International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Modeling for Uncertainty Assessment in Human Health Risk Quantification: A Fuzzy Based approach

Abdullah Mofarrah[1] and Tahir Husain[2]

[1]Graguate student, Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, A1B 3X5 Canada abdullah.mofarrah@gmail.com (corresponding author)
[2]Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, A1B 3X5 Canada

Abstract: Health risk analysis to the contaminated water involves the use of mechanistic models that include many uncertain and variable parameters. Currently, the uncertainties of risk analysis models are treated using statistical theories considering the randomness in data distribution. However, not all uncertainties in data or model parameters are due to randomness. Other sources of imprecision that may lead to uncertainty include scarce or incomplete data, measurement error or subjective interpretation of available information. These kinds of uncertainties cannot be treated solely by statistical methods. This paper uses fuzzy set theory (FST) together with probability theory (PT) to incorporate uncertainties into risk analysis model. Based on the form of available information, FST, PT, or a combination of both is used to incorporate parameter uncertainty and variability into risk assessment models. The pollutants concentration, cancer and non-cancer risk potency factors are highly uncertain parameters in risk analysis model and treated as fuzzy variables while the remaining model parameters are treated as random or constant function. Triangular fuzzy function (TFN) is integrated with random variables at different alpha-cut levels to produced cumulative distribution function (CDF) of individual's risk. The methodology is explained through a case study related to the human health risk posed by produced water discharge from petroleum industries.

Keywords Health risk analysis, fuzzy set, probabilistic risk assessment, triangular membership functions.

1. INTRODUCTION

Risk assessment model is used to quantify the human health impacts due to exposure to contaminants via multiple exposure routes such as ingestion, and dermal contact. The goal of risk assessment is to estimate the severity and likelihood of harm to human health from exposure to a substance or activity that under plausible circumstances can cause harm to human health. The quantitative risk characterization involves exposure dose estimates against a benchmark of toxicity, such as a cancer slope factor (SF), reference dose (RfD). Uncertainties in risk estimates may arise from different sources such as measurement or estimation of parameters, natural variability in individual's response, variability in environmental concentration of toxicants over time and space and unverifiable assumptions in dose-response models or extrapolations of the results of these models. In order to take into account the major uncertainties in risk quantification, in recent years, the probabilistic risk assessment (PRA) studies have become popular. PRA is the general term for risk assessment that uses probability theory to represent the likelihood of different risk levels in a population (Maxwell et. al. 1998; Maxwell and Kastenberg 1999; Ma et. al. 2002). In PRA physical parameters and input loads are considered as random variables. The output of a PRA is a probability distribution of risk that reflects the combination of the input probability distributions. Using this approach, if the input distributions represent variability in a probabilistic sense, then the output risk distribution may provide variable information. If the input distributions reflect uncertainty in a probabilistic sense, then the output risk distribution may provide uncertainty in the risk estimate (U.S. EPA 2001). However, resalable and sufficient data is the prerequisite to estimate and characterize the probability distribution of the input variables. Moreover, if uncertainty is not due to randomness or if the available information is in the form of an expert judgment or subjective interpretations, than probabilistic analysis may not be sufficient to represent the true nature of uncertainty. In such cases, fuzzy set theory Zadeh (1965) can be used to incorporate uncertainty associated in the computational models. In this paper, the fuzzy set theory is integrated with mathematical modeling in order to assess uncertainties in human health risk. The proposed model permits to use other types of information such as, expert knowledge or fuzzy information. This approach integrates probability theory and fuzzy arithmetic in treating variability and uncertainty. The demonstration of this methodology is explained by a hypothetical case in human risk estimation through food chain.

