IJSAR, 3(1), 2016; 23-31

International Journal of Sciences & Applied Research

www.ijsar.in

A comparative case study for cancer risk assessment under probability - generalised fuzzy number

Rupjit Saikia*

Department of Mathematics, Dibrugarh University, Dibrugarh-786004, Assam, India.

Correspondence Address: *Rupjit Saikia, Department of Mathematics, Dibrugarh University, Dibrugarh-786004, Assam, India.

Abstract

An evaluation is required to determine the possible impact of hazardous substances on human health. Due to this reason, a comparative case study for cancer risk assessment under probability-generalised fuzzy number is presented here. The case study is presented here by using two principles, viz, one is by Chen's Functions principle and the other is by Dutta and Ali. Then we compare the results to find which is more effective.

Keywords: Aleatory uncertainty, epistemic uncertainty, generalised fuzzy number, probability theory, fuzzy set theory, Chen's function principle

Introduction

In the decision making system, risk assessment is an important and popular tool. As per the IAEA report [1] the major categories of sources of the uncertainty are (i) Input parameter uncertainty: The parameters of the various models are not exactly known because of scarcity or lack of data, variability with the populations of plants and/or components and assumption made by experts. (ii) Modelling uncertainty**:-** These uncertainties are introduced by the relative inadequacy of the conceptual models, the mathematical models, the numerical approximations, the coding errors and the computational limits[2].

In the input parameters of the model, uncertainties are characterized and propagated to quantify their impact on the model output. There are various methods for the treatment of uncertainties. These are : Method of Moments (Granger and Henrion [3]), Monte Carlo simulation (Jackson et al.[4]), Fuzzy arithmetic (Tanaka et al.[5]); Durga Rao et al. [6], Dempster-Shafer theory (Bae et al.[7]) and Probability bounds (p-box)(Fenson and Hajagos [8]).

Risk assessment is done by using model and a model is a function of parameters which are usually affected by aleatory uncertainty and epistemic uncertainty. When both aleatory uncertainty and epistemic uncertainty affect the model parameters, then hybrid method or combination of probability and fuzzy number is required.

When some model parameters are affected by aleatory uncertainty and other some parameters affected by epistemic uncertainty, one can either transform all the uncertainties to one type of format or need for joint propagation of uncertainties for computation of the risk. Different effort have been made by different researchers for joint propagation of aleatory and epistemic uncertainty, viz, Guyommet et al.[9], Baudrit et al.([10],[11]), Kentel and Aral [12], Li et al.[13], Dutta et al.[14]. In all their effort it is observed that, some input parameters are probabilistic and some are normal fuzzy number.

Recently, Dutta [15] made an analysis to combine probability distribution with normal fuzzy number and generalised interval valued fuzzy number within the same framework.

Environmental or human health risk assessment is an important tool in any decision making system in order to minimise the effects of human actives on the environment. Generally, environment data are vague and imprecise. Therefore, uncertainty is related to these data. Model parameters can be affected by epistemic uncertainty. To handle this type of uncertainty, fuzzy set theory or possibility theory can be used [14]. Guyonnet et al. [9] have proposed hybrid method for combinning probability and possibility distributions. The hybrid method proposed by Guyonnet et al.[9] joins the random sampling of probability distribution functions with fuzzy interval analysis on the α-cut and performed a post-processing of this result in order to combine random fuzzy set with a tolerance threshold. Baudrit et al.[10] work on the theory of Dempster-Shafer. Kental and Aral [12] combined utilisation of fuzzy and random variables. Li et al. [13] has proposed different hybrid method to join probability and possibility distribution. Dutta et al.[14] proposed a hybrid method to deal with both variability and uncertainty within the same framework of computation of risk.

In this paper, an effort has been made to combine probability distribution and generalised fuzzy number. Here we used two approaches to present the hypothetical case study for health risk assessment and finally we compare the results.

Basic idea of uncertainty

The ideas and concepts uncertainty begin in the 4th century B.C. From the Greek epistemic, meaning `knowledge' and logos, meaning `theory', the word epistemology is derived. Epistemology associates with the possibilities and limits of human knowledge. Uncertainty impacts decisions, designs and behavior in a wide variety of fields. Uncertainty in the physical sciences primarily concentrated on error analysis and quantum physics. Uncertainty influences a wide range of fields in the physical sciences and engineering.

