Fuzzy model to assess the water quality in the Meia Ponte river watershed under different standpoints

Vinícius SEBBA PATTO[1], Gelson DA CRUZ JUNIOR[2], Luis Maurício BINI[3], Marco Antonio ASSFALK DE OLIVEIRA[2]

[1]Laboratoire d'Informatique de Paris 6 - Université Pierre et Marie Curie
104, Av. du Président Kennedy, 75016 Paris, France

[2]Electrical and Computing Engineering School of Federal University of Goiás
Setor Universitário, s/n, 74605-220, Goiânia, Brazil

[3]Biological Sciences Institute of Federal University of Goiás
Campus Samambaia, 74001-970, Goiânia, Brazil.

vinicius.sebba-patto@lip6.fr, {gcruz,assfalk}@eee.ufg.br,

bini@icb.ufg.br

Abstract. *This paper addresses a fuzzy model to assess the quality of the water through the use of three different indices. To compose these indices, we considered the use of the water and the characteristics of the Meia Ponte river watershed. At first, we will present the context, the assessment of water quality considering the complexity of the water, its parameters, and its application, as well as the non-significant relationship between parameters employed to assess its quality. Our main goal is to help the assessment of water quality, avoiding the influence of parameters on each other and taking into account the destination of the water. Then, we will present the model and the fuzzy inference process. At the end, a case study is presented in order to help to compare the results of the three proposed indices with the results obtained from the classical model proposed by the National Sanitary Foundation of the United States to determine the index of the water quality. To finish, we discuss it.*

1. Introduction

In Germany, in the middle of the 20th century, some attempts emerged in order to try to synthesize informations about the water quality. The idea was to relate the quality of water with the occurrence of certain communities of macro and microorganisms [Derísio 2000]. More ahead, in the middle of 60 years, numeric indices appeared, like the Horton index [Horton 1965]. These indices emerged in order to provide numeric scales which could translate the value of some parameters of water in a unique value. This way, it would be possible to represent briefly the informations about a sample of water, independently of the the individual evaluation of an expert in water quality [Nives 2003]. Some regulatory agencies proposed some criteria for water quality. However, most of these indices were based on the WQI (Water Quality Index) proposed by the National Sanitation Foundation (NSF) of the United States of America, in order to designate the general condition of the quality of water, through the use of a numeric value [Said 2004]. Basically, these indices are composed by the following parameters: biochemical oxygen demand within 5 days (BOD5), dissolved oxygen (DO), fecal coliform, potential of hydrogen (pH), temperature, nitrate, total of phosphor, total mass of solid material in the water, and turbidity. In general, these parameters are selected by a team of experts in water quality of a specific local [Nives 2003].

In Brazil, since 1975 year, some indices of water quality were proposed, based on the NSF WQI. However, some modifications were done aiming to adequate them to the local context [CETESB 2001, CETESB 2004]. These modifications are done, in general, according with the opinion of a group of experts in water quality [Karamouz 2004]. This means that these indices become specific, according with the local ecosystem and also, according with the kind of use of that water. So we can realize that they do not really follow a standard. Through this verification, we can observe that these indices for assessment of water quality are pretty influenced empirically and heuristically. So, this research proposes a model to determine three indices to assess the quality of water of the Meia Ponte river watershed (in Goiás State), according with three specific kinds of use of water: to drink (with no treatment), to protect the aquatic biota, and tratament of water.

2. Method

2.1. Data base used in the study

The data employed in this study refers to the parameters read through the 12 months of the year 2004 by Sanitation of Goiás Company (SANEAGO). The water sample were collected in one of the points of water collection for treatment, in Meia Ponte river watershed, near Goiânia (see data in page 10, Annex A).

