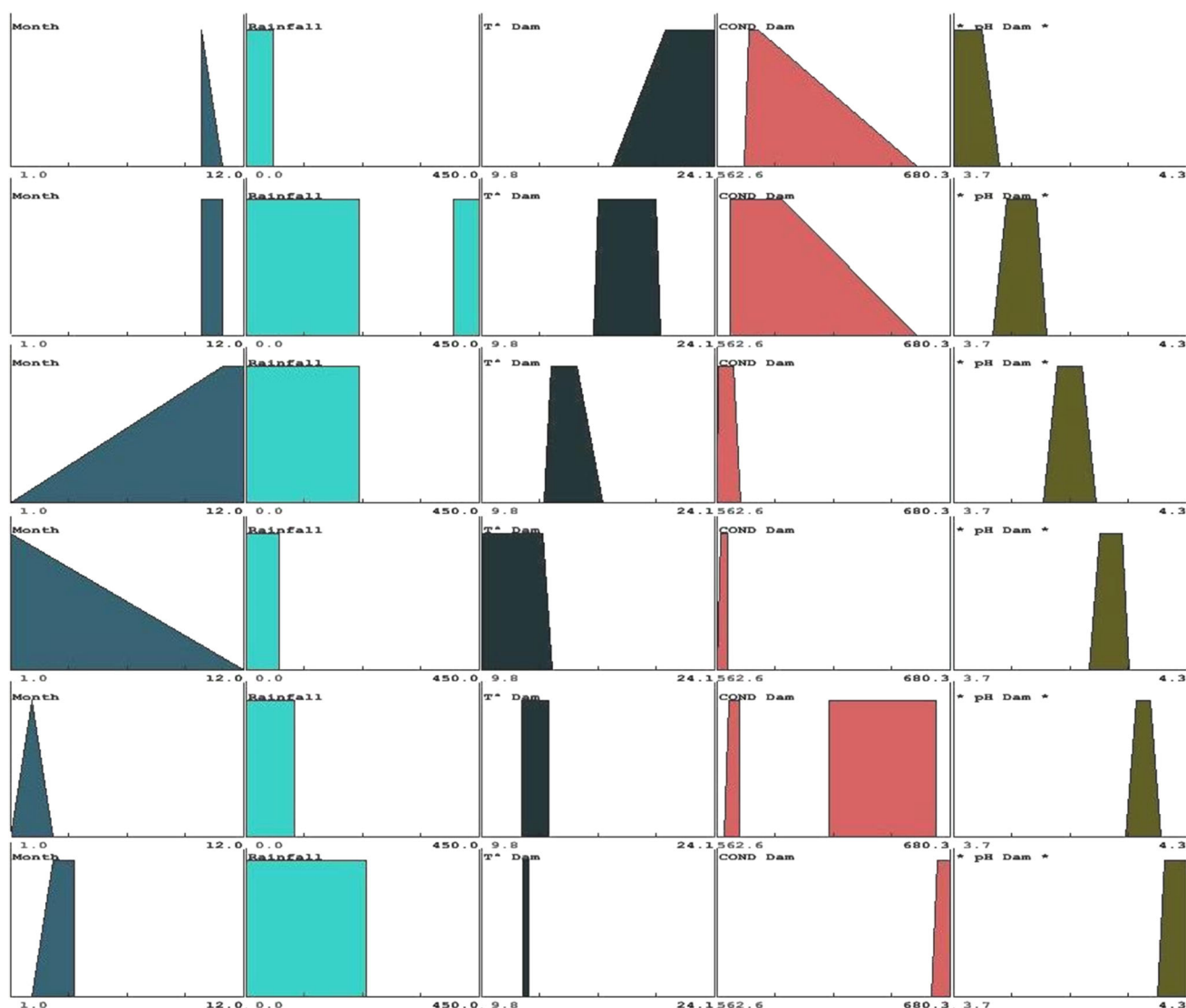

Fig. 6 Fuzzy rules taking pH as the consequent and the rest of the variables at the dam as the antecedent
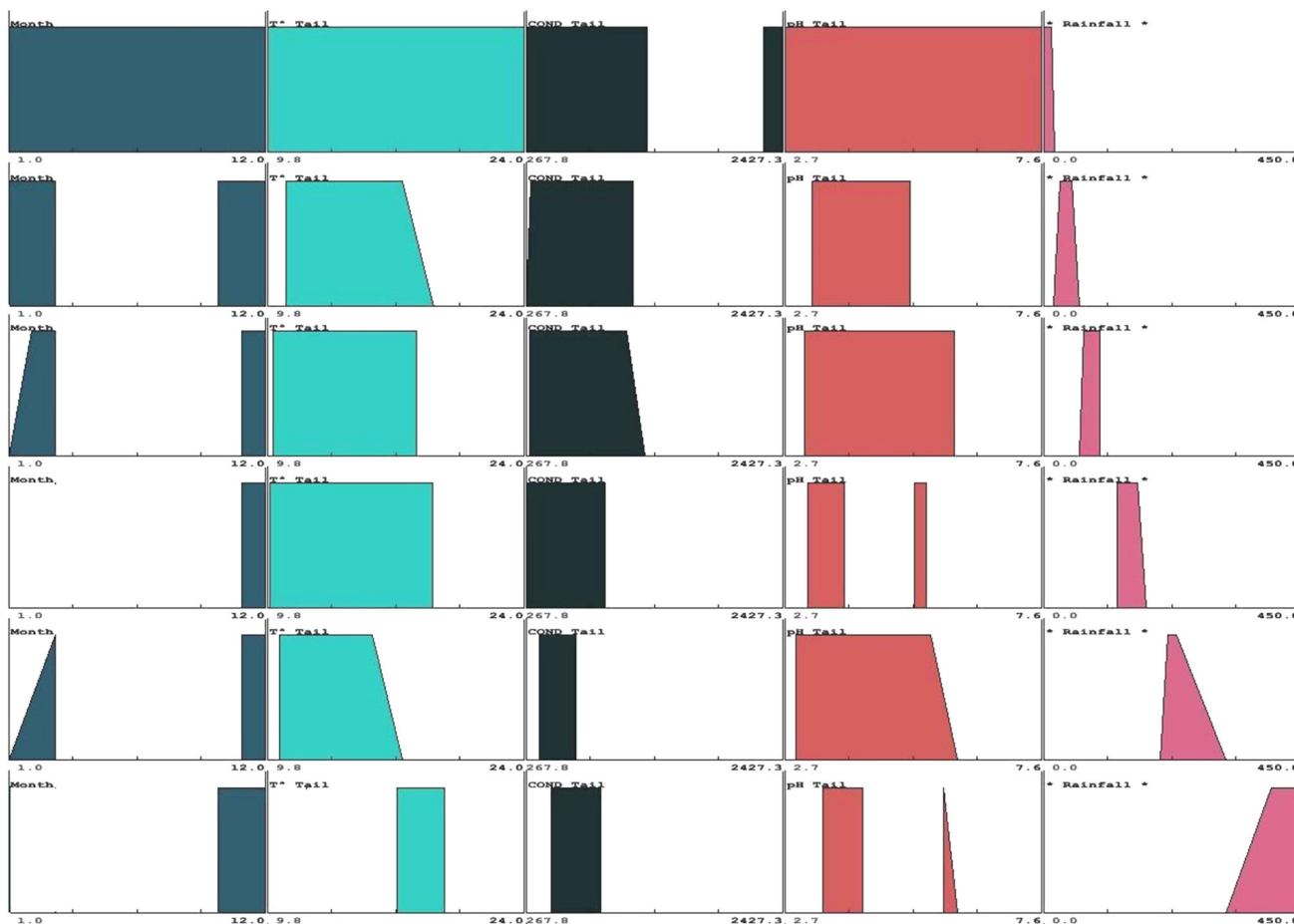
Fig. 7 Fuzzy rules taking rainfall as the consequent and the rest of the variables at the tail as the antecedent

general, it can be seen that when rainfall is extremely low, the other parameters appear random, with no clear relationship between rainfall and the other parameters. However, looking at months and temperature, we can see that the behaviour at the two monitoring points is similar, although unusual in both cases.

Discussion and Conclusions

This application of fuzzy logic and data mining for AMD characterization confirmed operational models that had been previously proposed for this area using classical statistics. PreFuRGe provided a remarkable dimension of overall efficiency for qualitative diagnosis. It was also used to establish cause–effect relationships that, in contrast with classical statistical treatments, made understanding the processes involved easier.

