Expert system for applying coagulant in water treatment: case study in Nobres (Brazil)

Ricardo Augusto Moraes Zaque[a], Welitom Ttatom Pereira da Silva[a,b,*] and Aldecy de Almeida Santos[a,b]

[a] Postgraduate Program in Water Resources, Federal University of Mato Grosso, Cuiabá 78090-600, Brazil

[b] Department of Sanitary and Environmental Engineering, Federal University of Mato Grosso, Cuiabá 78090-600, Brazil

*Corresponding author. E-mail: wttatom@gmail.com

Abstract

In most Brazilian municipalities, the water supply comes from surface water sources, requiring infrastructure and human resources, especially water treatment. Where infrastructure and human resources are lacking, auxiliary methods and/or substitutes are valuable. The aim of this work was to develop an expert system (ES) to determine and apply coagulant dosing, in real time, in a water treatment plant (WTP). The ES was developed and evaluated at the WTP in Nobres, Mato Grosso (MT) state, Brazil. It is called AIASD (artificial intelligence for aluminium sulphate dosage) and was validated using the Turing test. It can determine and apply coagulant dosing in water treatment. AIASD contributed successfully to WTP operations in Nobres and other WTPs; making decisions and applying the coagulant dosing necessary for water treatment.

Key words: aluminium sulphate, artificial intelligence, dosing control

INTRODUCTION

In approximately 56% of Brazilian municipalities, the water supply comes from surface water sources (IBGE 2008). Thus, by and large, water needs to be treated to fit the country's drinking water quality standards, as regulated under Consolidation Ordinance (PRC – Portaria de Consolidação) No. 5, dated September 28, 2017, Annex XX, according to the Brazil Ministry of Health (2017a).

IBGE (2008) also reports that about 54% of municipalities with water treatment plants (WTPs) use complete treatment cycles, including coagulation, flocculation, decantation, filtration and disinfection processes.

In many Brazilian cities, significant deficiencies exist in water and sanitation infrastructure, and the planning process. Abers (2010) notes that among the factors causing this are the lack of modern technology and a shortage of human resources.

In order to reduce the response time in determining coagulant dosing and avoid dependence on expensive water treatment structure, this study was designed to develop an expert system (ES) to determine appropriate coagulant dosing.

STUDY AREA

The municipality of Nobres, Mato Grosso (MT) state, Brazil (Figure 1), was chosen as the study area on the basis of issues raised in the Sanitation Sector Modernization Program (Brazil Ministry of Cities 2008): (1) common problems with sanitation in small municipalities – MT is a state composed of 97%
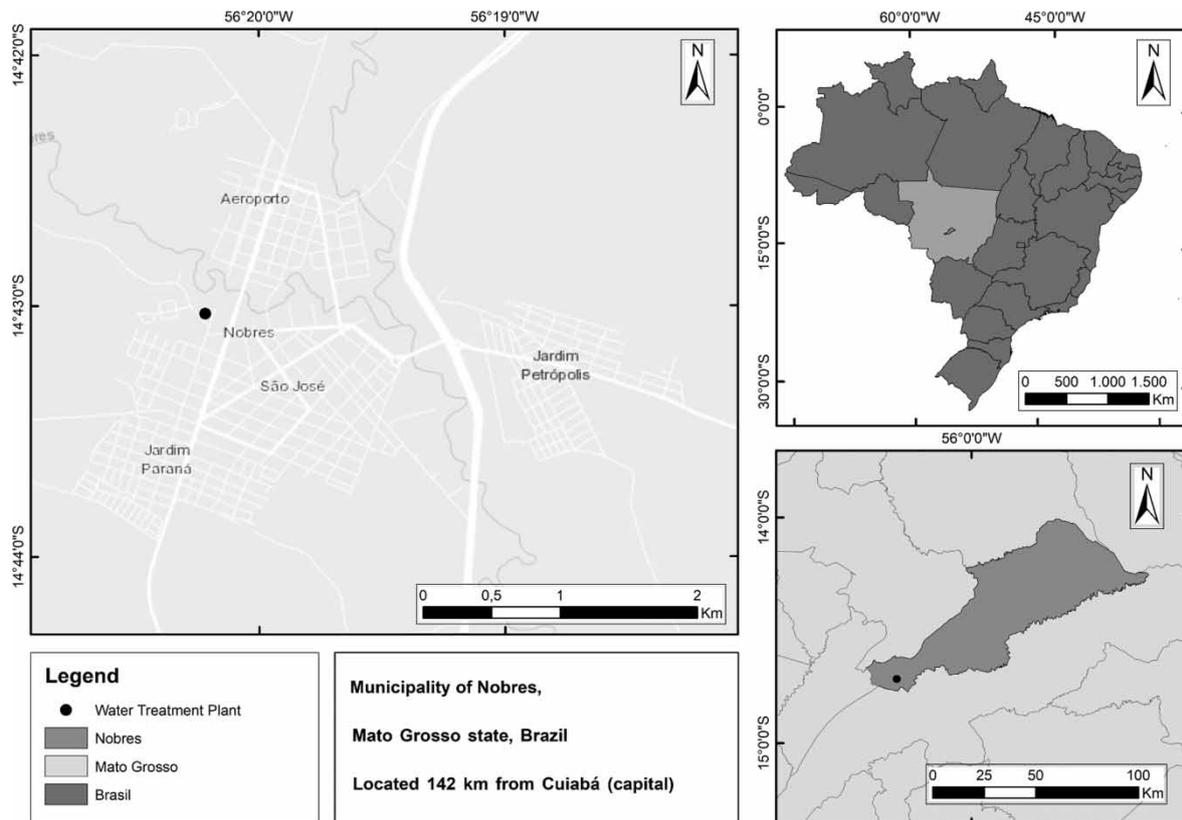
Figure 1 | Case study location.