To determine the three water quality indices proposed, we used six parameters available and provided by SANEAGO: oxygen percentage saturation (OS), biochemical oxygen demand (BOD), total of phosphor (TP), potential of hydrogen (pH), fecal coliform (FC), and turbidity (TURB). Obviously, there are some other parameters also important, that reveal other features of water, however they were not considered because it would increase the complexity of the inference process and it would augment considerable the number of rules (to better understand, see subsection 2.3.2). In our study, these six parameters were enough because they are potentially strong to express important features of water, according with the three proposed types of use. It is important to keep in mind that the conceptualization of these parameters takes into account the features of the water of the Meia Ponte river watershed.

Since SANEAGO does not perform reading of phosphor regularly on this point of the watershed, this parameter was simulated randomly based on the measurements done in other local near to that one from which SANEAGO uses to do regular measurements of the other parameters (OS, BOD, pH, FC, TURB). The simulation of these data were also done with the endorsement of experienced experts with the measurement of physical, chemical, and biological parameters in the Meia Ponte river watershed. The phosphor data did not exceed the maximum measure and was not inferior to the minimum measure done in other point of the river through the same period of this study.

2.2. Proposed Indices

The water quality assessment can be better done if indices are elaborated to verify the quality of water according with the specific use of it. This way, we propose three indices to evaluate water designated to (1) drink, since it is a non-treated water; (2) treatment of water; and (3) the protection of the aquatic biota.

As result of the inference process, the model outputs concepts for each kind of use of water: Bad, Medium (-), Medium, Medium (+), and Good. Then, in order to help the comprehension of the results (which are linguistic concepts) and in order to give a more precise index, the result is defuzzificated for a scale which varies from zero to hundred. The defuzzification process is very advisable since numeric scales are very helpful to precise results and also because human subjects can better understand numeric scales as outputs due their large use.

2.3. Fuzzy logic: general principles and application

As well as in other models that also use fuzzy logics, *e.g.* inference system for prediction of water level in reservoir [Chang 2006], model to estimate pseudo steady state chlorophyll-a concentrations in a very large and deep dam reservoir [Soyupak 2004], and model to diagnose water quality in reservoir [Lu 2002], these indices were proposed based on the representation of the knowledge of experts through the inference rules (also known as production rules).

In a nutshell, the parameters employed in the system were classified by experts in certain linguistic concepts (fuzzy sets). Then, the input data are classified into these concepts. After the activation of the inference rules by these concepts, the indices of water quality are determined. So, after find these indices, we map them to a numeric scale (this process is called defuzzification) [Ross 1997].

2.3.1. Membership functions

In order to create partitions for the universe of discourse of each variable (this process is known in fuzzy logic as fuzzification) and for rules output, we created membership functions. Each variable addresses each parameter of water (presented in section 2.1). For most of our parameters (except for pH) we divided the universe of discourse in five fuzzy sets (Bad, Medium (-), Medium, Medium (+), and Good; see Figures 1 to 6).

Due to the nature of the problem and of the parameters present in our model, we realized that the trapezoidal membership function can better represent the fuzzy sets. In fact, it happens due to the existence of large intervals from which we can verify the total membership of certain parameters in certain fuzzy sets or the contrary - the total absence of these parameters in certain fuzzy sets. This occurrence justifies the existence of plateaus.

However, three observations deserve our attention:

1. The pH parameter has just three membership functions (Bad, Medium, and Good), where just one, Good membership function, has one support (area of the universe of discourse that has elements with non-null membership values), whereas Bad and Medium membership functions have two supports (each one);
2. The membership functions of the universe of discourse for each variable, that are not totally represented in the graphic, extends through the abscissa axis; and

3. The membership function that represents the partition Good for fecal coliform parameter starts in zero with total membership and decays. There is no plateau in this set because there is only one situation where fecal coliform can be considered good: when it is zero. Remembering, we consider three types of use of water.

