

Fuzzy logic in air pollution: Revisited

A Deshpande^{1,2}

¹Berkeley Initiative in Soft Computing (BISC) - University of California Berkeley CA

²Indian Institute of Technology, Mumbai India

Abstract

In everyday life and field, people mostly deal with concepts that involve factors that defy classification into crisp sets. The decisions people usually make are perceptions without rigorous analysis of numeric data. Like in other field of studies, there may exist imprecision in air quality parametric data collected and in the perception made by air quality experts in defining these parameters in linguistic terms such as: very good, good, poor. This is the reason why 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 to deal with aleatory as well as epistemic uncertainty in air quality related issues. This paper has highlighted mathematical preliminaries of air pollution studies like Similarity Measures (Cosine Amplitude Method), Fuzzy to Crisp Conversion (Alpha cut method), Fuzzy c Mean Clustering, Zadeh-Deshpande (ZD) Approach and linguistic description of air quality. Similarly, the applications of fuzzy similarity measures and fuzzy c mean clustering with defined possibility (- cut) levels in case air pollution studies for Delhi, India have been reflected. Though the approach of using fuzzy logic in pollution studies are not of common practice, the comprehensive approach that involves air pollution exposure surveys, toxicological data, and epidemiological studies coupled with fuzzy modeling will go a long way toward resolving some of the divisiveness and controversy in the current regulatory paradigm.

Key words: air pollution, alpha cut, fuzzy logic, fuzzy c mean clustering, fuzzy modeling

Introduction

In everyday life and field such as environmental health people deal with concepts that involve factors that defy classification into crisp sets: *safe, harmful, acceptable with mitigation measures*, and so on. Example: A regulator, carefully explaining the result of a detailed quantitative risk assessment report to a community group, only to be asked over and over again. *But are we safe?* In this case, *safe* defies crisp classification because it is a multivariate state with gradation that varies among different individuals and groups.

Human cognition plays an important role in the decision making process, which is invariably based on his/her experience and environment management systems are no exception. The majority of the decisions in the real world are made by perceptions without rigorous analysis of numeric data. For example, a physician after examining patients over a period of time, gives opinion in linguistic terms about the incidence of respiratory disease(s) due to possible air

pollution, without referring to any health records. Air quality, water quality and alike are fuzzy terms.

There exist imprecision in air quality parametric data collected and in the perception of air quality experts in defining these parameters in linguistic terms such as: *very good, good, poor* in relation to their air pollution potential. Fuzzy set theory could be one of the right kind of mechanism to deal with aleatory (imprecision in field data and in measuring instruments) as well as epistemic (perception based) uncertainty which are resident in air quality related issues.

There are several fuzzy logic based instruments developed and therefore, the arguments in favour and against of fuzzy logic has reached a steady state.

Air Pollution-A Man Made Disaster

Air pollution is the fifth leading cause of death, with 0.62 million per year premature deaths in India (2010) from

outdoor air pollution related diseases. Air quality assessment has assumed significance in the context of present times due to increased air pollution. The sequel address two issues with the results of a case study in air pollution wherein fuzzy logic has been successfully applied by the author with his team of researchers.

Issue 1

Sizing Optimum number of Air Quality Stations

How many air quality monitoring stations should be installed in a city is being researched for many years (Mazi et al., 2014; Ross, 2009). The issues in air quality monitorin network (AQMN) could be broadly classified into following three categories:

1. Designing AQMN for a new developing city where no air quality parametric data is available is not a trivial task. In this category, network design is based on projected pollution load for the design period of, say, 30 year and micro-metrological parameters. Various mathematical formulations are suggested by the researchers in designing AQMN. The authors believe that these efforts are at best a brilliant academic exercise with less / no practical utility.
2. With an objective to assess air pollution status in a city, the regulatory agency (say, Pollution Control Boards) had already installed a few (may be 2-5) AQMS and collected one or two year criteria pollutant data. The agency would need to know the adequacy of these stations, and suggest a few more if needed, which will represent the overall air pollution status of the city. The authors of the present paper discourages such an ad-hoc approach.
3. The regulatory agency installed air quality monitoring stations in mega cities at defined places – may be on the basis of population density, proposed pollution loads due to mobile transportation and relevant micro meteorological parameters. The authorities would like to classify these AQMS based on their pollution data, with a view to reduce the number of stations without comprising on the objective of presenting overall air pollution status of the city.

