International Journal of Mathematical Archive-2(8), 2011, Page: 1300-1309
Available online through www.ijma.info ISSN 2229 – 5046

RANDOM SET BASED UNCERTAINTY EVALUATION OF HAZARD INDEX THROUGH INGESTION OF TOXIC CHEMICALS

A. K. Ranade*, M. Pandey and D. Datta

Computational Radiation Physics Section, Health Physics Division, Bhabha Atomic Research Centre, Mumbai Maharashtra- 400085, Telephone and fax number: 022 2559 0378, 022 2550 5151

*Email: *akranade@barc.gov.in, pandeym@barc.gov.in, ddatta@barc.gov.in*

(Received on: 01-07-11; Accepted on: 10-07-11)

ABSTRACT

Monitoring of concentration level of uranium in groundwater provides the basic knowledge of the activity of the uranium in groundwater. Quantification of the hazard index due to ingestion of uranium through drinking of ground water will facilitate the health impact of the uranium. Model representing the computation of this hazard index contain parameters which are uncertain due to their insufficient information and hence uncertainty analysis of the hazard index necessitates. Monte Carlo simulation is generally applied to compute the uncertainty .Monte Carlo simulation requires large amount of realizations of parameter uncertainty which are addressed by their respective probability density function. Insufficient information of the parameter does not allow to characterize their proper probability distribution. In this paper, probability distribution of the uncertain parameter with insufficient information has been achieved. The computational methodology has been devised to generate the probability density function of the uncertain parameter using imprecise probability. Random set theory along with vertex method have been implemented to evaluate the target uncertainty. In this context, the present paper describes the random set theory and its application for uncertainty analysis of hazard index.

Keywords: *Random set, vertex method, uncertainty, concentration, uranium, monitoring.*

INTRODUCTION

Risk assessments at present circumstances have become important aids in the decision making process related to the management of sources of pollutant, the issue of uncertainty with respect to model parameter value is of primary importance. Uncertainty affecting parameters in risk assessments can be of different nature (ambiguity, vagueness, imprecision, ignorance, etc.) and can be represented in various fashions (probability density functions, fuzzy numbers, uncertainty intervals, etc.). While different type of uncertainty may warrant different modes of uncertainty representation, the question arises as to how several modes of representation can be accommodated in the same estimation of risk. Researcher's have used some of the modern mathematical theories of uncertainty in the context of epistemic uncertainty (lack of knowledge) and their relevance in risk analysis, using probability distribution function (Labieniec et al 1997) , probability theory (Prado et al 1999) and fuzzy theory (Doc C 1995;Bardossy 1995; Freissinet 1998). Traditional Monte Carlo method of uncertainty analysis demands the probability density function (PDF) of the model inputs and these PDFs results by a large number of experimental studies. However imprecise probability input parameters of representative risk model has less information and are addressed by fuzzy set theory. All the uncertain fuzzy input parameters in this case are described as triangular fuzzy numbers. Reason behind the consideration of triangular fuzzy number is due to the linguistic phrase of measurement as "around ξ", where ξ is an imprecise measurement. Uncertainties in risk estimates might arise from many different sources such as measurements or estimation of parameters, environmental monitoring of data, natural variability in individual response, variability in environmental concentration of toxicant/radionuclides over time and space. Uncertainty analysis was the part of risk assessment that focuses on uncertainties in the assessment. Results of environmental monitoring are reported in an interval as minimum and maximum, In real practice several such intervals should have an appropriate frequency that can be further expressed in terms of a cumulative probability number called as basic probability assignment. In the probability domain basic probability mass assignment is also called as cumulative probability. Imprecise probability is defined as set of random intervals with the corresponding basic probability mass assignment. Alternatively, this set is known as random set. Random set can be used to find the epistemic uncertainty of the risk involved. Random set has been used in several area of science and engineering viz design of structure, analysis of expert opinion, reliability bounds, decision making etc. (Tonon et al 1999; Oberguggenberger 2008) .The risk management plan should propose

---

***Corresponding author:** A. K. Ranade*, ***E-mail:** akranade@barc.gov.in*

applicable and effective controls for managing and mitigating the risks. In the field of nuclear technology risk analysis/management due to exposure from the drinking of contaminant is mandate. Deterministic or complete probabilistic analysis is ruled out with respect to the available information on the model input parameters. Section 2 describes the problem in detail and addresses uncertainty affecting the information available on each parameter. The extension principle for random set is utilized to propagate uncertainty through a model response which is discussed elsewhere (Dubois 1991; Klir 1989).