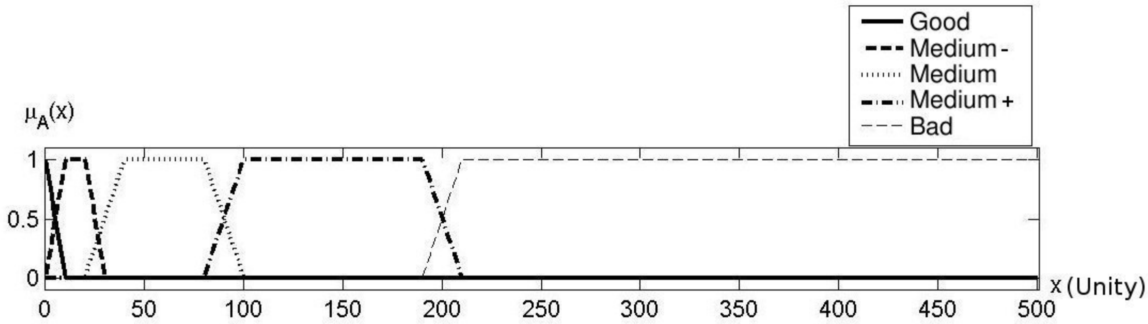

Figure 1. Membership function for fecal coliform

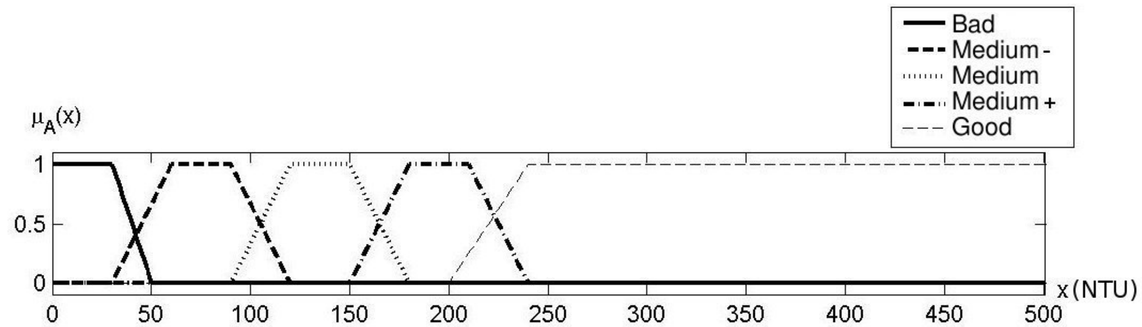

Figure 2. Membership function for turbidity

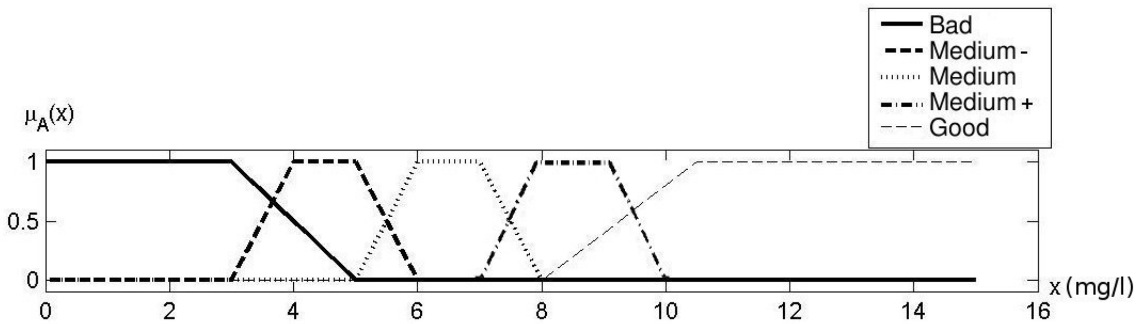

Figure 3. Membership function for biochemical oxygen demand

2.3.2. Decision rules

The number of rules varies according with the possibility of combinations of all possible states of each parameter. The index for assessment of (non-treated) water to drink is composed by three parameters, and each one can assume five possible concepts, then we can attain one hundred twenty-five rules (because $5^3 = 125$). For the conservation of the