of small municipalities (Brazil Ministry of Cities 2009); (2) the relative homogeneity, in terms of technical and economic limitations, experienced by these municipalities; and (3) the inability of most municipalities in MT to provide, regulate or control basic sanitation services satisfactorily. In addition, the fact that the water supply service concessionaire in Nobres was interested in this case swayed the choice for the study location.

Nobres has an estimated population of 15,000 inhabitants (IBGE 2010), distributed across about 3,900 km². The local economy is driven by the cement and limestone industry, livestock and dairy farming, agriculture, trade, perennial cultures and ecotourism (Brazil Ministry of Cities 2008).

In Nobres, about 76% of the population receives a water supply service from the municipality and it is considered reasonably adequate, although it needs adjustments to fulfil its social function (Brazil Ministry of Cities 2008).

A floating pump system was installed in the Rio Nobres to transfer water to the WTP, where the coagulation, flocculation, decantation, filtration and disinfection processes take place. The distribution network serves the entire city, which has an urban population estimated at 12,500 (IBGE 2010).

The water supply service is operated by the Nobres Sanitation Company (ESAN). The WTP has two water treatment modules (WTP-I and WTP-II), of which WTP-I is constructed of masonry and WTP-II is metal. Raw water is conveyed to a Parshall gutter with a throat section of 22.9 cm, which distributes 16 L/s to WTP-I and 34 L/s to WTP-II. Both modules have vertical hydraulic flow flocculators with baffles, two laminar flow decanters, and four down-flow filters with anthracite, sand and gravel. The discharge from the down-flow filters goes into a single – i.e., combined – contact chamber.

Currently, the WTP is operated manually. The operator: (1) collects water samples throughout the WTP; (2) determines physico-chemical parameters (turbidity, colour, pH and free residual chlorine); (3) determines the aluminium sulphate dosage (jar-test or experiment); (4) applies the coagulant

