# Zadeh-Deshpande Approach for Fuzzy Description of Air and Water Quality

**Ashok Deshpande[1], Jyoti Yadav[2]** and **Vilas Kharat[3]**

*Submitted in September, 2013; Accepted in February, 2014*

***Abstract - Ever increasing pollution levels due to rapid urbanization and industrialization, especially in many developing countries with minimal focus on adequate pollution abatement strategies, has resulted in impairing the natural environment such as river, air, land, and alike. It is, therefore, important to classify the environmental quality. The practice in vogue for classifying air or water quality for variety of usage is by computing Air Quality Index (AQI) or Water Quality Index (WQI). Why compute a numeric AQI or WQI, and then describe air/ water quality linguistically? Human brain does not compute any numbers. Why not describe air or water quality, for the defined usage, straightway in linguistic terms with some linguistic degree of certainty attached to each linguistic description?***

***In this paper, we present two need based research studies. The first case study refers to fuzzy air quality description in Pimpri-Chinchwad Municipal Corporation (PCMC) monitoring location, while the other relates to linguistic classification of water quality with degree of certainty in PCMC area, India,***

***Index Terms – Air quality, water quality, uncertainty, Linguistic term, degree of match, fuzzy logic, fuzzy rule base system, degree of certainty***

## 1.0 INTRODUCTION

Information on the status of environmental quality is necessary to formulate sound public policies and effective implementation of environmental quality programs. One of the effective ways of communicating such information in general, and air and water quality in particular to policy makers and ultimately to the end users, is with indices.The efforts on the development of water quality indices -expressed in an interval scale in numeric terms were initiated in 1990's in USA. Air/ water quality index (AQI or WQI) is in essence a function used to simplify large quantities of data into a more useful form which might convey an image of overall air/water quality to variety of users.

[1]*Berkeley Initiative in Soft Computing (BISC) - Special Interest Group (SIG) - Environment Management Systems (EMS), College of Engineering Pune (COEP) Pune India*
[2,3]*Department of Computer Science, University of Pune*
*E-mail: [1]ashokdeshpande@hotmail.com,*
*[2]jyo_yadav@yahoo.co.in and [3]laddoo1@yahoo.com*

The limitation of the selected indices is their lack of power in dealing with uncertainty and subjectivity present in the expert's perception in describing air/water quality in linguistic terms such as: *Good, Fair, Poor and so on* for the defined purpose. In spite of the best efforts, there has been no general acceptance of AQI/WQI for the intended usage.

There exists *imprecision / fuzziness* in the perception of domain experts' about rating environmental quality parameters in various linguistic terms as *epistemic uncertainty* -an unavoidable feature of most humanistic systems. Uncertainty is also present due to inherent variability in randomness in environmental quality parametric data and is known as -statistical or *aleotary uncertainty*. We believe that fuzzy logic concepts could be considered as a useful formalism for modeling, both, *aleotary and epistemic uncertainty.*

The paper is organized as follows: Section 1 is Introduction covering the relevance of the proposed need based research. The write up in Section 2 describes fuzzy logic based formalism used in the study. Application of fuzzy logic concepts in two case studies is presented in Section 3. The first part of the study refers to air quality description in PCMC with linguistic degree of certainty attached to each linguistic description while the second part describes water quality in Pawana river near PCMC intake well straightway in linguistic terms with some degree of certainty The results and discussion are also included in this section. Section 4 relates to the conclusion and the need for further research.

## 2.0 THE METHODOLOGY

Fuzzy logic based approach used in describing air and water quality in PCMC area is detailed in this section.

### 2.1 Type 1 Fuzzy Inference System (FIS) with Degree of Match (DM)

Randomness in air quality data or for that matter any parametric data, can be represented using Gaussian distribution with probability density function given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

where $x$ is a random variable taking values on real line, $\mu$ is the mean and $\sigma$ is the standard deviation. The continuous random variable can be discretised using distribution

$$\Phi(x) = P[X \le x] \quad (2)$$

such that probability of any $x_i$ in the parametric domain, denoted by $p(x_i)$ is :

$$p(x_i) = \emptyset\left(x_i + \frac{dx}{2}\right) - \emptyset\left(x_i - \frac{dx}{2}\right) \quad (3)$$

where $dx$ is a small interval about $x_i$ for $i$ = 1,2,3,4,……n.
Thus if $x_i$ is some point on the parametric domain for which p ($x_i$) is maximum, then define function $\mu_A(x)$ as:

$$\mu_A(x) = p(x)/p(x_i) \quad (4)$$

The operation in equation 6 transforms a random variable into a convex normalized fuzzy number A with membership grade $\mu_A(x)$, thereby characterizing the dynamic behavior of the governing parameter. The construction of fuzzy number or convex fuzzy sets for modeling perception of the experts' in classifying each parametric domain linguistically, involves:

- Selection of linguistic terms such as Very Good, Good, Fair etc which allows for referencing all possible parametric values to be described.
- Classification of the parametric domain and assigning linguistic terms to each class linearly by the experts', reflecting imprecision in their perception. The set of values for which all the experts' assign the same linguistic term are given membership value as $\mu = 1.0$ while none of the expert assigning that term are given membership value as $\mu = 0.0$. The breakeven point membership grade 0.0 and 1.0 are connected by continuous monotonic function which presupposes that the degree of consensus amongst the experts' goes on increasing as the parametric values approach the core of the fuzzy number for the specified linguistic term.

### 2.2 Matching between two fuzzy values

The fuzzy number for pollutant data (A) on parameters and the fuzzy number characterizing linguistic terms ($\tilde{A}$) are matched together to arrive at a measure called Degree of Match (DM) (Figure 1 and 2) defined by :

$$DM_{ij}(A, A') = \frac{\int \mu_{A \cap A'}(x)\,dx}{\int \mu_A(x)\,dx}, \; x \in X \quad (5)$$

in which X denotes the universe, and $\mu_{A\cap\tilde{A}}(x)$ is the membership grade for $A \cap \tilde{A}$. A (fact or the data) and $\tilde{A}$ (expert's perception) are the possibility distributions the measure is defined as:

$$DM_{ij}(A, A') = \frac{\sum \mu_{A \cap A'}(x)\,dx}{\sum \mu_A(x)\,dx}, \; x \in X \quad (6)$$