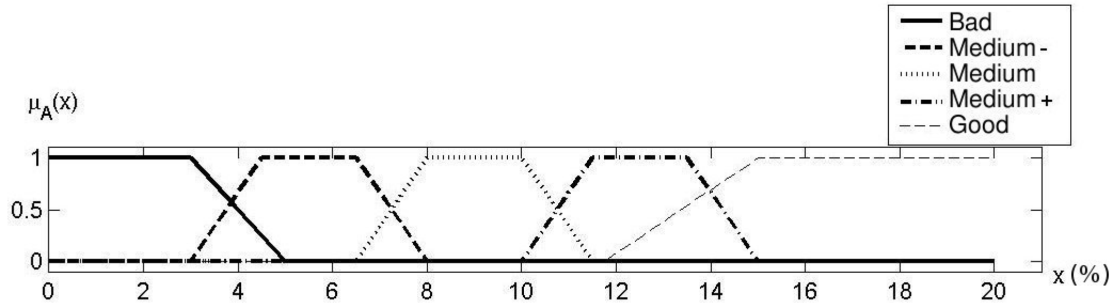

Figure 4. Membership function for oxygen percentage saturation

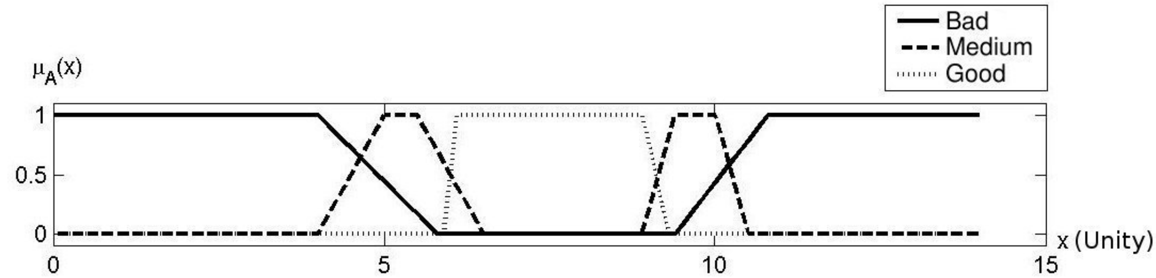

Figure 5. Membership function for potential of hydrogen

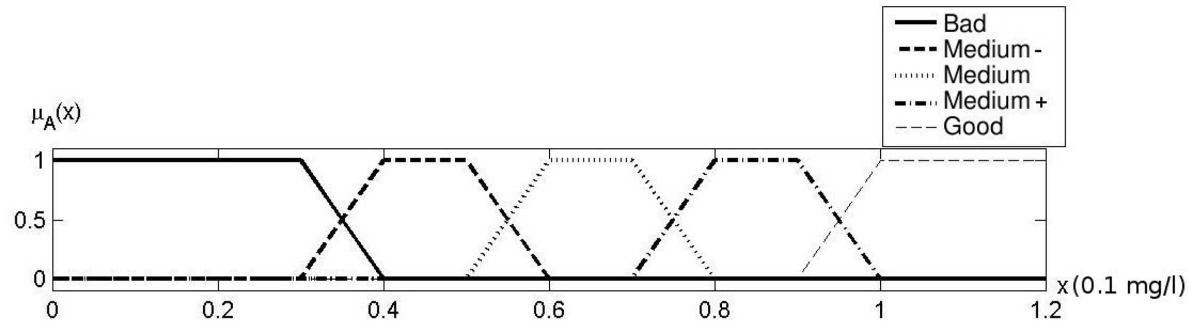

Figure 6. Membership function for total of phosphor

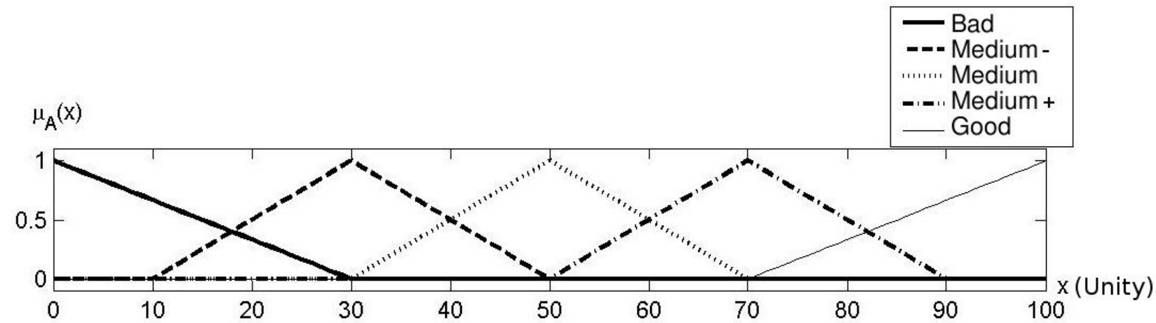

Figure 7. Membership function for output

aquatic biota, the number of combination increases for three thousands one hundred and twenty-five rules ($5^5 = 3125$). Considering now water treatment as the analysis criterion, the number of rules is equal to one thousand eight hundreds and seventy-five possible rules ($5^4 \cdot 3 = 1875$). We have to remember that pH parameter has just three possible concepts (Bad, Medium, and Good) and that is why the calculation to find the number of rules is a little bit different for water treatment.

Two remarks also deserve our attention:

1. Always that a variable belongs to more than one fuzzy set, it triggers more than one rule. Then, if there is only one concept for each variable, these variables trigger just one rule. If one variable has two concepts and the other ones have just one concept, these concepts together trigger two rules, and so on.
2. Even that the number of combinations of concepts gives the number of rules to cover completely the universes of discourse of our model, it is possible to represent all possible situations with a number of rules small than the total of combinations obtained from all possible concepts.

To help understanding this situation, let's consider, for example, non-treated water to drink. In this case, we have the following rule:

$$if\ FC\ is\ Bad\ then\ QI\ is\ Bad$$

Even that we can reach twenty-five combinations with other two variables (BOD and TURB), it is enough that FC is Bad so that the concept of the quality index is also considered Bad. Then, even that the other two variables, BOD and TURB, were classified as Good, the quality index still remains Bad.