MATERIALS AND METHODS

Following the guidelines as given in US EPA 2000 (Dawoud 1996), risk due to the ingestion of a chemical can be written as

$$\text{LADD of drinking water} = \text{CDI} \times \text{IR} \times \text{EF} \times \frac{\text{ED}}{\text{AT} \times \text{BW}} \tag{1}$$

where $LADD$ signifies the lifetime average daily dose (μg·kg^{-1}·day^{-1}); CDI, the cumulative daily intake (μg·l^{-1}); IR, the ingestion water rate (l·day^{-1}); EF, exposure frequency (days·year^{-1}); ED, total exposure duration (years); AT, average time (days); BW, body weight (kg).

Assuming all parameter are independent, i.e. knowledge about the value of one parameter implies nothing about the value of the other. The main objective is to quantify the uncertainty in $LADD$ and also to assess the uncertainty in Hazard quotient (HQ)/Hazard Index (HI).

Among the input parameters as described in the model (equation (1)) experimentally measured values of uranium concentration in groundwater has been used for the uncertainty analysis of LADD. The data related to the body weight, breathing rate and relevant dose conversion factor reported elsewhere (Dawoud 1996; Babu 2008) are used in the analysis. The total exposure duration (set in the computation as 76 years), averaging time and risk coefficient were all used as deterministic input to the present model (Dawoud 1996). The ratio of $LADD$ and the reference dose (RfD) of 0.6 μg·kg^{-1}·day^{-1} (Gilman 1998) relates to the hazard quotient. Hazard quotient for chemical risk therefore can be written as

$$\text{Hazard quotient(HQ)or Hazard Index (HI)} = \frac{\text{LADD}}{\text{RfD}} \tag{2}$$

All the parameters (equation (2) and equation (1)) are modeled by a triangular probability density function and information on EF is given by independent and two equally credible sources of information.

Cumulative daily intake (CDI) is taken as triangular probability density function (PDF) and discretisation is done on the intervals [CDI_{min}, CDI_{mod}] and [CDI_{mod}, CDI_{max}] into n_1 and n_2 subintervals Z_{CDI} = [a_i, b_i] respectively. Discretization is nothing but the division or partition of the intervals into several focal elements each with different basic probability assignment. Each subinterval Z_{CDI} is treated as a focal element. If p(CDI) be the PDF of CDI and P (CDI) the Cumulative Distribution Function(CDF) of CDI then basic probability assignment (m) for focal element Z_{CDI} is evaluated as

$$m_{CDI}(Z_{CDI}) = P_{CDI}(b_i) - P_{CDI}(a_i) \tag{3}$$

The area between the two points on a normalized triangle evaluated from similar triangle method is given by

$$m_{CDI} = \frac{(2k-1)n^2}{(B-A)(C-A)} \qquad for\ Z_{CDI} \in [CDI_{min}; A, CDI_{mod}; B] \tag{4}$$

$$m_{CDI} = [n - \frac{(2k-1)n^2}{2(C-B)}] \Big/ (\frac{C-A}{2}) \qquad for\ Z_{CDI} \in [CDI_{mod}; B, CDI_{max}; C] \tag{5}$$

where A, C are extreme bound and B is the middle most value, and the index k = 1, 2, …., n_1 ;n being the length of subinterval. Similar equation can be written for IR and BW.

As far as exposure frequency is considered, two bodies of evidences are assigned and each body of evidence was affected by both dissonance and imprecision (Dubois 1991; Walley 1991). Accordingly aggregation or fusion of knowledge on exposure as evidence is combined using Dempster's rule of combination (Sentz 2002). One should take the upper and lower CDF for ith source of information which was defined by a triangular distribution. As we are estimating focal element from CDF of distribution it leads to

$$U_{EF,i} = \frac{(EF - EF_{min\,1})^2}{(EF_{mod\,1} - EF_{max\,1})(EF_{max\,1} - EF_{min\,1})} \quad EF \in [EF_{\min 1}, EF_{\text{mod}\,1}] \tag{6}$$

$$= 1 - \frac{(EF_{mod\,1} - EF)^2}{(EF_{mod\,1} - EF_{min\,1})(EF_{mod\,1} - EF_{max\,1})} \quad EF \in [EF_{mod\,1}, EF_{\max 1}]$$