2. BACKGROUND AND SOURCES OF UNCERTAINTY

Produced water (PW) is the most significant source of waste generated in the production phase of oil and gas operations. Once discharged into the ocean, a number of heavy metals and poly aromatic hydrocarbon (PAHs) in PW may introduce toxicity and bioaccumulation in marine organisms (Neff, 2002). These compounds are therefore harmful to fish and humans. There are numbers of models have been using to predict the dilution as well as the predicted environmental concentration (PEC) concentration, but none of those models can not predict PEC accurately. Outfall dilution depends on the several factors including the discharge rate, current speed, wind, temperature, salinity range of receiving water etc. According to the Rye et al. (1996), the initial dilution can be considered 1:1000 at a distance of 500 meters. Responses of populations exposed to a given dose of contaminant, risk analysts conduct mathematical extrapolations. Clear-cut relations are rare since epidemiological studies are not very sensitive in detecting health effects from relatively low levels of exposure (Hattis and Kennedy 1986). Most of the cases the epidemiological evidence is incomplete or ambiguous, mathematical models are considered to predict human doses using animal studies data that leads serious uncertainties. To overcome these uncertainties fuzzy membership function may be used in selecting the toxicological response (i.e., slope factors) of the individuals.

3. HEALTH RISK ASSESSMENT MODELS

Human health risk assessment involves evaluating the effect of toxins, contaminants and other environmental hazards on human health. Among the other model USEPA (2001) developed a comprehensive food chain risk assessment model as:

$$CDI = \frac{C_f \times FIR \times FR \times EF \times ED \times CF}{BW \times AT} \qquad (1)$$

where, CID = Chronic daily intake (mg/kg-day), FIR = fish ingestion rate (g/day); according to USEPA (1996), FIR = 170 g/day); FR = fraction of fish from contaminated source (a value of 0.50 (50%) may be used); EF = exposure frequency (days/year; according to USEPA (1991), EF = 350 days) BW = average human bodyweight over the exposure period (kg) according to USEPA (1991), BW = 70 kg); ED = exposure duration (years); CF = 10^{-9} conversion factor for fish tissue concentration and fish ingestion; AT = averaging time in days; (the life expectancy is assumed 70 x 365 days for carcinogen risk and 30 x 365 days for non-carcinogen risk); and C_f = chemical concentration in fish tissue (mg/kg). The chemical concentration in fish tissue (C_f) can be computed as (Marie et. al., 1994).

$$C_f = PEC \times BCF \qquad (2)$$

Where PEC = predicted environmental concentration (mg/l) and BCF is the chemical bioconcentration factor in fish (l/kg) .

This study proposed a systematic health risk assessment approach which allows variables use in the risk equations as crisp, random, or fuzzy variables depending on the available information. Assuming same public response to all the pollutants, additive mode of action is applied and combined risk model for fish ingestion is expressed in fuzzy form as:

$$Risk_{cancer} = \sum \underbrace{C_f}_{fuzzy} \times \underbrace{\left[f(\beta^1,....,\beta^n)\right]}_{random} \times \underbrace{CSF_k}_{fuzzy} \tag{3}$$

$$Risk_{non-cancer} = \sum \underbrace{C_f}_{fuzzy} \times \underbrace{\left[f(\beta^1,....,\beta^n)\right]}_{random} / \underbrace{Rfd_k}_{fuzzy} \tag{4}$$

Where $f(\beta_{ing}^1,......,\beta_{ing}^n)$ represents the function involving random variables occurring in the Equation 1, n is the number of random variables and k represents the pollutants. The parameters in Equation 4 could be crisp, random, or fuzzy variables depending on the available information.

4. ANALYSIS TECHNIQUES

The Monte Carlo Simulation (MCS) was used to propagate information supplied by probability density functions of the random variables. On the other hand, fuzzy arithmetic and interval analysis was used to integrate uncertainty associated with the fuzzy variables.

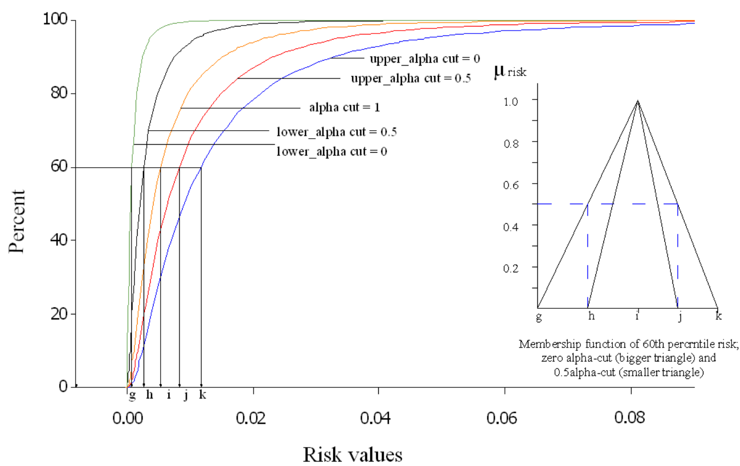

Figure 1. Illustration of risk membership function development

The risk Equation 3 and 4 is monotonic functions which allow interval analysis to carry out fuzzy calculations. Interval analysis involves converting the membership domain of the fuzzy variable into a specified number; alpha-cut (α_c) technique was applied. The lower and upper bounds of a TFN can be found as (Jie et al., 2006):