2.3.3. Fuzzy reasoning

The process of fuzzy reasoning evolves the combination of composition methods with defuzzification methods [Jang 1997]. The composition methods are applied in order to make "addition" of outputs activated by different rules. The defuzzification process consist of converting fuzzy outputs in non-fuzzy values, it means crisp values [Cox 1994]. In our study we applied three composition methods: max-min, max-max, and limited sum. In order to verify different reasoning processes we used, for each composition method, four defuzzification methods: weighted average, average of maxima, first of maxima, and last of maxima.

For the max-min composition, the consequent fuzzy set is restricted to the minimum of the true antecedent, and the output consequent fuzzy set is composed getting the maxima of these minimized fuzzy sets [Cox 1994]. For the max-max composition, the consequent fuzzy set is restricted to the maximum of the activated antecedent, and the output consequent fuzzy set is composed by the maxima of these maximized fuzzy sets [Ross 1997]. For the limited sum composition, the consequent fuzzy set is given by the minimum of the true antecedent and the output consequent fuzzy set is given by the sum, limited to 1, of these minimized fuzzy sets [Cox 1994].

The weighted average defuzzification method is given by the following equation:

$$x^* = \frac{\sum \mu_A(\overline{x}) \cdot \overline{x}}{\sum \mu_A(\overline{x})} \tag{1}$$

Where x^* is the defuzzified value. To find x^* we have to sum the multiplication of the degree of membership of average values of output plateaus by the average values of plateaus. Then this result should be divided by the sum of the degree of membership of average values of output plateaus.

The average of maxima defuzzified value is given by the following equation:

$$x^* = \frac{a + b}{2} \tag{2}$$

The average is given by the extremities of the plateau that presents the highest membership value.