Mathematical Preliminaries

Similarity Measures (Cosine Amplitude Method):

There are different ways to develop the numerical values that characterize a relation, but similarity measures in one of the most prevalent forms of determining the values in a relation. Let data samples of a set for m a data array, say X and that can be expressed in the following equation for n data:

$$X = \{x_1, x_2, \dots, x_n\} \dots \dots \dots (1)$$

Each of the elements , in the data array X is itself a vector of length m, i.e.;

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{im}\} \dots \dots \dots (2)$$

Each element of a relation, r_{ij} , results from a pair wise comparison of two data samples, say x_i and x_j , where the strength of the relationship between data sample x_i and data sample x_j is given by the membership value expressing that strength, that is, $r_{ij} = \mu_R(x_i, y_j)$. Therefore, the cosine amplitude method calculates r_{ij} by the following equation based on the comparison of two data arrays.

$$r_{ij} = \frac{\left| \sum_{k=1}^m x_{ik}x_{jk} \right|}{\sqrt{\left(\sum_{k=1}^m x_{ik}^2 \right) \left(\sum_{k=1}^m x_{jk}^2 \right)}} \text{ where } i, j = 1, 2, \dots, n. (3)$$

This method calculates pair-wise relational strength (r_{ij}) based on the comparison of two data arrays and range of r_{ij} values that varies from 0 to 1 ($0 \leq r_{ij} \leq 1$).

A fuzzy equivalence relation must satisfy y all three matrix conditions viz., reflexivity, symmetry and transitivity. Let's R, a similarity relation and x, y be elements of a set X and $\mu_R(x, y)$ denote the grade of membership of the ordered pair (x, y) in R. Then R is a similarity relation in X if and only if, for all x, y and z in X, $\mu_R(x, x) = 1$ for all x in X (reflexivity), $\mu_R(x, y) = \mu_R(y, x)$ for all x and y in X (symmetry), and $\mu_R(x, z) \geq \min\{\mu_R(x, y), \mu_R(y, z)\}$ for all x, y and z in X (transitivity) (Zadeh, 1971; Ross, 2004; Zimmer mann, 2001). If fuzzy relation matrix only have the properties reflexivity and symmetry then it is called fuzzy tolerance relation matrix. Before defuzzification (fuzzy to crisp conversion), fuzzy tolerance relation has to be converted to fuzzy equivalence relation by composition. The computation using expression 1 will result into a matrix of size size $n \times n$ which will invariably a fuzzy tolerance relation matrix in similarity relations. Fuzzy tolerance relation is transfor med into fuzzy equivalence relation using different composition methods.

Any fuzzy tolerance relation matrix, R_1 can be reformed into a fuzzy equivalence relation by at most (n-1) compositions.

That is $R_1^{n-1} = R_1 \circ R_1 \circ \dots \circ R_1 = R \dots \dots \dots (4)$

Max-min composition: Considering $R_1(x,y), (x,y) \in X \times Y$ and $R_2(y,z), (y,z) \in Y \times Z$ be two fuzzy tolerance relations. The max-min composition $R_1 \max\text{-min } R_2$ is then the fuzzy set.

$$R_1 \circ R_2 = \{[(x,z), \max_y \{ \min \{ \mu_{R_1}(x,y), \mu_{R_2}(y,z) \} \}] | x \in X, y \in Y, z \in Z \} \dots \dots \dots (5)$$

$\mu_{R_1 \circ R_2}$ is the membership function of a fuzzy relation on fuzzy sets. Fuzzy to Crisp Conversion (Alpha cut method): Begin by considering a fuzzy set R, then define a lambda-cut set R_α , where $0 \leq \alpha \leq 1$. The set R_α is a crisp set called

the alpha (α)-cut (lambda (λ)-cut) set of the fuzzy set A , where $R_\alpha = \{x | \mu_R(x) \geq \alpha\}$. The α -cut set R_α is a crisp set derived from its parent fuzzy set, any particular fuzzy set R can be transformed into an infinite number of α -cut sets, because there are an infinite number of values α on the interval [0, 1]. Any element x in R_α belongs to fuzzy set R with a grade of membership that is greater than or equal to α .