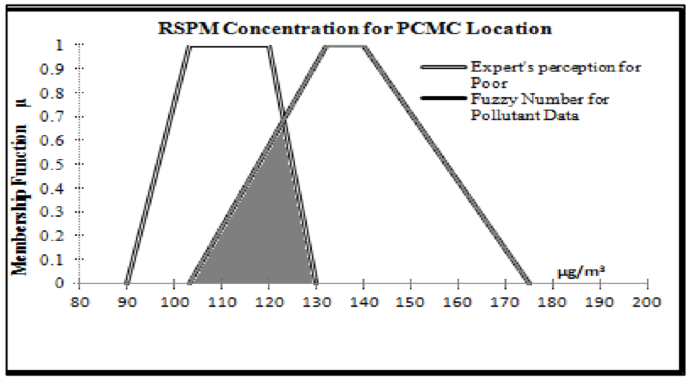


**Figure 1: Degree of Match of RSPM with Linguistic Term Poor for PCMC location**

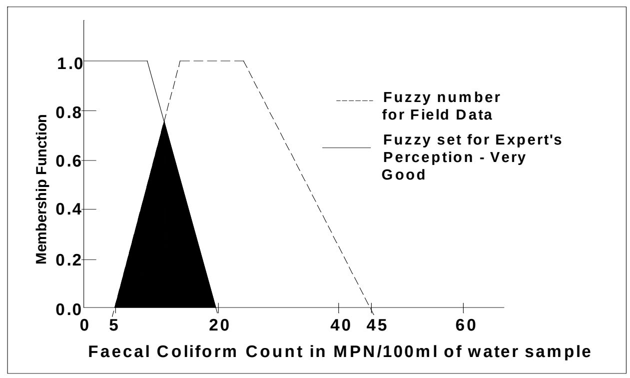


**Figure 2: Fuzzy matching between A and A'**

### 2.3 Fuzzy Rule Based System

The computational scheme of Degree of Match (DM) can be used with a view to estimate matching between expert's perception and the antecedent part of the rule, in order to describe air quality fuzzily with certain degree of certainty [3]. The degree of match of each classification rule indicates the certainty value of classification, in the present case: air and water quality. The greater the degree of match, the greater is the possibility that air and water quality is classified in that class. A fuzzy rule based system was developed for knowledge representation or reasoning process. A set of fuzzy rules is constructed for classifying air quality as: *very good, good, fair, poor and very poor* in order to aggregate the set of attributes. These linguistic descriptions are invariably imprecise / vague / fuzzy keeping in view the inadequate information on the health implications of each parameter on the users and the aggregated effect of all the parameters on human health. Sample rules displayed in Table 1 were stored in the knowledge base. The rules are processed using conjunction and disjunction operators. The optimal acceptance strategy is usually that for which the degree of assertion is the maximum [4]. Figure 3 presents the hierarchical structure for water quality classification [6] and Figure 4 represents the Fuzzy Inference System for Air Quality classification.

.

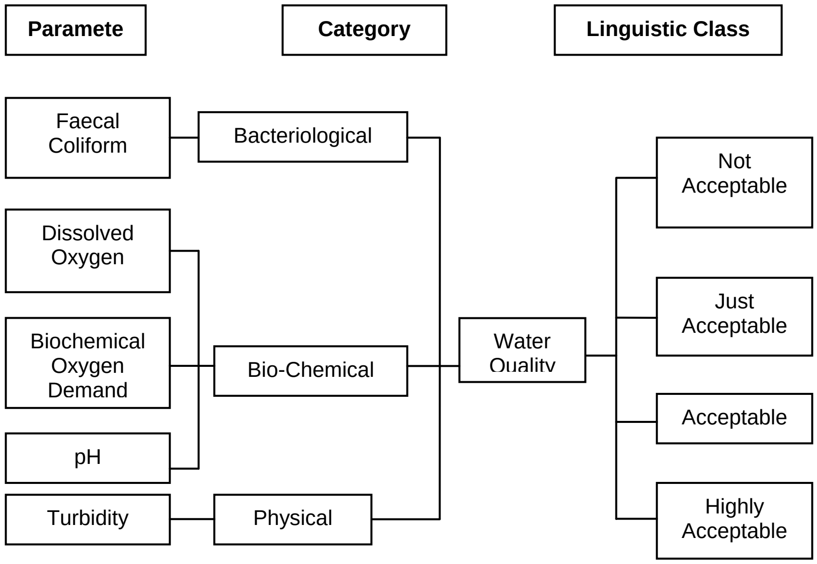


**Figure 3: Hierarchical structure for water quality classification**

| Sr. No. | Air Quality Fuzzy Rules | Water Quality Fuzzy Rules |
|---|---|---|
| 1. | If RSPM is Good AND<br>$NO_x$ is Poor AND<br>$SO_x$ is Good THEN<br>Air Quality is Good | IF BOD is *Good* AND<br>DO is *Very Good* AND<br>pH is *Very Good* THEN<br>Biochemical status of water is *Good* |
| 2. | If RSPM is Fair AND<br>$NO_x$ is Good AND<br>$SO_x$ is Fair THEN<br>Air Quality is Fair | If Bacteriological status of water is *Fair* AND<br>Biochemical status of water *Very Good* AND<br>Physical status of water *Very Good* THEN<br>over all water quality at water intake well is *Fair*. |