In order to find the defuzzified value of the output, either for the first or for the last of the maxima, we should find before the highest membership value of the output:

$$h(A_k) = \sup_{x \in X} \mu_{A_k}(x) \tag{3}$$

And then, we find the first of the maxima:

$$x^* = \inf_{x \in X} \{x \in X \mid \mu_{A_k}(x) = h(A_k)\} \tag{4}$$

Or the last of the maxima:

$$x^* = \sup_{x \in X} \{x \in X \mid \mu_{A_k}(x) = h(A_k)\} \tag{5}$$

3. Experiments to analyze our indices

The main focus in this study is not to compare fuzzy reasoning processes, but to help the assessment of water quality avoiding the influence of parameters on each other and taking into account the use of water. In this sense, to better understand and analyze our indices, we present in Table 1 the results of one of our studies case. We present for each index conceptual and numeric outputs as well as a NFS WQI based index for all months of year 2004.

The (mathematical) WQI based indices vary from zero to hundred and were obtained from the following equation:

$$WQI = \prod_{i=1}^{n} q_i^{Wi} \tag{6}$$

Where:

- n = number of parameters;
- q_i = quality of the *ith* parameter (number between 0 and 100, got from the respective quality graphic, which is related to its concentration); and
- Wi = weigh corresponding to the *ith* parameter (number between 0 and 1, attributed according with the relevance of each parameter).

Table 1. Indices to assess the quality of water

Case Study							
Date	Treatment		Biota		Drink		WQI
(m/y)	Concept	Numer.	Concept	Numer.	Concept	Numer.	(mat).
01/2004	Medium-	30	Medium-	30	Bad	4,5	15,45
02/2004	Medium-	26,25	Medium-	26,25	Bad	0	16,61
03/2004	Medium-	21,73	Medium-	21,73	Bad	0	15,94
04/2004	Medium-	30	Medium-	30	Bad	0	18,52
05/2004	Medium-	30	Medium-	30	Bad	2,5	17,6
06/2004	Medium	55,88	Medium	50	Medium	50	25,65
07/2004	Medium	59,49	Medium-	30	Bad	4,28	26,25
08/2004	Medium+	70	Medium-	40	Medium-	18,75	26,96
09/2004	Medium+	70	Medium-	30	Bad	0	23,85
10/2004	Bad	5,5	Medium-	30	Bad	5,5	17,08
11/2004	Medium-	30	Medium-	30	Bad	0	12,29
12/2004	Medium-	30	Medium-	30	Bad	0	16,01

4. Results and discussion

For this fuzzy system, the max-max composition showed more widespread results and its outputs were concentrated around the maximum membership points of the output functions, or between them. The max-min and the limited sum showed less widespread results and their outputs were less concentrated around the maximum membership points of the output functions, or between them. The max-min method showed more smooth output curves comparing with other compositions methods.

About the defuzzification step, we observed that results were smaller or greater than the expected ones for the first and last of maxima, respectively, in comparison with the average of maxima and weighted average. Considering these last two defuzzification methods, the first one (average of maxima) showed influenced by maximum membership points in the output, since the second one (weighted average) showed results influenced by points which are not necessarily the maximum membership points in the output.

Since our main focus is not to compare fuzzy reasoning processes, we presented just one case study based on the max-min composition and the weighted average defuzzification method.

As we can see, the difference of the quality of water can vary strongly, depending on the use of the water (see for example, indices for 2004 September). Then, the use of indices like the NSF WQI can lead people to mistake and to fail when they try to assess the quality of water. It happens because in these models we can not disaggregate the influence of one parameter on the other one and vice-versa. This effect is known as eclipse, and it means that one information can be hidden by the general result, which emerges from the assemble of informations provided by all parameters together.

Since we employed fuzzy techniques to determine the indices, we can assure that an important information is not hidden - because we can disaggregate influences of parameters on each other. The rule about fecal coliform (page 6) illustrate very well what it means: it is enough that FC is bad to classify the water to drink as bad. It does not matter how good can be the other parameters, this kind of water is always considered bad to drink. Weighted parameters, like in Equation 6 (for NFS WQI based indices) can be helpful, but it can not really eradicate influences of parameters on each other.