A Brief account of Fuzzy c Mean Clustering

In fuzzy c-mean clustering, a frequently used pattern recognition method in which each sample point in a set of data belong to more than one cluster with membership value between zero (completely unlike) and one (completely like) and the sum of all membership values of each sample point must be one (Bezdek,1981). The number of cluster, both in hard and fuzzy models, may either be predetermined on empirical ground, or models for different number of clusters may be compared on the basis of suitable validity function.

In the absence of universally accepted approach for determining number of clusters in fuzzy c mean, the authors have adopted another route using the concept of fuzzy similarity measure especially for deciding air quality monitoring stations.

The algorithms used in fuzzy c-mean cluster analysis for non-empty data set, fuzzy c-partition matrix U, for grouping an assembling of data sets into classes, is grounded on minimization of objective function J_m

$$J_m (U,V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2 \dots\dots\dots(6)$$

$$d_{ik} = d(X_k-V_i) = \left[\sum_{j=1}^m (X_{kj}-V_{ij})^2 \right]^{1/2} \dots\dots\dots(7)$$

weighting parameter, m' is any real number ($1^2m'^{20}$) which control the amount of fuzziness in the classification process and μ_{ik} is membership of k^{th} data point in i^{th} class. The Euclidean distance d_{ik} is between i^{th} cluster center and k^{th} data set (data point in m-space). For i^{th} cluster center V_i , have m features, put in vector form as $V_i = \{V_{i1}, V_{i2}, \dots, V_{im}\}$. Each of cluster coordinate for each class calculated by following equation-

$$V_{ij} = \frac{\sum_{k=1}^n \mu_{ik}^m \cdot x_{kj}}{\sum_{k=1}^n \mu_{ik}^m} \quad 1 \leq i \leq c; j=1,2, \dots, m. \dots\dots\dots(8)$$

FCM is accomplished through an iterative optimization of the objective function, with the update of membership μ_{ik} and the cluster center v_{ij} .

The steps in FCM algorithm are as follows:

1. Set up c value ($2^2c < n$) and choose a value of parameter m'. Initial partition matrix, $U^{(0)}$ and each step in this algorithm will be marked as r, where $r = 0, 1, 2, \dots$
2. Compute the centers $\{v_i^{(r)}\}$ for each step.
3. Update partition membership matrix for r^{th} step, $U^{(r)}$ as equation

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^c \left(\frac{d_{jk}^{(r)}}{d_{ik}^{(r)}} \right)^{2/(m'-1)} \right]^{-1}; 1 \leq k \leq n; 1 \leq i \leq c \dots\dots\dots(9)$$

4. Compare $U^{(r+1)}$ to $U^{(r)}$ in any convenient matrix norm. If $\|U^{(r+1)} - U^r\|^2 \leq \epsilon_L$ (prescribed level of accuracy, say, 0.05), stop; otherwise set $r = r + 1$ and return to step 2. This operation meets to a local minimum or a saddle point of J_m .

Application of Fuzzy Operations in a Case Study: Delhi

Attempt are made to use fuzzy similarity measures and fuzzy c mean clustering to address the issue mentioned in category 3. In the first stage, based on 2 or 3 year criteria pollutant's data, Cosine Amplitude- one of the fuzzy similarity measures is used to classify or for grouping monitoring stations for the most critical air pollutant PM₁₀ and NO₂ for the defined possibility (α - cut) levels. While In the second stage, average values of PM₁₀, NO₂, SO₂, CO and SPM are used in fuzzy clustering for the winter months- a worst case scenario. The weighting factor m' value used in fuzzy clustering algorithm is decided based on the concept of reference group. Finally the optimal number of air quality monitoring locations is selected using combination of the single group for med in fuzzy similarity measure for PM₁₀ and NO₂, and having highest membership value in fuzzy clustering. The case study relates to Delhi Metro City in India. The air quality data of criteria pollutants was collected for two years from the installed 41 stations . The final outcome of the study reveals that the city of Delhi needs only 16/ 41 monitoring stations which will result into sizeable reduction in the capital cost and the recurring expenses in ambient air quality monitoring in Delhi (Mazi et al., 2014).