**Table 1: Fuzzy Rules by Air and Water Quality Domain Expert's**

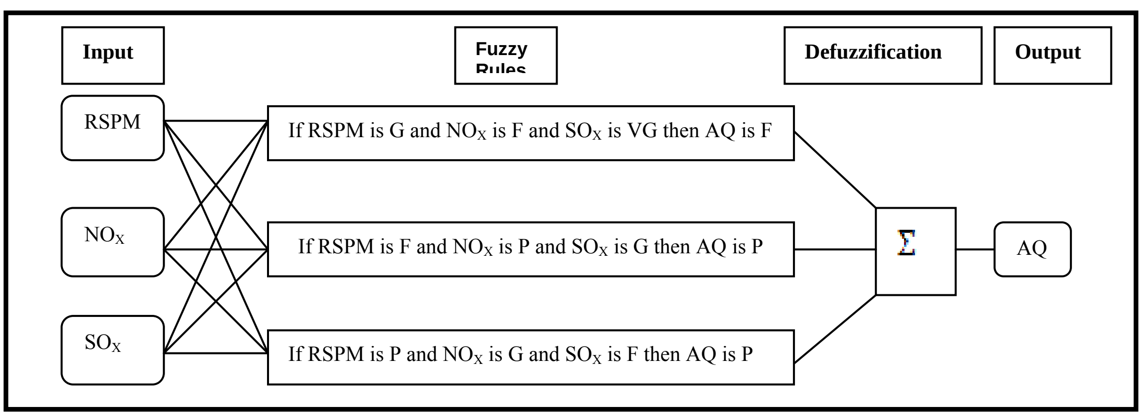


**Figure 4: Fuzzy Inference System for Describing Air Quality**

## 3. 0 CASE STUDY

Application of fuzzy set theory has been a long standing need while dealing with engineering system in general and air and water quality classification in particular.

### 3.1 Linguistic description of Air Quality with Linguistic Degree of Certainty

Air pollutants are added in the atmosphere from variety of sources that change the composition of atmosphere and affect the biotic environment. The concentration of air pollutants depends not only on the quantities that are emitted from air pollution sources but also on the ability of the atmosphere either to absorb or disperse these emissions. The air pollution concentration vary spatially and temporarily causing air pollution pattern to change with different locations and time due to changes in meteorological and topographical condition. The sources of air pollutants include: vehicular traffic, emissions of hazardous gases from industries, domestic sources and natural sources [7].

Air pollution is the *fifth* leading cause of death in India after high blood pressure, indoor air pollution, tobacco smoking and poor nutrition, with about 620,000 premature deaths occurring from air pollution related diseases (Times of India News dated 14/2/2013). Especially in the developing countries, a large portion of vehicles on city roads use diesel fuel, and contribute greatly to the emissions of particulates, especially those that are less than 10 microns in size and are respirable. In many cities, transportation fuels contain lead and high amounts of sulphur and use older engine designs that emit more toxic pollutants than modern ones. Various other factors like predominance of old outdated vehicles, lack of maintenance, limited use of emission control technologies, poor traffic management and poor road conditions helped accentuate the level of automobile pollution. The number of two/three wheeler vehicles which use two stroke engines and produce up to 10 times size hydrocarbons are on increase [7].

Air quality monitoring network design needs structured approach which involves : identification of pollutants, selection of locations, frequency and duration of sampling, sampling techniques, infrastructural facilities, man power operation, maintenance costs, and alike. The design also depends on type of pollutants in the atmosphere through various common sources, called common urban air pollutants- such as Suspended Particulate Matter (SPM), Respirable Suspended Particulate Matter (RSPM), Sulphur Dioxide ($SO_2$), Oxides of Nitrogen ($NO_x$), and Carbon Monoxide (CO) etc. Factors to be considered for the identification of air quality monitoring locations include: high traffic density, industrial growth, human population and its distribution, emission source, public complaints, if any, the land use pattern etc. Realising the increase in air pollution levels, it was considered relevant to initiate a study to classify the air quality in PCMC.

The case study relates to fuzzy air quality description with the available air quality data. The locations of air quality monitoring stations are: 1. Pimpri-Chinchwad Municipal Corporation, and 2. Bhosari in PCMC. Auto exhaust pollution is on increase in PCMC area. From air quality standpoint, winter months are more important due to temperature inversion phenomenon. November 2010 air quality parametric data is considered for the assessment of ambient air quality at PCMC location. Following three pollutants monitored by the Maharashtra State Pollution Control Board (MPCB) viz. Oxides of Sulphur ($SO_x$), Oxides of Nitrogen ($NO_x$), and Re-Suspended Particulate Matter (RSPM) is used for the assessment of air quality following the concept of fuzzy inference system with degree of match as detailed in this paper.