dosing (rotor control installed in the dosing pump set-up pipe); and (5) doses chlorine in the contact chamber. Figure 2 is a schematic representation of Nobres WTP.
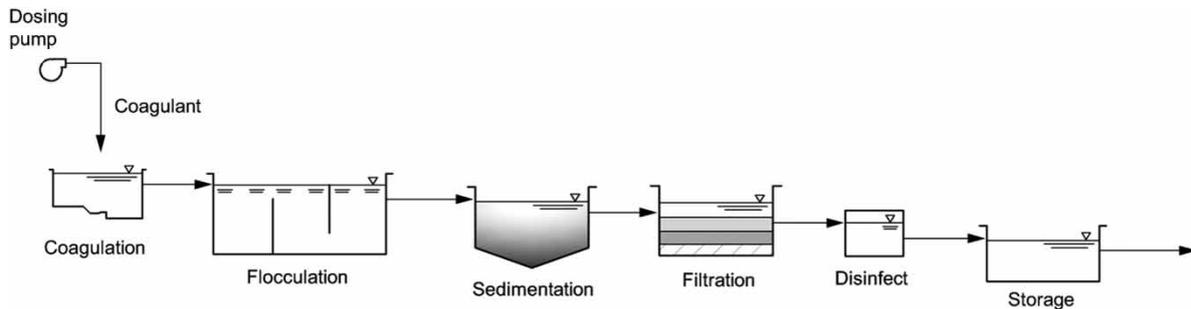
Figure 2 | Schematic representation of Nobres WTP.

METHODOLOGY

The coagulant dosing ES was developed in stages, following recommendations made by Giarratano & Riley (2004), which include: (1) planning; (2) knowledge extraction; (3) knowledge coding; and (4) assessment and adequacy of the ES.

Expert system planning

In the planning stage a formal plan for ES development is produced. Table 1 shows the planning activities and their respective objectives (Giarratano & Riley 2004).

Table 1 | ES planning activities

Activity	Objective
Feasibility assessment	Determining whether the ES approach is appropriate.
Resource management	Assessment of human resources, time, financial resources, software and hardware required.
Preliminary functional layout	Definition of requirements of the system by specifying system functions. Specifying the system's purpose.

Source: Adapted from Giarratano & Riley (2004).

Knowledge extraction

The knowledge extraction stage is that needed to solve the problem (knowledge domain). The activities required include: data collection and tabulation, and knowledge acquisition and assessment.

For data collection, periodic visits were made to Nobres WTP to obtain data about the treatment process. Operating data, especially on raw and treated water quality and quantity, and coagulant dosing were collected.

Details of the method used by the operators to define the dose rate were obtained. These included the variables considered and the form of data analysis. The operating data collected refer to a one-year period, enabling the study to cover both the rainy and dry seasons; typical climatic periods in the region (Brazil Ministry of Health 2017b). The information relating to turbidity (NTU), pH, sulphate dosage (mg/L) and rainfall in the previous 24 hours (24-hour rainfall) were tabulated using a spreadsheet (MS Excel), ready for processing in data mining (DM) software (Bouckaert *et al.* 2010).

Knowledge acquisition was based on the tabulated data using the DM WEKA software (Waikato Environment for Knowledge Analysis) described by Bouckaert *et al.* (2010). WEKA is the only toolkit

that has been widely used over a long period. Jain *et al.* (2010) report that it is a reference system in DM and machine learning research.

It was during knowledge acquisition that the domain knowledge needed to develop the ES was obtained. Production Rules (PR) is the most commonly used knowledge representation model, and was adopted because it has modularity and uniformity advantages (Artero 2009). Three classification algorithms were used: J48, REP Tree and Random Tree. These three were chosen for several reasons:

- J48, a classic algorithm for decision tree (DT) generation;
- REP Tree, an algorithm that generates DT from logical relations between variables; and
- Random Tree, an algorithm that generates DT from random relations between variables.

It was anticipated that some of these approaches (classical, logical, and random) would be ideal for this study.

The Kappa statistics and confusion matrix (CM) were used as precision indicators in knowledge assessment, to select relevant knowledge. Following the approach by Landis & Koch (1977), knowledge was considered adequate if it presented almost perfect agreement (Kappa statistic $0.81 \leq \kappa \leq 1.00$). The algorithm used had its DT converted to PR, following Han & Kamber (2011).

Knowledge coding

The MS Excel spreadsheet (macro development, form utilization and other resources) was used for knowledge coding, on the Windows 10 Platform in an Intel Core i7-4810MQ 2.80 GHz microcomputer with 16GB of RAM. MS Excel is widely known and used, and is thus easy to replicate.

Assessment and adequacy of the ES

Artero (2009) reports that the classic Turing test can be used to verify whether a machine has intelligence at the human level, so this method was used to assess the adequacy of the ES. The test was performed on two people – employees A and B – and one machine, the ES (C). There was no communication between any of them. Employee A was asked to find out which of B and C was the machine. If A could not determine this with at least 50% accuracy, the machine (ES) would pass the Turing test.