Issue 2

Fuzzy Description of Air Quality

The approach proposed by the author has been labled as Zadeh –Deshpande (ZD) Approach/ This is because of the fact the final outcome of the formalism is somewhat llike Z number or Zadeh number.

This approach directly describes air quality in linguistic tems with linguistic degree of certainty attached to each description. The identification of air quality parameters, number of sampling stations, time and frequency of observations etc., are crucial and are invariably based on the experts knowledgebase. Collection of the pollution

parametric data from sampling locations is the first step. On the basis of probability distribution fitting, mean and variance values could be used in further analysis. Many a times, collected data from the sampling locations is inadequate for carrying out statistical analysis (in fact, may not be needed). In such a situation, the defined statistical technique- a bootstrap method could be employed which ensures reliability of the data. Based on the data obtained using bootstrap method, probability distribution fitting to air quality parametric data is carried out. As we intend to compare probability distribution with the fuzzy sets drawn for the selected parameters, it is necessary to transform probability distribution into possibility distribution of the parametric data, using the concept of Convex Normalized Fuzzy Number (CNFN). Probability distribution can be transformed into a CNFN with membership grade function $\mu(x)$ thereby characterizing the dynamic behavior of the pollutants into a possibility distribution. Fuzzy sets are generated using GA for selected parameters. A computational scheme of Degree of Match (DM) can be used with a view to estimate matching between fuzzy sets and the antecedent part of the rule, in order to describe air quality fuzzily with certain degree of certainty. 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. The degree of match of each classification rule indicates the certainty value of classification. The greater the degree of match, the greater is the possibility that air quality is classified in that class. 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.

Reliability of monthly mother sample data of the pollutants and the identification of the worst winter month has been ensured using bootstrapping. 1000 bootstrap samples were compared for mean and standard error of the original and the bootstrap samples. In the present study, bootstrap mean for the month of November 2013 for all the pollutants monitored in Mumbai and New York is considered for further analysis. Convex Normalized Fuzzy Numbers (CNFN) was constructed for the data sets of the parameters. Linguistic description of air quality was obtained from the experts.

The case study relates to comparative evaluation of air quality of air quality monitoring stations in New York City and Mumbai. Table 1 presents the final outcome of the study.

Table 1 Linguistic description of Air Quality and Degree of Certainty

AQ description with expert Fuzzy sets (perception based)						
VG	1 DC-VH	1 DC-VH	1 DC-VH	1 DC-VH	0	0
G	0	0	0	0	0	0
F	0	0	0	0	0.23	0
P	0	0	0	0	0.46	0.82
VP	0	0	0	0	VH	VH

Table 1 describes air quality linguistically with linguistic degree of certainty using fuzzy sets based on the perceptions of the domain experts'. Using expert fuzzy sets the air quality at New York could be linguistically described as *Very Good* with degree of certainty as *Very High*.

It can be concluded that the air quality in the defined locations Mumbai is generally described as *Very Poor*. It can be concluded that the AQ at all the monitoring stations in New York is *Very Good* with degree of certainty as *Very High*. The air quality at Sion city in Mumbai can be directly described linguistically as *Very Poor* with a *Very High* (0.95) degree of certainty value. The next higher value of 0.78 for *Poor* indicates that the air quality at Sion is definitely deteriorating as the DC is 0.78 which is higher than 0.46 for the linguistic hedge *Fair*.

New York City has initiated strict pollution control norms. New York City enacted a law in 2004 requiring CO alarms in residential and many public buildings bringing CO with permissible limits. New York City's air quality has improved over time as regulations have made Federal, State, and local air quality standards more stringent over the last two decades. Federal and State regulatory efforts to reduce emissions from the transportation, off-road, and stationary source sectors have driven continued national improvements in air quality.

Outlook in Environmental Policy

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. Here, we consider some obvious advantages and limitations.