Uncertainty refers to the imprecision in the estimate concerning a parameter value. Some uncertainty may be associated with variability, for e.g., the variation in consumption rates of a particular vegetable may be somewhere between a particular range of coefficient of variation.In such case the range expressing the uncertainty in coefficient of variation. The uncertainty associated with this particular value may be reduced in further measurements. There are mainly two types of uncertainties which are discussed below.

Aleatory uncertainty

If a system behaves in random ways, this type of uncertainty arises. Aleatory uncertainty is also called stochastic uncertainty, Type A uncertainty, variability, randomness and irreducible. It may be tied to variation in physical and biological processes and cannot be reduced with additional research or information although it may be known with greater certainty. This type of uncertainty relates with the data like demographic data on food intake, water intake which depends on the height, body weight, socio-economic status life style and inherent variations in dietary habits. Aleatory uncertainty cannot be reduced by additional study, improvement in questions, increase in sample size whatsoever. Only this type of uncertainty can be reported with higher level of confidence.

Epistemic uncertainty

Epistemic uncertainty evolves from the incompleteness of knowledge about the system. Epistemic uncertainty is also called subjective uncertainty, Type-B uncertainty, state of knowledge uncertainty. It arises due to faulty sampling design, sample size, sample collection methodology, measurements, processing of samples, detection limits of instruments used in study and data analysis, censoring, ignorance about the physical mechanisms and processes involved and other imperfection in scientific understanding. This type of uncertainty can be reduced by additional experiments and collections of additional bit of data.

Basic concept of probability theory

Probability theory is frequently used to study uncertainty analysis if the parameters used in prescribed models are random in nature and followed well defined distribution. Probabilistic methods quantify variation in exposure by using distribution to describe variation in consumption, concentration and body weight, then combining these two produce a distribution for exposure.

A random variable is a variable in a study in which subjects are randomly selected. It frequently occurs in performing an experiment that we are mainly interested in some functions of the outcome as opposed to the outcome itself. Formally, these are real valued functions defined on the sample space.

A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable X, we $p(a)$ define the probability mass function of X by

$$p(a) = P(X = a)$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a, i.e.,if x must assume one of the values $x_1, x_2, \cdots$,then

$$p(x_i) > 0 \quad i = 1, 2, \ldots$$

$p(x) = 0$ for other values of x.

Since X must take on one of the values of x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{all x_i \le a} p(x_i)$$

X is a continuous random variable if there exists a nonnegative function $f(x)$, defines for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P(X \in B) = \int_B f(x)$$

The function $f(x)$ is called the probability density function of the random variable X. Also

$$\int_{-\infty}^{\infty} f(x) = 1$$

The CDF of a continuous random variable X is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

Basic idea of fuzzy set theory

The publication of a paper by Lotfi A. Zadeh [16] was the important point in the evaluation of modern concept of uncertainty. In his paper, Zadeh introduced a theory whose objects are fuzzy sets. Fuzzy sets are the sets with boundaries that are not precise. Definition of some terms related to fuzzy set theory is given below.

Fuzzy Set

Let X be a universal set. Then the fuzzy subset A of X is defined by its membership function

$\mu_A(x): X \to [0,1]$

$\mu_A(x)$ gives the grade of the membership of x in A.

α-cut of a Fuzzy Set

The concept of an α-cut is one of the most important concepts of fuzzy sets. Let A be a fuzzy set on X and $\alpha \in [0,1]$. Then the α-cut is defined as the crisp set

$\alpha_A = \{x : A(x) \geq \alpha\}$

The strong α-cut is also the crisp set defined by

$\alpha_A^+ = \{x : A(x) > \alpha\}$

Support of a Fuzzy Set

The support of a fuzzy set A defined on X is the crisp set that contains all the elements of X having non-zero membership grades in A. That is, it is defined as

$\sup(A) = \{x \in X : \mu_A(x) > 0\}$

Height of a Fuzzy Set

The height of a fuzzy set A defined on X is the largest membership grade obtained by any element in that set. That is, it is defined as

$h(A) = \sup_{x \in x} \mu_A(x)$

Fuzzy Number

A fuzzy number A is a fuzzy set of the real line which is normal, convex and continuous membership function of bounded support.