RESULTS AND DISCUSSIONS

After planning, the factors and returns were verified, as suggested by Giarratano & Riley (2004), as a first stage in the feasibility assessment. All of the returns proposed by Giarratano & Riley (2004) were favourable to development of the ES – see Table 2 – enabling it to be undertaken.

Table 2 | ES feasibility assessment

Item	Factor[a]	Return[b]	Assessment[c]
1	Can the problem be solved efficiently by conventional programming?	No	No
2	Is the domain of the problem well defined?	Yes	Yes
3	Is there a need for/interest in ES?	Yes	Yes
4	Is/Are there (a) human expert(s) to cooperate?	Yes	Yes
5	Can the expert(s) transmit their knowledge?	Yes	Yes
6	Does the solution to the problem involve mainly heuristics and uncertainty?	Yes	Yes

[a]Factors suggested by Giarratano & Riley (2004).
[b]expected return for the ES approach to be viable.
[c]return found after feasibility assessment of ES development.
Source: Adapted from Giarratano & Riley (2004).

The assessment of item 1 was negative as there is no efficient algorithm to solve the problem, according to Giarratano & Riley (2004). For item 2, the evaluation was positive because the knowledge collected by the operators, as well as historical data on water quality (turbidity and pH of the raw and treated waters, and 24-hour rainfall), can be used as PR and algorithm procedures. The domain is well-defined, which is beneficial for ES development. Concerning item 3, an ES to determine the ideal coagulant dosage is both necessary and interesting to ESAN, as it would reduce water loss and the risk of distributing non-potable water caused by incorrect coagulant dosage. Item 4 yielded a positive evaluation as ESAN made the WTP operating database available, which was considered the expert to help develop the ES. For item 5, the evaluation was also positive as DM techniques can draw the knowledge from the database. The evaluation of item 6 was also positive as the coagulant dosage can be determined on the basis of observation (operator experience), which involves heuristics and uncertainty.

The resource management activity was carried out using a survey of computer (software and hardware), human and financial resources for ES development. The resources available were sufficient according to a comparative analysis based on studies by Zhang & Luo (2004), Wu & Lo (2008) and Santos et al. (2017).

The preliminary functional layout activity should define the system target by specifying its functions. Careful analysis of the ES objectives was carried out to define the system's functions, following Giarratano & Riley (2004).

Daily data were obtained over several visits to Nobres WTP in 2012, including details of turbidity, pH, aluminium sulphate dosage, 24-hour rainfall, etc. The colour characteristic was not considered for the development of the ES, as the operators determine the dosage from the turbidity and pH values of the raw and treated water, and 24-hour rainfall. These data were obtained by interviewing the operators and an operations engineer. Information was also obtained on the executive project, project and current flows, flowcharts, etc.

The data spreadsheet had seven columns containing 61,320 items. The data ranges, etc., are shown in Table 3.

Table 3 | Variables of interest

Variable	Unit	Possible responses
Raw water turbidity	NTU	0.62 to 1,129
Treated water turbidity	NTU	0 to 3.08
Aluminium sulphate dosing	mg/L	10, 11, 12, … , 41
24-hour rainfall	dimensionless	Yes (Y) or No (N)
pH of raw water	pH units	6.5 to 7.7
pH of treated water	pH units	5.8 to 7.3

MD WEKA software was used for knowledge acquisition. Three DTs were generated, using the J48, REP Tree, and Random Tree algorithms. The best of the DTs was selected later and the output attribute was the aluminium sulphate dosage.