Several authors have reported the absence of a correlation between pH and conductivity in AMD environments (Dogan 1999; Grande et al. 2005; Kwong and

Lawrence 1998; Liew and Sheppard 2001; Younger 1997; Younger et al. 2002). But according to Grande et al. (2010), pH has a high, negative correlation with EC. This should be expected since the sulfate oxidation process generates, on the one hand, sulfate, which increases EC, and hydrogen ions, which acidifies water and decreases pH (Grande et al. 2010). This highly negative correlation between pH and EC was confirmed at the tail of the reservoir, but was less evident near the dam. This can be explained by dissolution processes associated with contaminant transit and the high volume of water in the reservoir. As expected, higher pH corresponded with lower EC values, as a direct result of sulphate concentrations in the virtual absence of chloride (Grande et al. 2010). The behaviour of the variables versus rainfall was similar at both monitoring locations except that there were more outliers near the dam due to the distance between the AMD source area and the sampling point.

As a final conclusion, we can say that the PreFuRGe tool, already implemented in other AMD environments, is
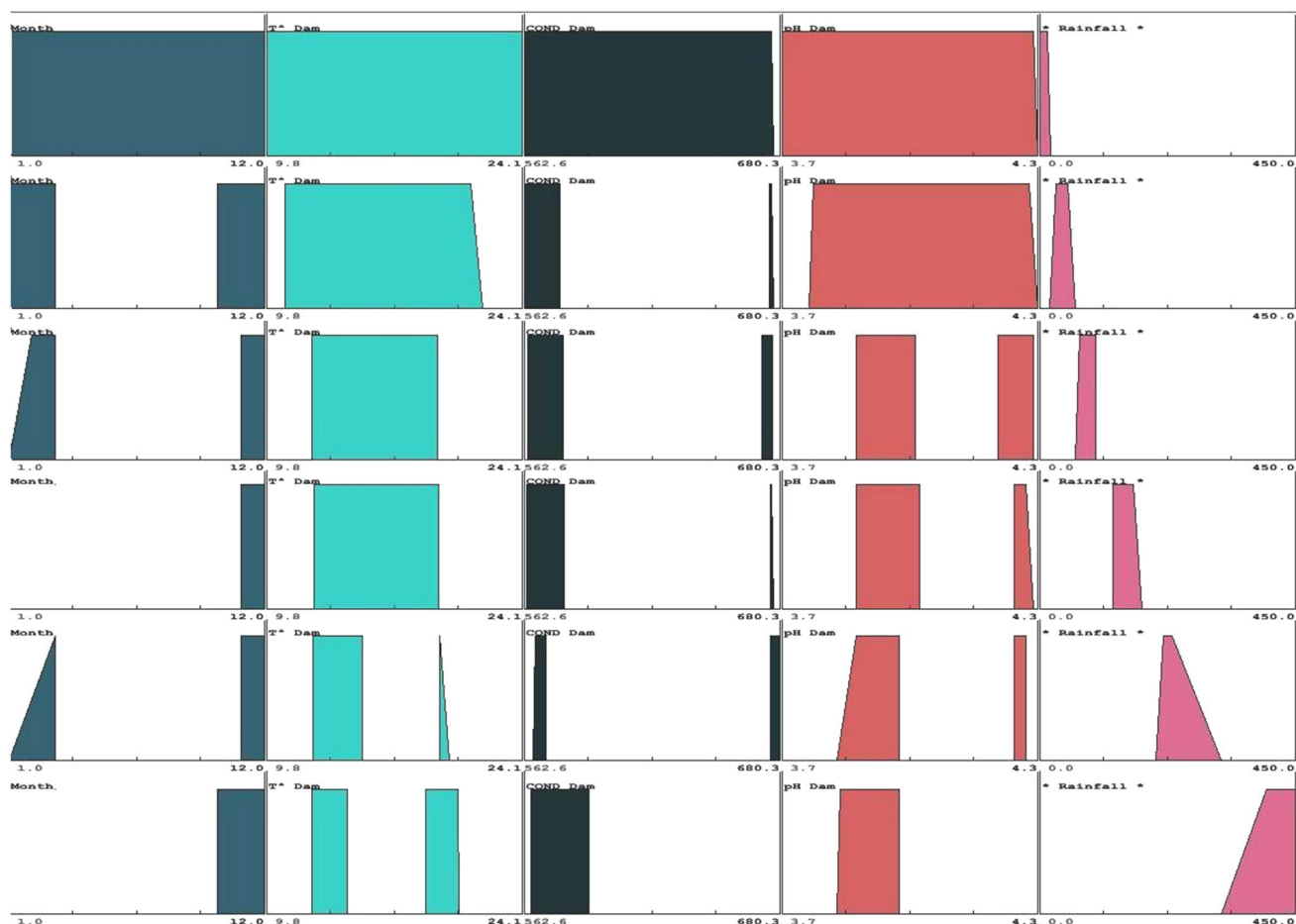
Fig. 8 Fuzzy rules taking rainfall as the consequent and the rest of the variables at the dam as the antecedent