Generalised Fuzzy Number

The membership function of generalised fuzzy number $[a,b,c,d : w]$ where $a \leq b \leq c \leq d, 0 < w \leq 1$ is defined as ([17],[18])

$$\mu_A(x) = \begin{cases} 0; x \leq a \\ w\dfrac{x-a}{b-a}; a \leq b \\ w\dfrac{d-x}{d-c}; c \leq d \\ 0; x > d \end{cases}$$

The generalised fuzzy number A will be a normal trapezoidal fuzzy number if $w=1$. If $a=b$ and $c=d$, then A will be crisp interval. It will be a generalised triangular fuzzy number if $b=c$. When $w=1$ and $a=b=c=d$ then A will be a real number. Since the parameter w represents the degree of confidence of opinions of decision makers, generalised fuzzy number can deal with uncertain information in a more flexible manner compared to normal fuzzy number.

The α cut of the generalised fuzzy number is given by

Chen's function principle

S. H. Chen [17] develops the theory and possible application of generalised fuzzy numbers. By proposing the function principle, different arithmetic operations on generalised fuzzy number were formulated in his work. In order to overcome of the complication arise in due to the use of extension principle, Chen investigated the function principle. Though function principle was used to develop arithmetic operations on generalised fuzzy nuber, in practice it has been realised that there are some limitations of Chen's method. From the mathematical point of view, it can be said that, computing different arithmetic operation using function principle is basically a pointwise operation (addition, subtraction, multiplication and division). Due to this reason, it has been observed that arithmetic operations of generalised trapezoidal (triangular) fuzzy number with the function principle cause the loss of information and do not give the exact result.

Proposed Hybrid Approach

In this section, first a hybrid has been proposed in which fuzzy numbers are combined by using function principle. After this we discuss the hybrid approach proposed by Dutta [19] and finally we compare the results.

Hybrid Approach I

Let us consider an arbitrary model

$M = g(P_1, P_2, ..., P_m, Q_{1,} Q_2, ..., Q_r)$

where $P_1, P_2, ..., P_m$ are m parameters represented by probabilistic distribution and $Q_{1,} Q_2, ..., Q_r$ are r parameters represented by generalised triangular fuzzy number with heights $w_1, w_2, ..., w_r$ respectively. Thus,

$Q_2 = [a_2, b_2, c_2; w_2]$ $Q_1 = [a_1, b_1, c\ ; w_1]$,..., $Q_r = [a_r, b_r, c_r; w_r]$

The approach is explained below:

Step I: Calculate α-cut for each generalised triangular fuzzy number. Since the height of each generalised triangular fuzzy number is different, so to perform the computation we consider Chen's function principle [17].Thus

$Q_1 = [a_1, b_1, c_1; \min(w_1, w_2, ..., w_r)]$, $Q_2 = [a_2, b_2, c_2; \min(w_1, w_2, ..., w_r)]$, ..., $Q_r = [a_r, b_r, c_r; \min(w_1, w_2, ..., w_r)]$

For this first we calculate for $\alpha = 0$ i.e., we will find 0-cut. From this we get r closed intervals. Thus we get $2r$ numbers of values.

Step II: Generate m numbers of uniformly distributed random number from $[0,1]$ and we perform Monte Carlo simulation to obtain m numbers of random numbers by sampling probability distribution.

Step III: Assign all $2r$ values and m random numbers in the model and calculate minimum and maximum value of the model M . i.e.,

$M_1^{\sup} = \sup(M)$ and $M_1^{\inf} = \inf(M)$

Step IV: Repeat step I to step III for 5000 times.

Step V: Plot CDF of $\left(M_1^{\sup}, M_2^{\sup}, ..., M_{5000}^{\sup}\right)$ and $\left(M_1^{\inf}, M_2^{\inf}, ..., M_{5000}^{\inf}\right)$

Step VI: Consider other α-levels to calculate α-cut for each generalised fuzzy number. It is to be noted that the greatest value of α is w, i.e., α values within the interval $[0, w]$

Step VII: Repeat step II to step V.

If we proceed in this way, we get a family of CDF's. From these CDF's membership functions at different fractiles can be generated. In this hybrid approach the shape of the generalised triangular type fuzzy number remain same. At different fractiles the shape of the resultant membership function will be generalised triangular type fuzzy number.

Hybrid Approach II

Let us consider a arbitrary model $M = g(P_1, P_2, ..., P_m, Q_{1,} Q_2, ..., Q_r)$

where $P_1, P_2, ..., P_m$ are m parameters represented by probabilistic distribution and $Q_{1,} Q_2, ..., Q_r$ are r parameters represented by generalised fuzzy number with heights $w_1, w_2, ..., w_r$ respectively.