The Kappa statistics (κ) and CM results were determined during the knowledge assessment stage. Table 4 shows the Kappa statistics and the percentages of correct and incorrect classifications obtained from each algorithm. The J48 and Random Tree algorithms yielded κ values of almost perfect agreement ($0.81 \leq \kappa \leq 1.00$). The Random Tree algorithm, however, produced the highest proportion of correct classifications at 99.2%, while J48 managed 91.4%. Thus, both the Kappa statistic and Correct Classification Percentage indicators suggest that the Random Tree algorithm would be best for building the knowledge base. Kalmegh (2015) says that the Random Tree algorithm

Table 4 | Algorithm assessment

Adjustment indicator	J48	REP Tree	Random Tree
Correct classifications[a]	6,006 (91.4%)	5,539 (84.3%)	6,520 (99.2%)
Incorrect classifications[b]	564 (8.6%)	1,031 (15.7%)	50 (0.8%)
Kappa statistic (κ)	0.89	0.79	0.99
Total number of classifications[c]	6,570	6,570	6,570

[a]Coincidence of the human expert's response with that of the classification model.
[b]Non-coincidence of the human expert's response with that of the model.
[c]n.

produces a random dataset to construct a DT. Thus, it seems that the database used to develop the ES in this study is not based on defined logic, but has a marked level of heuristics and uncertainty, confirming the result of the ES planning step.

Figure 3 shows the Random Tree algorithm's CM, which confirms that most of the incorrect classifications were between dosages of 20 and 15 mg/L, and 15 and 10 mg/L. There were nine incorrect classifications of each type.

```
  a    b  c  d  e  f   g   h  i  j  k   l   m   n  o  p   q  r  s  t  u   v  w  x  y  z aa ab ac ad ae af ag  <- classified as
2494   0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | a = 0
   0 1599  0  1  0  0   0   0  0  0  0   1   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | b = 10
   0    0  2  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | c = 11
   0    5  0 44  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | d = 12
   0    0  0  0  5  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | e = 13
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | f = 14
   0    9  0  1  0  0 550   0  0  0  0   2   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | g = 15
   0    0  0  0  0  0   1  31  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | h = 16
   0    0  0  0  0  0   0   0 18  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | i = 17
   0    0  0  0  0  0   1   0  0 67  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | j = 18
   0    0  0  0  0  0   1   0  0  0 14   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | k = 19
   0    7  0  0  0  0   9   0  0  2  1 812   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | l = 20
   0    0  0  0  0  0   0   0  0  0  0   2   9   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | m = 21
   0    0  0  0  0  0   0   0  0  0  0   0   2 129  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | n = 22
   0    0  0  0  0  0   0   0  0  0  0   0   0   0 21  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | o = 23
   0    0  0  0  0  0   0   0  0  0  0   0   0   1  0  4   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | p = 24
   0    0  0  0  0  0   2   0  0  0  0   0   0   0  1  0 359  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | q = 25
   0    0  0  0  0  0   0   0  0  0  0   0   0   1  0  0   0 22  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | r = 26
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0 65  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | s = 27
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0 34  0   0  0  0  0  0  0  0  0  0  0  0  0 | t = 28
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | u = 29
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0 150  0  0  0  0  0  0  0  0  0  0  0 | v = 30
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  3  0  0  0  0  0  0  0  0  0  0 | w = 31
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0 50  0  0  0  0  0  0  0  0  0 | x = 32
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0 15  0  0  0  0  0  0  0  0 | y = 33
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | z = 34
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0 19  0  0  0  0  0  0 | aa = 35
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | ab = 36
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | ac = 37
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | ad = 38
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  0 | ae = 39
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  2  0 | af = 40
   0    0  0  0  0  0   0   0  0  0  0   0   0   0  0  0   0  0  0  0  0   0  0  0  0  0  0  0  0  0  0  0  2 | ag = 41
```

**Figure 3** | Random Tree algorithm confusion matrix.

Knowledge base development resulted in 1,397 PR. One of the rules forming part of the knowledge base is presented in Equation (1):

$$
\begin{aligned}
\text{IF} \quad & \text{RT} >= 0.31 \text{ and} \\
& \text{pHT} < 6.95 \text{ and} \\
& \text{RT} < 23.95 \text{ and} \\
& \text{RT} < 4.63 \text{ and} \\
& \text{TT} < 0.39 \text{ and} \\
& \text{Rain} = \text{Y and} \\
& \text{pHT} < 6.85 \text{ and} \\
& \text{pHR} < 7 \text{ and} \\
& \text{TB} < 4 \\
\text{THEN} \quad & \text{Dosage} = 17 \text{ mg/L}
\end{aligned}
\tag{1}
$$

where: RT is the raw water turbidity (NTU); TT the treated water turbidity (NTU); pHT the pH of the treated water; pHR the pH of the raw water; and, Rain the 24-hour rainfall.