It is important to use different types of indices to verify the water quality under different perspectives of use. It can identify for example that, a non-advisable water to drink can be advisable for treatment, or vice-versa. Certainly, other indices can be proposed *e.g.*, indices to asssess water quality for agriculture irrigation, for fish culture, to identify most advisable kind of treatment for water, etc.

In natural systems, sometimes we can not assess features through the use of mathematical formulas which aggregate parameters, such as in water quality assessment, because the influence of parameters or variables can hide important informations at the final result. It can lead to errors and mistakes that can be catastrophic.

And last but not least, we believe that the composition of water assessment indices, like these proposed above, can be helpful and more efficacious because they consider the kind of use of water and they disaggregate the influence of parameters in the inference process.

Acknowledgments

We thank SANEAGO for all supplied data. This research was funded by CNPq CT-Hidro.

References

CETESB (2001). Relatório de qualidade das águas interiores do estado de São Paulo. Technical report, Companhia de Tecnologia de Saneamento Ambiental.

CETESB (2004). Relatório de qualidade das águas interiores do estado de São Paulo. Technical report, Companhia de Tecnologia de Saneamento Ambiental.

Chang, F. J., C. Y. T. (2006). Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. volume 9, pages 1–10. Elsevier.

Cox, E. (1994). *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*. Elsevier Science & Technology Books, 1rst edition.

Derísio, J. C. (2000). *Introdução ao controle de poluição ambiental*. Signus Editora, 2nd edition.

Horton, R. K. (1965). An index-number system for rating water quality. pages 300–306. Water Pollution Control Fed.

Jang, J.-S. R., S. C.-T. M. E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach To Learning And Machine Intelligence*. Prentice-Hall.

Karamouz, M., J. N. K. R. (2004). River water quality zoning: a case study of karoon and dez river system. volume 1, pages 16–27. Iranian Journal of Environmental Health Science and Engineering.

Lu, R.-S., L. S.-L. (2002). Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. *Water Research*, 36:2265–2274.

Nives, S.-G. (2003). Comparison of dalmation water evaluation indices. volume 75, pages 388–405. Water Environment Research.

Ross, T. J. (1997). *Fuzzy Logic with Engineering Applications*. McGraw-Hill, Inc., 1st edition.

Said, A., S. D. K.-S.-G. (2004). An innovative index for evaluating water quality in streams. volume 34, pages 406–414. Environmental Management.

Soyupak, S., C. D.-G. (2004). Fuzzy logic model to estimate seasonal pseudo steady state chlorophyll-a concentrations in reservoirs. volume 9, pages 51–59. Springer.

A. Data provided by SANEAGO from 2004 January - 2004 December

The data used in the case study, presented in page 8, Table 1, were obtained from the Table 2.

Table 2. Data employed for the case study

Meia Ponte River Watershed Parameters						
Date.	Parameters					
(m/y)	pH	%OS	BOD	TP*	TURB	FC
01/2004	7,54	5,5	3,6	0,1	152	30
02/2004	7,11	4,5	0,3	0,2	82,8	900
03/2004	7,16	4,1	0,3	0,3	132	8000
04/2004	7,19	5,4	0,3	0,4	80	5000
05/2004	7,62	6,1	0,3	0,5	55	2600
06/2004	7,68	7,4	0,2	0,6	39	70
07/2004	7,69	7,3	0,8	0,7	35,7	2200
08/2004	7,77	8	1,2	0,8	23,3	200
09/2004	7,85	7,9	0,3	0,9	16,4	400
10/2004	7,54	5,6	1,2	1	161	2700
11/2004	7,74	6	0,4	0,1	85	1700
12/2004	7,36	5	0,5	0,2	140	2600

* In its scale TP is given by $x \cdot 10^{-1}$