The approach is given below:

Step I: Calculate α-cut for each generalised fuzzy number. The height of each generalised fuzzy number is different.

Thus to perform computation we consider α values within the interval $[0,w]$ where $w=\min(w_1,w_2,...,w_r)$. For this first we calculate for i.e., we will find 0-cut. From this we get r closed intervals. Thus we get $2r$ numbers of values.

Step II: Generate m numbers of uniformly distributed random number from $[0,1]$. Now we perform Monte Carlo simulation to obtain m numbers of random numbers by sampling probability distribution.

Step III: Assign all $2r$ values and m random numbers in the model and calculate $M_1^{\sup}=\sup(M)$ and $M_1^{\inf}=\inf(M)$

Step IV: Repeat step I to step III for 5000 times.

Step V: Plot CDF of $\left(M_1^{\sup},M_2^{\sup},...,M_{5000}^{\sup}\right)$ and $\left(M_1^{\inf},M_2^{\inf},...,M_{5000}^{\inf}\right)$.

Step VI: Consider other α-levels to calculate α -cut for each generalised fuzzy number. It is to be noted that the greatest value of α is w i.e., α values within the interval $[0,w]$

Step VII: Repeat step II to step V.

If we proceed in this way, we get a family of CDF's. From these CDF's membership functions at different fractiles can be generated.

Case study

A hypothetical case study for cancer risk assessment is presented for the demonstration of proposed hybrid method. Taking water became contaminated due to the release of radionuclide to water; we need to calculate cancer risk for ingestion pathway.

The risk assessment model due to the ingestion of radionuclides in water as provided by EPA [19] is follows:

$$Risk=\frac{C\times IR\times EF\times ED}{BW\times AT}\times CSF$$

Where C is the concentration (mg/L), IR is the ingestion rate (L/day), EF is the exposure frequency (days/year), ED is the exposure duration (years), BW is the body weight (kg), AT is the average time and CSF is the cancer slope or potency factor associated with ingestion $(mg/kg-day)^{-1}$.

The uncertain variables concentration C, intake rate IR and body weight BW are depicted in the Fig1.1 -Fig1.3. The result of the cancer risk assessment due to the ingestion of radionuclide in water using hybrid approach I is depicted in Fig.1.4. Here we take α =0.0, 0.4, 0.8 for simple and clear representation. At 80th fractile, risk value lies in the generalised fuzzy number [128.4729, 144.2585, and 162.2816]. The graphical representation of this generalised fuzzy number is depicted in Fig.1.5.

The result of the cancer risk assessment due to the ingestion of radionuclide in water using hybrid approach II is depicted in Fig.4.6. Here we take α =0.0, 0.4, 0.8 for simple and clear representation. At 80th fractile, risk value lies in the generalised fuzzy number 128.4705, 142.3367, 146.1837, 162.2786; 0.8]. The graphical representation of this fuzzy number is depicted in Fig.1.7.

Table:1 Values of the parameters for calculating cancer risk.

Variable	Representation	Value
Concentration C	Probabilistic	Normal(0.15,0.0005)
Intake Rate IR	Fuzzy	[4.75,5,5.25;0.8]
Exposure Frequency EF	Deterministic	350
Exposure Duration ED	Determinitic	30
Body Weight BW	Fuzzy	[0.14,0.15,0.16;1]
Averge Time AT	Deterministic	25550
Cancer Slope Factor CSF	Deterministic	70