Some 75% of the database was used during the knowledge extraction stage, and the remainder (25%) for the assessment and adequacy of the ES.

During knowledge coding, the interface components of acquisition, knowledge base (Equation (1)), inference machine (part responsible for processing the user's questions, in this case, Visual Basic for Applications in MS Excel) and user interface were integrated in MS Excel – see Figure 4. This integration – the ES – is called AIASD (artificial intelligence for aluminium sulphate dosage).
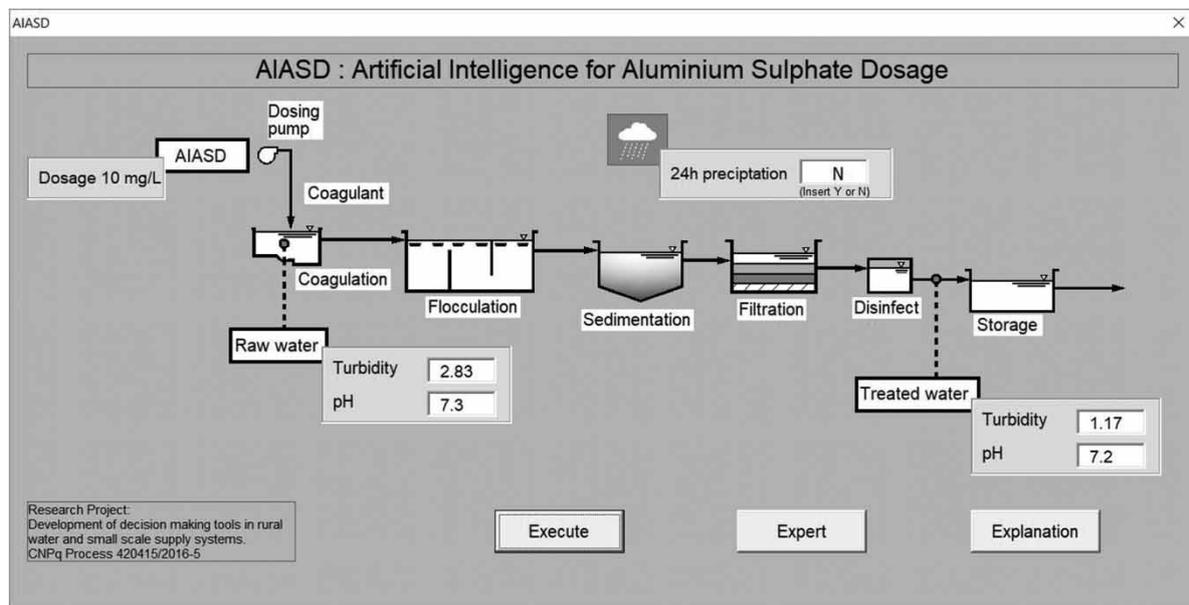
Figure 4 | User interface and responses.

AIASD processing is summarized by comparing the input values with the stored knowledge (production rules) and, finally, supplies the aluminium sulphate dosage that must be applied at Nobres WTP.

In order to use AIASD, the 24-hour rainfall, turbidity and pH data of the raw and treated waters need to be entered. Clicking on the 'Execute' option yields the aluminium sulphate dosage required. Clicking 'Explanation' displays a new screen explaining the production rule enabled (the 'Expert' option allows the addition of new PR, via an Excel worksheet).

In order to demonstrate AIASD, the data shown in Table 5 were entered. The aluminium sulphate dosage obtained was 10 mg/L. Figure 4 shows the response on the user interface.

Table 5 | Input data

Variable	Data
24-hour rainfall	N
Raw water turbidity	2.83
Raw water pH	7.3
Treated water turbidity	1.17
Treated water pH	7.2

The Turing test was applied to the ES using 100 randomly selected cases from the 25% of data not used in developing AIASD. The number of sample units ($n = 100$) was based on the premise that 10 employees from ESAN could respond to the Turing test and that each could analyse 10 cases.

AIASD determined the correct dosage in 95 of the 100 cases – i.e., it obtained the same coagulant dosage as the human expert – and its results could not be differentiated from those determined by humans. The five cases in which the AIASD dosage differed from that determined by a human were also evaluated with the Turing test and indicated that three results were in error. Thus, employee A had a 60% error rate in identifying employee B and the ES.