### 3.2 Fuzzy description of Pawana river water quality near PCMC intake well

The developing countries have been witnessing pollution of water resources which has assumed a serious threat to mankind due to increase in the incidence of water related diseases. The World Health Organization (WHO) in their report (2002) states that around 21% of communicable diseases in India are water related, and out of these, diarrhoeal diseases alone killed over 7,00,000 Indians in the year 1999 [2]. The present study refers to the water quality near intake well of water treatment plant of PCMC India. PCMC depends mainly upon Pawana river water. Water quality surveillance from the source to consumer tap is routine activity of major water supplies. In India, the issue of sample size using statistical methods is invariably ignored. How many samples to be collected from a sampling location? In this study, we have implemented the concept which is described below:

Total coli form is one of the important parameters in describing water quality for drinking purpose. A monitoring study was intended to estimate the mean concentration of total coli form near the intake well of PCMC near Water Treatment Plant. A preliminary survey consisting of fourteen (n=14) representative observations, were: 1200, 1300, 1600, 1100, 400, 600, 1800, 300, 1000, 440, 350, 120, 460, 1250 expressed in MPN/100ml of sample. Biochemical Oxygen Demand (BOD) expressed as mg/l were also in the range of 6- 18 mg/l, indicating pollution at the intake point on Pawana river. Looking at these high values, the then Municipal Commissioner (MC) who is also an engineer, requested the first author on suggesting the sample size as he was keen to direct PCMC officials to carry out sampling based on basic statistics. The purpose of these investigations was to initiate action on the defaulters -a sugar mill owner, sewage from nearby colonies, villages to take necessary corrective measures.

1. What sample size is needed to estimate the true mean within ± 200 units? Also, suggest a remedial measure to bring down drastically the bacterial load i.e. say, ≥ 500 MPN/100 ml of water sample.

*Sizing the Experiment or sample size*

Assume: data will follow a normal distribution.

$$\text{Number of Samples } N = \frac{(Z_{\alpha/2}\,\sigma)^2}{E^2} \quad (7)$$

where, σ is the standard deviation( 524 MPN/100ml of water sample); Z is $[(x_i - \bar{x})/s]$ =1.96 ; for α/2 equals 0.025 and E (true mean within ± 200 units) = half length E=200. The computed number of samples will be = 26.37, say a total of 30 samples should be collected. Therefore PCMC officials should collect additional (30-14) = 16 sample for the estimation of total coli form.

Similar procedure has been followed for deciding the number of samples for all the other water quality parameters, viz. Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), pH, Turbidity and was used in the deciding the water quality near PCMC intake well using fuzzy logic based formalism.

## 4.0 RESULTS AND DISCUSSION

The Air Quality description with Degree of Match method describes air quality straightway in linguistic terms with *linguistic* degree of certainty. Uncertainty of uncertainty is modelled using type I fuzzy logic. Firstly the uncertainty in the expert's perception is modelled and then the second uncertainty related to Degree of Certainty (DC) is modelled by defining fuzzy sets for DC as Very Good, Good, Fair, Poor and Very Poor.

| Location | CAQI | AQ Description and DC in Linguistic terms | | | |
|---|---|---|---|---|---|
| | | Expert 1 | Expert 2 | Expert 3 | Expert 4 |
| PCMC | 151.13 (VP) | Very Poor (1.00-VG) | Very Poor (0.92-VG) | Poor (0.92-VG) | Poor (0.88-G) |
| Bhosari | 100.38 (P) | Poor (0.53-F) | Poor (0.30-P) | Fair (0.30-P) | Fair (0.42-F) |

**Table 2: CAQI and Fuzzy description of Air Quality with DC in PCMC location by 4 AQ Experts'**

From Table 2 it can be seen that the air quality computed using degree of certainty method is at par with the air quality description using the Conventional Air Quality Index (CAQI). We can say that air quality at *PCMC location is Very Poor with Very Good degree of certainty* .Variability in experts perception is also modelled using fuzzy logic. The four air quality experts are almost close in their opinion about describing air quality at Pimpri-Chinchwad monitoring location.

| Linguistic Description | Degree of certainty (DC) |
|---|---|
| Very Good | 0 |
| Good | 0.2 (Fair) |
| Fair | 0.75 [Good] |
| Poor | 0.1 |
| Water Quality at PCMC is Fair with DC = 0.75, Good with DC = 0.2 & Poor with DC = 0.1 | |