$$\alpha^{lower} = \left[\alpha * (m - l) + l\right] \tag{5}$$

$$\alpha^{upper} = n - \left[\alpha * (n - m)\right] \tag{6}$$

where the parameters l, m, and n, respectively, indicate the smallest possible value (left), the most promising value (middle), and the largest possible value (right) of a TFN. Considering each level of alpha cut, joint cumulative distribution function (CDF) is generated for the associated risk by applying MCS. For decision making purpose, the CDF of risk $(Risk_{cdf})$ is generated for different alpha-cut level. The membership function of the associated risk to a specific percentile was generated by drawing a horizontal line cutting through $(Risk_{cdf})$ curves and risk values corresponding to $(Risk_{cdf})$ can be read. Finally, risk membership functions were developed by combining all risk values. The concept of risk membership function is explained in Figure 1.

5. CASE STUDY

To demonstrate the methodology a hypothetical case study for evaluation of non-cancer human risk is represented here. The parameters in Equations 1 were considered constant, random variable and fuzzy numbers. According to USEPA (1996) average fish ingestion rate (FIR) by the people is 170 g/day, but this number could be varied depending on the fish consumption habit, allowing for uncertainty, this parameter was considered in probabilistic mode with mean 170 and std. 50. The fraction of fish from contaminated source was assumed a constant of 0.50 (50%). The exposure frequency, exposure duration, average human bodyweight and the life expectancy were assumed according to USEPA (1991) shown in Table 1.

Table 1. parameters used in the risk calculations

Parameter	Units	Type of variable	Value/Distribution*
Average time (AT)	days	Constant	25550 (70 years)
Body weight (BW)	kg	Constant	70
Exposure duration (ED)	years	Constant	30
Exposure frequency (EF)	Day/year	Constant	350
Fraction of contaminated fish (FR)	-	Constant	0.5
Fish ingestion rate (FIR)	g/day	random	normal ~ (170, 50)
PEC for As	ug/l	Fuzzy	Triangular (1.5, 4.7, 9.0)[a]
PEC for Cd	ug/l	Fuzzy	Triangular (20, 6670, 10000)[a]
PEC for Cu	ug/l	Fuzzy	Triangular (2, 128.8, 600)[a]
PEC for Pb	ug/l	Fuzzy	Triangular (50, 112.5, 270)[a]
BCF for As	l/kg	Fuzzy	Triangular (30, 44, 60)[b]
BCF for Cd	l/kg	Fuzzy	Triangular (70, 81, 90)[b]
BCF for Cu	l/kg	Fuzzy	Triangular (150, 200, 250)[b]
BCF for Pb	l/kg	Fuzzy	Triangular (30, 49, 70)[b]
Oral Rfd for As	mg/(Kg.day)	random	Triangular (3.0E-05,3.0E-04, 3.0E-03)[c]
Oral Rfd for Cd	mg/(Kg.day)	random	Triangular (1.0E-04,1.0E-03, 1.0E-02)[c]
Oral Rfd for Cu	mg/(Kg.day)	random	Triangular (4.0E-01,5.0E-01, 6.0E-01)[c]
Oral Rfd for Pb	mg/(Kg.day)	random	Triangular (1.0E-04,1.0E-03, 1.0E-02)[c]

[a] *Data compiled from Roe et al. (1996), Stephenson (1992) and Neff (2002)*
[b] *Data middle values compiled from Marie et. al., 1994*
[c] *Data middle values compiled WHO (1987), IRIS (1995)*