an effective tool for studying this kind of environment. It made the dependency relationships of variables, as well as the cause–effect relationship with the rainfall, obvious.

References

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

Aroba J, Grande JA, Andujar JM, de la Torre ML, Riquelme JC (2007) Application of fuzzy logic and data mining techniques as tools for qualitative interpretation of acid mine drainage processes. Environ Geol 53:135–145

Bezdek J, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203

Ceron JC, Grande JA, de la Torre ML, Borrego J, Santisteban M, Valente T (2013) Hydrochemical characterization of an acid mine drainage-water reservoir affected: the Sancho Dam (Huelva, SW Spain). Hydrol Sci J 59(6):1213–1224. doi:10.1080/02626667.2013.834341

Dogan PA (1999) Characterization of mine waste for prediction of acid mine drainage. In: Azcue JM (ed) Environmental impacts of mining activities. Springer, Berlin

Grande JA, Andújar 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 Monitor 7(4):325–334

Grande JA, de la Torre ML, Ceron JC, Beltrán R, Gómez T (2010) Overall hydrochemical characterization of the Iberian Pyrite Belt. Main sources generating acid mine drainage (Huelva, SW Spain). J Hydrol 390:123–130

Grande JA, Valente T, de la Torre ML, Santisteban M, Ceron JC, Perez-Ostalé E (2013) Characterization of acid mine drainage in the Iberian Pyrite sources belt: basic methodology for quantifying affected areas and for environmental management. Environ Earth Sci. doi:10.1007/s12665-013-2652-0

Jimenez A, Aroba J, de la Torre ML, Andújar JM, Grande JA (2009) Model of behaviour of conductivity versus pH in acid mine drainage water, based on fuzzy logic and data mining techniques. J Hydroinformatics 11(2):147–153

Kwong YTJ, Lawrence JR (1998) Acid generation and metal immobilization in the vicinity of a natural acid lake in Central Yukon Territory, Canada. In: Geller W, Klapper H, Salomons W (eds) Acid mining lakes. Springer, Toronto

Liew D, Sheppard J (2001) Use of conductivity to monitor the treatment of acid mine drainage sulphate reducing bacteria. Water Resour 35(8):2081–2086

Santisteban M, Grande JA, de la Torre ML, Valente T, Cerón JC (2013) Acid mine drainage in semi-arid regions: the extent of the

problem in the waters of reservoirs in the Iberian Pyrite Belt (SW Spain). Hydrol Res. doi:10.2166/nh2013086

Sarmiento M, Olías M, Nieto JM, Cánovas C, Delgado J (2009) Natural attenuation processes in two water reservoirs receiving acid mine drainage. Sci Total Environ 407:2051–2062

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

Valente TM, Leal Gomes C (2007) The role of two acidophilic algae as ecological indicators of acid mine drainage sites. J Iber Geol 33(2):283–294

Younger PL (1997) The longevity of mine water pollution: a basis for decision-making. Sci Total Environ 195:457–466

Younger PL (2001) Mine water pollution in Scotland: nature, extent and preventative strategies. Sci Total Environ 265(1–3):309–326

Younger PL, Banwart SA, Hedin RS (2002) Mine water, hydrology, pollution, remediation. Kluwer Academic Publ, Dordrecht

Zadeh LA (1965) Fuzzy sets. Inf Control 83:338–353