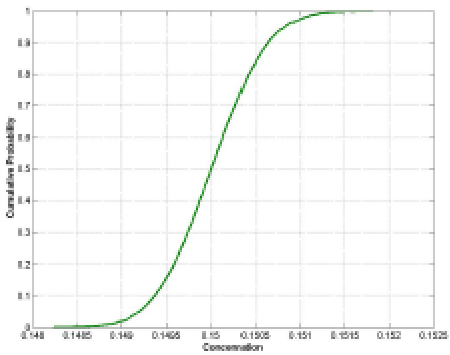

Figure 4.1: Uncertain Variable C

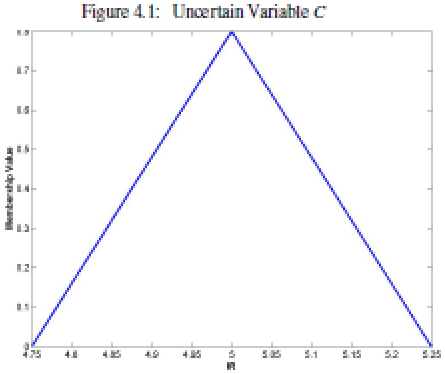

Figure 4.2: Uncertain Variable IR

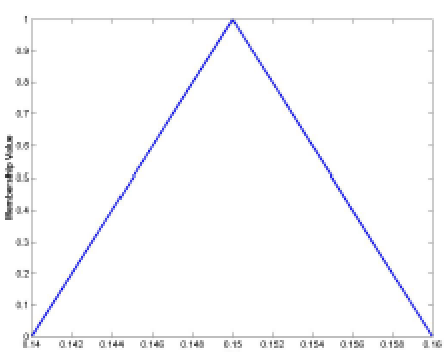

Figure 4.3: Uncertain Variable BW

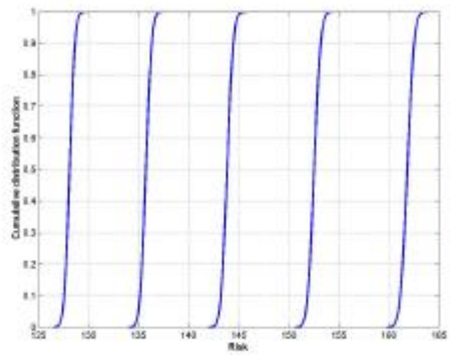

Figure 4.4: Cumulative distribution function of Risk for different α values

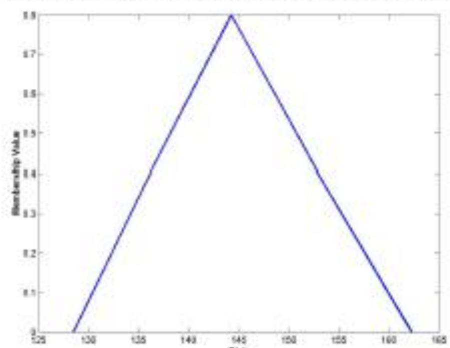

Figure 4.5: Membership function of Risk at 80th fractile

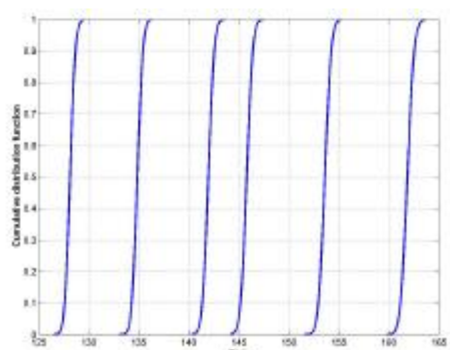

Figure 4.6: Cumulative distribution function of Risk for different α values

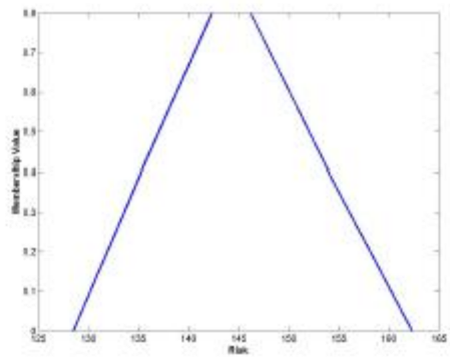

Figure 4.7: Membership function of Risk at 80th fractile

Conclusion

In this paper, we have considered risk model where input parameters are tainted with epistemic and aleatory uncertainty and risk assessment has been carried out using the hybrid approach. We have studied generalised fuzzy number to represent epistemic uncertainty and tried to fuse with probability distribution. To demonstrate and make use of generalised fuzzy number and the combination of generalised fuzzy number with probability distribution, a hypothetical case study for health risk assessment is presented here by applying two hybrid approaches.

Using the first hybrid approach based on Chen's function principle, risk has been obtained in the form of CDF's and from which membership functions of the risk have been generated at different fractiles. The membership function at different fractiles is generalised triangular fuzzy numbers. The result of the cancer risk assessment due to the ingestion of radionuclide in water using hybrid approach I is depicted in Fig.1.4. Here we take α =0.0, 0.4, 0.8 for simple and clear representation. At 80th fractile, risk value lies in the generalised fuzzy number [128.4729,144.2585, 162.2816;0.8]. The graphical representation of this generalised fuzzy number is depicted in Fig1.5.