There are number of organic and inorganic pollutants present in PW but, this paper considered only four heavy metals namely, arsenic (As), cadmium (Cd), copper (Cu) and lead (pb) because of its toxicity and high concentration in PW. The maximum, average and minimum pollutants concentration data were collected from literature (Table 1), a dilution factors 1000 folds were used to calculate PEC. The calculated PECs were represented by TFN. The chemical bioconcentration factor (BCF) in fish is highly uncertain depends on many others parameters; form the available literature review this factor was also considered TFN. Table 1 is shown all the parameters used in risk calculations along with the type of variability. The membership functions of the fuzzy variables are shown in Figure 2. Monte Carlo simulations (MCS) with 1000 iterations were used to generate of cumulative distribution function (CDF) of non-cancer risk. Total non-cancer risk was calculated by adding individual non-carcinogenic risk for each chemical. Non-carcinogenic risk is expressed in terms of a hazard quotient (HQ). This is simply the ratio of the estimated chronic daily intake to the reference dose (RfD).

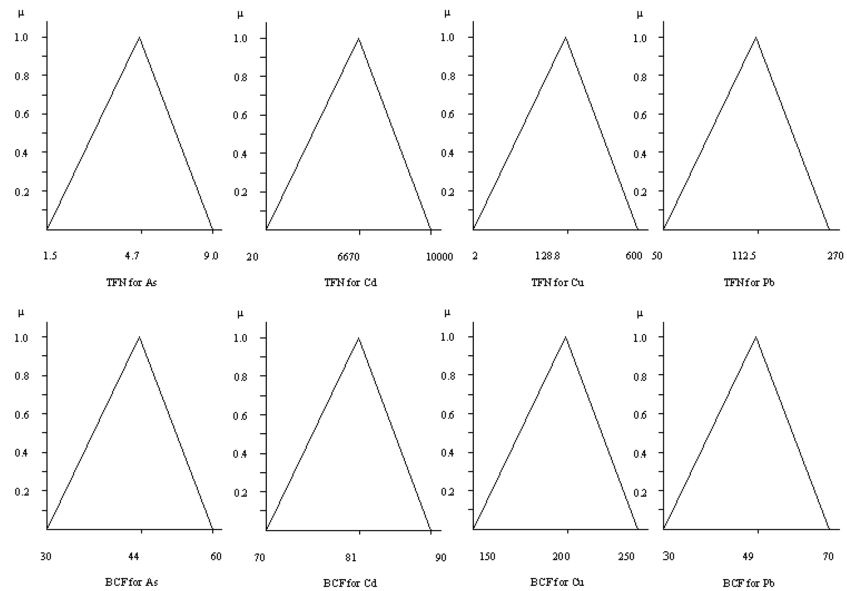

Figure 2. Membership function of contaminant water and fish BCF

The CDFs of risk, corresponding to lower and upper limits of different alpha-cuts (0, 0.4 and 1.0) were generated shown in Figure 3. Using these non-cancer risks CDF, the membership function of the total risk to a specific percentile was generated. Figure 4&5 is respectively shown the fuzzy membership function of 60 and 80 percentile risks for alpha-cut 0.4 along with zero alpha-cut. By analysing the Figure 4 & 5 it is clear that, the zero alpha –cut is giving more wide range of risk values (i.e., 0.001 to 0.28 for 60 percentile and 0.001 to 0.625 for 80th percentile) than alpha-cut equal to 0.4 (i.e., 0.11 to 0.22 for 60 percentile and 0.15 to 0.48 for 80th percentile).

As can be seen from Figure 4&5, different percentile of risk the membership functions have triangular distributions. Triangular membership functions of risk can be interpreted as risk to individuals at a certain percentile of risk being around the peak value (i.e., value corresponding to a membership function of 1.0). A membership value of one reflects the most likely value for the variable. The selection of input membership functions for the fuzzy variables and the probability distribution functions of the random variables will impact the shape of the membership function of risk obtained for a certain percentile. The shape of the membership function may also represent valuable information to the risk analyst. It is very difficult to make decision especially when there is a wide range of risk values (i.e., lower and upper limit). On the other hand fuzzy risk membership functions generated ranges of risk values, but at the same time it can be converted into crisp values whenever needed for decision making purpose. This is the main advantage of fuzzy theory over probability analysis. One of the most popular fuzzy defuzzification method is the center of gravity method (Yager, 1980). Alpha-cut plays an important role in decision making process for example decision with lower alpha-cut will provide wide range of uncertainty but better representative than the lower alpha-cut.