In total, in 98 of the 100 cases it was not possible to identify the human expert. Thus, AIASD passed the Turing test because expert A could not differentiate it, with at least 50% accuracy, from the human expert.

CONCLUDING REMARKS

An expert system – AIASD – was developed in this study and was considered satisfactory, according to the Turing test. In general, AIASD contributed positively to operations at Nobres WTP. It therefore could be applied to other WTPs, determining the coagulant dosage suitable for water treatment.

ACKNOWLEDGEMENTS

REFERENCES

Abers, R. 2010 Pensando politicamente a gestão da água (Thinking politically about water management). In: *Água E Política – Atores, Instituições E Poder nos Organismos Colegiados de Bacia Hidrográfica no Brasil* (Abers, R. N. ed.). Annablume, São Paulo, Brazil, pp. 13–36.

Artero, A. O. 2009 *Inteligência Artificial: Teórica E Prática (Artificial Intelligence: Theory and Practice)*. Livraria da Física, São Paulo, Brazil.

Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2010 WEKA – experiences with a Java open-source project. *The Journal of Machine Learning Research* **11**, 2533–2541.

Brazil Ministry of Cities 2008 *Sanitation Sector Modernization Program, Municipalization of Water Supply and Sanitary Sewage Services in Mato Grosso State: Diagnosis, Lessons and Perspectives*. Technical report. Available at: http://www.pmss.gov.br/downloads/apoio-a-estados/relatorio-final/nobres.rar (accessed 27 April 2018).

Brazil Ministry of Cities 2009 *National Habitation Plan*. Available at: http://bibspi.planejamento.gov.br/bitstream/handle/iditem/285/Publiicacao_PlanHab_Capa.pdf?sequence=1&isAllowed=y (accessed 25 March 2018).

Brazil Ministry of Health 2017a *Consolidation Ordinance No. 5*, dated September 28, 2017, Annex XX. Available at: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2017/prc0005_03_10_2017.html (accessed 01 August 2018).

Brazil Ministry of Health 2017b *Basic Sanitation Municipal Plan of Nobres – MT*. Available at: http://pmsb106.ic.ufmt.br/wp-content/uploads/2018/04/PMSB_Nobres.pdf (accessed 27 April 2018).

Giarratano, J. C. & Riley, G. D. 2004 *Expert System: Principles and Programming*, 4th ed. Course Technology, Boston, USA.

Han, J. & Kamber, M. 2011 *Data Mining: Concepts and Techniques*, 3th ed. Elsevier, San Francisco, USA.

IBGE 2008 *Pesquisa Nacional de Saneamento Básico 2008 (National Survey of Basic Sanitation 2008)*. Available at: https://sidra.ibge.gov.br/tabela/1364#resultado (accessed 8 August 2017).

IBGE 2010 Instituto Brasileiro de Geografia e Estatística. *Censo Demográfico* 2010 (*Demographic Census* 2010). Available at: https://sidra.ibge.gov.br/tabela/1505#resultado (accessed 8 August, 2017).

Jain, S., Aalam, M. A. & Doja, M. N. 2010 κ-means clustering using WEKA interface. In: *Paper Presented at the 4th National Conference, INDIACom-2010, Bharati Vidyapeeth's Institute of Computer Applications and Management*, New Delhi, India.

Kalmegh, S. 2015 Analysis of WEKA data mining algorithm REPTree, Simple Cart and RandomTree for classification of Indian news. *International Journal of Innovative Science, Engineering & Technology* **2**(2), 438–446.

Landis, J. R. & Koch, G. G. 1977 The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174.

Santos, F. C. R., Librantz, A. F. H., Dias, C. G. & Rodrigues, S. G. 2017 Intelligent system for improving dosage control. *Acta Scientiarum* **39**(1), 33–38.

Wu, G. D. & Lo, S. L. 2008 Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system. *Engineering Applications of Artificial Intelligence* **21**(8), 1189–1195.

Zhang, H. & Luo, D. 2004 Application of an expert system using neural network to control the coagulant dosing in water treatment plant. *Journal of Control Theory and Applications* **2**(1), 89–92.