Using the Dutta and Ali [14] hybrid approach risk has been obtained in the form of CDF's and from which, membership functions of the risk have been generated at different fractiles. The membership function at different fractiles are generalised trapezoidal fuzzy numbers.

The result of the cancer risk assessment due to the ingestion of radionuclide in water using hybrid approach II is depicted in Fig.1.6. Here we take α =0.0, 0.4, 0.8 for simple and clear representation. At 80th fractile, risk value lies in the generalised fuzzy number [128.4705,142.3367,146.1837,162.2786;0.8. The graphical representation of this generalised fuzzy number is depicted in Fig.1.7.

In this study, it is observed that in hybrid approach I, the shape of the resultant risk at certain fractile is generalised triangular fuzzy number while in approach II it is generalised trapezoidal fuzzy number.

References

[1] IAEA Safety series, 1992. IAEA procedure for conducting probabilistic safety assessment of nuclear power plants (Level 1). Safety Series, 50, pp. 4.

[2] K. D. Rao, 2009. Treatment of aleatory and epistemic uncertainty in safety assessment. Uncertainty modelling and analysis. Bhaba atomic research centre, pp. 125-152

[3] M.M. Granger and M. Henrion,1992. Uncertainty-A guide to dealing uncertainty in quantitative risk and policy analysis. Cambridge University Press.

[4] P.S. Jackson, R.W. Hockenbury and M.I. Yeater, 1981. Uncertainty analysis of system reliability and availability assessment.Nuclear engineering and design, 68, pp.5-29

[5] H. Tanaka, L.T. Fan, F.S. Lai and K. Toguchi, 1983. Fault tree analysis by fuzzy probability. IEEE transactions on reliability, 32, pp.453-457.

[6] K. Durga Rao, H.S. Kushwaha, A.K. Verma and A. Srividya, 2006. Uncertainty propagation in availability assesment of complex engineering

systems. International Journal on Dependability and Quality Management, 10, pp.1-17.

[7] H. Bae, R.V. Grandhi and R.A. Canfield,2004.Epistemic uncertainty quantification techniques including evidence theory for large scale structures. Computers and Structures, 82, pp.11101-1112.

[8] S. Ferson and J.G. Hajagos,2004. Arithmetic with uncertain numbers: rigorous and (often) best possible answers. Reliability Engineering and System safety, 85, pp.135-152.

[9] D. Guyonnet, B. Bourgine, D. Dubois, H. Fargier, B. Côme and J. P. Chilès. Hybrid approach for addressing uncertainty in risk assessments", Journal of Environmental Engineering, 126, pp. 68-78.

[10] C. Baudrit, D. Dubois, D. Guyonnet and H. Fargier, 2004. Joint Treatment of imprecision and Randomness in Uncertainty Propagation. Proc. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 873-880.

[11] C. Baudrit, D. Dubois and D. Guyonnet,2006. Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment. IEEE Transaction on Fuzzy System, 14, pp. 593-608.

[12] E. Kentel and M. M.Aral, 2004. Probabilistic-fuzzy health risk modelling. Stoch Envir Res and Risk Ass, 18, pp. 324-338.

[13] J. Li, G. H. Huang, G. M. Zeng, I. Maqsood and Y. F. Huang, 2007. An integrated fuzzy-stochastic modelling approach for risk assessment of groundwater contamination. Journal of Environmental Management, 82, pp.173-188.

[14] P. Dutta and T. Ali, 2012. A Hybrid Method to Deal with Aleatory and Epistemic Uncertainty in Risk Assessment. International Journal of Computer application. 42, pp. 37-44.

[15] P. Dutta,2013. An approach to deal with aleatory and epistemic uncertainty within the same framework: Case study in risk assessment. International Journal of Computer application, 80(12) , pp. 40-45.

[16] L. A. Zadeh , 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and system,1pp 3-28.

[17] S.H.Chen, 1985. Operations on fuzzy numbers with function principal. \textit{ Tamkang J.Manag.Sci}, 6, pp. 13-25.

[18] S.H.Chen,1999. Ranking generalized fuzzy number with graded mean integration. proc.8th Int.Fuzzy Syst.Assoc.World Congr, 2, pp. 899-902.

[19] P. Dutta,2013. Combined approach to propagate aleatory and epistemic uncertainty in risk assessment. International Journal of Mathematics and Computer applications Research}, 3, pp.29-35.

[20] EPA U.S., 2001. Risk assessment guidance for superfund, vol-1: Human health evaluation manual). Office of emergency and remedial response, Part E.