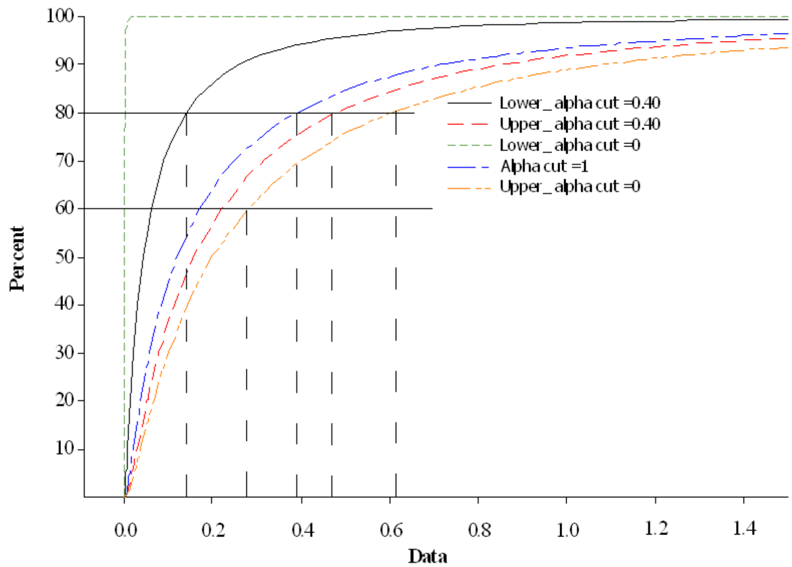

Figure 3. CDF of non-cancer risk

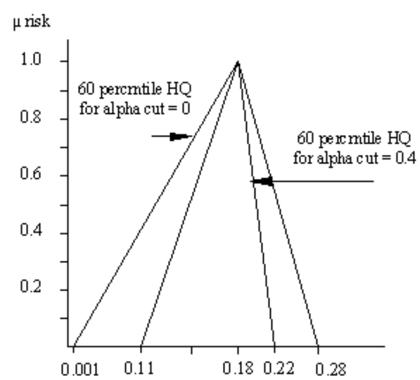

Figure 4. The membership function of non-cancer risk at 60th percentile

6. DISCUSSION

Fuzzy logic and probability theory are both powerful tools for handling uncertainty. In health risk assessment studies, it is very important to include all available information into the mathematical models. Traditionally, the available information is interpreted in a probabilistic sense and probability theory has been used to describe this information. Probability theory is a very strong and well established mathematical tool to treat variability. It has certain input requirements and whenever these requirements are met, probability theory will provide powerful results. However, it is clear that not all uncertainties in data or model parameters are random; other source of imprecision that may lead to uncertainty is scarce or incomplete data, measurement error or data obtained from expert judgment or subjective interpretation of available information. These kinds of uncertainties cannot be treated solely by probability theory. Thus, usefulness and applicability of other mathematical tools, such as fuzzy set theory or possibility theory should be explored.

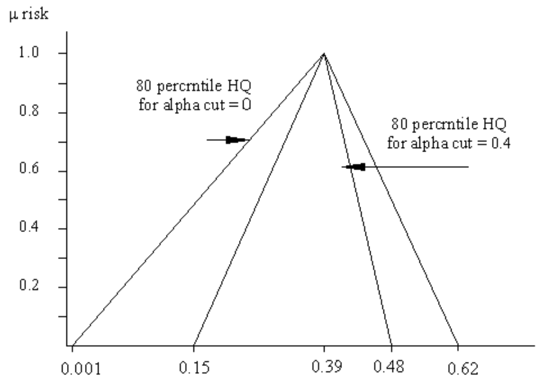

Figure 5. The membership function of non-cancer risk at 80th percentile

Representation of uncertainty using fuzzy or random variables will impact the form of the uncertainty in the calculated risk. The membership function of risk for a certain percentile may provide significant information for the decision maker. For example, the possibility of occurrence of risk values having zero membership values for a specific percentile are zero, while the risk value with a membership value of one is the most likely risk. The shape and the support base of the risk membership functions provide extra information about the resulting uncertainty which is a combined effect of the random and fuzzy input variables. For example, the uncertainty associated with a risk that has a small support base is respectively smaller than that of a risk which has a larger support base.