Quantitative models with explicit and crisp delineations of systems have long been the currency of discourse in engineering and the physical sciences, where basic physical laws form the foundations of analyses. These fields place high value on the causal linkages implicit in model structure and parameterization. But for problems that involve human perception, language, control theory, biology, and even environmental systems, researchers have had to rely more on descriptive and empirical approaches. When the goal is to summarize the observations in an efficient and useful manner, fuzzy-logic-based methods should be further investigated as alternative—and perhaps more appropriate—methods for addressing uncertain and complex systems.

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 an appropriate modeling tool. Because fuzzy arithmetic works well for addressing linguistic variables and poorly characterized parameters, fuzzy methods offer the opportunity to evaluate and communicate assessments on the basis of linguistic terms that could possibly match those of decision makers and the public. Moreover, approximate reasoning methods such as fuzzy arithmetic do not require well-characterized statistical distributions as inputs. Another key advantage of fuzzy logic in risk assessment is the ability to merge multiple objectives with different values and meanings, for example, combining health objectives with aesthetic objectives. It also provides rules for combining qualitative and quantitative objectives.

However, fuzzy logic has at least two limitations for expressing health risks and other environmental impacts. One problem is its strong reliance on subjective inputs. Although this is a problem in any type of assessment, fuzzy methods might provide more opportunity for the misuse of subjective inputs. Moreover, some argue that although fuzzy logic is well suited to address uncertainty, it has not been demonstrated to be superior to standard statistical descriptions for dealing with variability (heterogeneity). In considering health-based groundwater remediation, Ozbek and Pinder used statements and preferences of practicing toxicologists to construct fuzzy rules that relate a benzene exposure pattern to carcinogenic effects. Intake variability is typically represented by variability distributions in other risk assessments, but they used fuzzy logic methods to describe the variation of both intake and toxicological susceptibility among exposed individuals. More recently, Kentel and Aral combined fuzzy set theory with standard probabilistic methods into a procedure they call probabilistic–fuzzy risk assessment, which addresses both uncertainty and variability in their model of health risks from tap water contaminated by tetrachloroethylene.

Although probabilistic assessments based on tools such as Monte Carlo methods are analogous to assessments based on fuzzy logic, these two techniques differ significantly both in approach and in interpretation of results. Fuzzy logic confronts linguistic variables such as safe, hazardous, acceptable, and unacceptable, whereas Monte Carlo methods are forced to fit linguistic variables for probabilistic assessments. Fuzzy arithmetic combines outcomes from different sets in a way that is analogous to but still different from Monte Carlo methods. Possibility theory can be used as an alternative to probabilistic analysis, but this strategy creates the potential for misuse if membership functions are interpreted as probability distributions.

Conclusion

What do we mean by environmental quality in general and water quality for bathing in particular? There is no simple answer to this question as these are fuzzy terms. Then why should we calculate the index of a term which cannot be defined in more precise form? Why not directly describe water quality fuzzily with some degree of certainty? We believe that perception based modelling with focus on fuzzy logic could be one of the ways of addressing such problems.

Could we adapt to a system that relaxes crisp lines and sharp demarcations to fuzzy gradations? 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, 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.

Acknowledgement

The author is grateful to Mr K amal Mazi, Jyoti Yadav and Thomas McKone, D.V. Raje for their contribution Assistance of innumerable domain experts is gratefully acknowledged.

References

- Bezdek, J.C. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum, New York.
- Mazi, K.J., Dikshit, A.K., & Deshpande, A. (2014). Can Fuzzy Set Theory Bring Complex Issues in Sizing Air Quality Monitoring Network into focus? *International Journal of System Assurance Engineering and Management*. DOI: 10.1007/s13198-014-0327-1
- McKone, T., & Deshpande, A.W. (2005). Can Fuzzy logic Bring Complex Problems in to focus? *Environmental Science and Technology*, 39, 42A-45A.
- Ross, T.J. (2004). Fuzzy logic with engineering applications, 2nd edn. Wiley, Chichester
- Ross, T.J. (2009). Fuzzy Logic with Engineering Applications. Second Edition, Wiley-India, pp. 369-375.
- Zadeh, L.A. (1971). Similarity relations and fuzzy orderings. *Inf Sci* 3(2), 177–200.
- Zimmermann, H.J. (2001). Fuzzy set theory and its applications, 4th edn. Kluwer Academic Publishers, Boston, Dordrecht, London.