**Table 3: Fuzzy Description Of Water Quality With Degree Of Certainty**

- It could be inferred from fuzzy modelling that the water quality at Intake well of Water Treatment Plant is *fair* with degree of certainty 0.75 ( Good) and is tending towards *good* with degree of certainty 0.2 ( fair). It is necessary to improve Pawana river water quality as *Very good with high degree of certainty*. The strength of fuzzy logic concepts lies in defining, in this case, river water quality at intake, straightway in linguistic terms with some degree of certainty attached to each linguistic terms [6].
- There is visible organic pollution due to domestic waste water from the nearby villages. A few industrial discharges are also located in about 2 kms of the intake well. Therefore, there will be increase in Bio -chemical oxygen demand resulting in high bacteriological load in Pawana river water.
- In case the pollution remains unchecked, the possibility of introducing pre chlorination at the intake resulting into possible trihalomethane (THM) formation can not be overruled. PCMC, therefore, must to undertake pollution abatement measures on a war footing. The team of experts will look into the issue and locate the sources of pollution for further action.

## 5. 0 CONCLUDING REMARKS

The two valued logic and probability theory is the basis of prospect theory. It is a deep-seated tradition in science to employ the conceptual structure of bivalent logic and probability theory as a basis for formulation of definitions and concepts.

What is widely unrecognized is that, in reality, most concepts are fuzzy rather than bivalent, and, in general, it is not possible to formulate a co-intensive definition of a fuzzy concept within the conceptual structure of bivalent logic and probability theory. Fuzzy logic via computing with words is a human centric logic which has been successfully used where the perceptions of the human or domain experts are of primary concern. This is more relevant in decision making wherein the decision makers have no or less numerical information on the governing parameters. It is well known that the expert knowledge base uses only linguistic terms. Why not compute with words and why always numbers?

Over the past few decades, soft computing tools such as fuzzy-logic-based methods, neural networks, and genetic algorithms have had significant and growing impacts. But we have seen only limited use of these methods in environmental fields, such as risk assessment, cost–benefit analysis, and life-cycle impact assessment. Because fuzzy methods offer both new opportunities and unforeseen problems relative to current methods, it is difficult to determine how much impact such methods will have on environmental policies in the coming decades.

The research study has conclusively demonstrated the utility soft computing techniques with focus on fuzzy logic in water and air quality management. It could be an eye opener for those who are engaged in improving urban water and air quality

For the types of complex and imprecise problems that arise in environmental policy, the ability to model complex behaviors as a collection of simple if–then rules makes fuzzy logic via computing with words an appropriate modeling tool. Would decision makers and the public accept expressions of water or air quality goals in linguistic terms with computed degrees of certainty? Resistance is likely. In many regions, such as the

United States and European Union (EU), both decision makers and members of the public seem more comfortable with the current system—in which government agencies avoid confronting uncertainties by setting guidelines that are crisp and often fail to communicate uncertainty. Perhaps someday a more comprehensive approach that includes exposure surveys, toxicological data, and epidemiological studies coupled with fuzzy modeling will go a long way toward resolving some of the conflict, divisiveness, and controversy in the current regulatory paradigm.

**ACKNOWLEDGMENT**
Assistance from Asheesh Sharma and Parvin Ladkat in providing necessary data is gratefully acknowledged.

**REFERENCES**

[1]. WHO World Health Organization Health Report, 2002, centre for disease control, Geneva, Switzerland.
[2]. Ersoy, K., Cho, S., 1992. An algorithm to compute the degree of match in fuzzy systems. Journal of Fuzzy Sets and Systems. Vol. 49(1992), 285- 299.
[3]. Deshpande, A. W., Raje, D.V., Khanna, P., 1996. Fuzzy Description of River Water Quality. International Conference-EUFIT Germany.
[4]. Sharma, A., Yadav, J., Madlik, P., Ladkat, P., Deshpande, A., 2012.
[5]. Fuzzy logic in Water Supply System Management – A Case Study. Paper presented in NAFIPS, Berkeley, 2012.
[6]. Zadeh, L.A.,1973. The concept of a linguistic variable and its application to approximate reasoning. ERL-M411, Berkeley, USA.
[7]. Mckone, T. E., Deshpande, A.W., 2010. Can Fuzzy Logic Bring Complex Environmental Problems into Focus? IEEE. 11-17.
[8]. Yadav, J. Y., Kharat V., Deshpande, A. W., 2013. Evidence Theory and Fuzzy Relational Calculus in Estimation of Health Effects due to Air Pollution. International Journal of Intelligent Systems. Vol. 22(1), 9-22.