7. CONCLUSIONS

The proposed approach handles the uncertainties in health risk assessment using combination of probability and fuzzy set theory. Treating heavy metals BCF in fish, together with the contaminant concentration in PW as fuzzy variables allowed us to include uncertainties due to reasons other than randomness into the risk assessment model. Fuzzy set calculated resulted membership values of risk for individuals at a certain percentile. Instead of a single probability distribution of risk as provided by probabilistic risk assessment, proposed approach provides the probability distributions of risk for various alpha-cut levels.

For simplicity purposes triangular membership functions are used in this study; however, membership functions for fuzzy variables do not need to be triangular. Since the membership functions of the input parameters are chosen as triangular distributions the resulting fuzzy risks have triangular distributions.
If other membership functions are used for the input variables, the shape of resulting fuzzy risk will change. The shape of the membership function depends on the available information about the fuzzy variable. If the appropriate information is available, utilization of this approach will provide results which may help the decision maker to make more informed decisions.

ACKNOWLEDGE

Financial support provided by the Natural Science and Engineering Research Council of Canada (NSERC) is highly appreciated.

REFERENCES

Hattis, D., and Kennedy, D. (1986). "Assessing Risks from Health Hazards: An Imprefect Science." Technology Review, 89(4), 60-71.
IRIS (1995):Integrated risk information system. USEPA, oct. 1995.

Jie, H. L., Meng, M.C., and Cheong, C.W., 2006. Web Based Fuzzy Multicriteria Decision Making Tool. International Journal of the Computer, the Internet and Management Vol. 14.No.2, pp 1-14

Ma, H.-W. (2000). "The incorporation of stochasticity in risk analysis and management: a case study." Stochastic Environmental Research and Risk Assessment, 14, 195-206.

Marie, G, Rebecca, Z, Betsy, R, Elizabeth, H. and & Paul A., 1994. Review of mathematical models for health risk assessment: V. chemical concentrations in the food chain, Environmental Software 9, 115-131

Maxwell, R. M., and Kastenberg, W. E. (1999). "Stochastic environmental risk analysis: an integrated methodology for predicting cancer risk from contaminated groundwater." Stochastic Environmental Research and Risk Assessment, 13, 27-47.

Maxwell, R. M., Pelmulder, S. D., Tompson, A. F. B., and Kastenberg, W. E. (1998). "On the development of a new methodology for groundwater-driven health risk assessment." Water Resources Research, 34(4), 833-847.

Neff, J.M., (2002). Bioaccumulation in Marine Organisms: Effects if Contaminants in Oil Well produced water, J. M. Neff, Elsevier 2002, ISBN: 0-080-43716-8

Roe, T.I, Johnsen, S. (1996).Discharges of Produced water to the North Sea. Reed, M. and Johnsen S. (Editors), Produced Water 2: Environmental Issues and Mitigation Technologies, Plenum Press, New York, pp. 3-41.

Rye, H., Reed M., Melbye A., and Johnsen S. (1996). An Intercomparison between field measurements and three different calculation models for estimates of dilution factors. M. Reed and S. Johnsen (Edited), Produced Water 2; Environmental Issues and Mitigation Technologies, Plenum Press, New York, p 257-269.

Stephenson, M.T. (1992). A Survey of Produced Water Studies. in Produced Water, J.P. Ray and F.R. Englehart (eds.), Plenum Press, New York.

US EPA. 1989. Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual (Part A) Interim Final. Office of Emergency and Remedial Response, EPA/540/1-89/002. United States Environmental Protection Agency. December 1989.

US EPA. 2001. Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part E, Supplemental Guidance for Dermal Risk Assessment). Office of Emergency and Remedial Response, EPA/540/R/99/005, Interim, Review Draft. United States Environmental Protection Agency. September 2001.

WHO (1987): Air quality guidelines for Europe. WHO Regional Publications, European Series No. 23. WHO Regional Publications, European Series No. 23, WHO. Copenhagen.

Yager, R.R., 1980. On a general class of fuzzy connectives. Fuzzy Set Syst 4:235–242

Zadeh, L. A., 1965. Fuzzy sets. Information Control, 8(3), 338–353.