Similar equation can be written for $L_{EF,i}$ for $\left[EF_{\min 2}, EF_{\mathrm{mod}\,2}, EF_{max\,2}\right]$. Discretization is done in such a manner so that upper and lower probabilities U_{EF} and L_{EF} becomes as Belief and Plausibility function respectively. So the focal element A_{EF} can be written as

$$A_{EF} = \left[U_{EF}^{-1}\left(\frac{1}{n}\left(j - \frac{1}{2}\right)\right), L_{EF}^{-1}\left(\frac{1}{n}\left(j - \frac{1}{2}\right)\right)\right] \qquad j = 1, \ldots\ldots, n \tag{7}$$

and the basic probability assignment (bpa) for that focal element $m_{EF,i}(A_{EF,i,j})$ is given by

$$m_{EF,i}(A_{EF,i,j}) = \frac{1}{n} \tag{8}$$

This produces a random set whose plausibility averaged to U_{EF} and whose belief averaged to L_{EF}. As a first approach, basic probability assignment is assumed as of equal value obtained by using equation (8). The two random sets are combined into a unique random set by the same philosophy that justifies the joint probability distribution of independent variables starting from their marginal distribution (Klir 1995). Here we generate the random set using Dempster's rule of combination as follows (Klir 1995):

$$m_{EF}(A) = \frac{\sum_{A=B\cap C} m_{EF}(B) m_{EF}(C)}{1 - K} \tag{9}$$

For all$A \neq \phi$, and $m_{EF}(A) \neq \phi$ where $K = \sum_{B\cap C=\emptyset} m_{EF}(B)\, m_{EF}(C)$

Shafer's framework allows for belief about propositions to be represented as intervals, bounded by two values, *belief* (or *support*) and *plausibility*: *Belief* in a hypothesis is constituted by the sum of the masses of all sets enclosed by it (i.e. the sum of the masses of all subsets of the hypothesis). It is the amount of belief that directly supports a given hypothesis at least in part, forming a lower bound. *Plausibility* is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. It is an upper bound on the possibility that the hypothesis could be true, i.e. it "could possibly be the true state of the system" up to that value, because there is only so much evidence which contradicts that hypothesis. The Dempster-Shafer theory, also known as the theory of belief function, is a generalization of Bayesian theory of subjective probability. The corresponding aggregated basic probability assignment obtained is normalized by $1 - K$ to take care of empty set which may arise due to intersection of focal elements (equation 13). The extension principle for random sets is used to map the relation discussed elsewhere.

RESULTS AND DISCUSSION

Uncertainty due to variability of input and/or model parameters when the corresponding variability characterization is not available, or uncertainty due to an unknown process or mechanism leads to epistemic uncertainty. The chemical risk due to ingestion of uranium in drinking water is computed on the basis of random set theory as described in section 2. All the input parameters are modeled by a triangular probability density function as shown in figure 1, 2, 3, 4 (Table I). Random sets of parameter CDI, IR and BW for ten focal elements, discretization is obtained from equation (4) and (5) as given in Table II, III, and IV respectively and corresponding cumulative basic probability for the parameters are given in figure 5, 6, 7. In this case HQ was mapped through four uncertain random parameters, namely CDI, EF, IR and BW. Columns 2 and 3 of Table V are relevant to first and second source of information for EF and they are obtained from equation (6, 7). The values of bpa for EF are obtained by equation (8). The CDFs for first and second source are shown in figure 8 and 9 respectively.

$$U_{EF,1} = \frac{(EF - 180)^2}{25500} \qquad if\ EF \in [180,330] \tag{10}$$
$$= 1 - \frac{(350-EF)^2}{3400} \quad if\ EF \in [330,350]$$

$$L_{EF,1} = \frac{(EF - 190)^2}{25500} \qquad if\ EF \in [190,340] \tag{11}$$
$$= 1 - \frac{(360 - EF)^2}{3400} \quad if\ EF \in [340,360]$$

$$U_{EF,2} = \frac{(EF - 175)^2}{25200} \qquad if\ EF \in [175,315] \tag{12}$$
$$= 1 - \frac{(355 - EF)^2}{7200} \quad if\ EF \in [315,355]$$

$$L_{EF,2} = \frac{(EF-200)^2}{22275} \qquad if\ EF \in [200,335] \tag{13}$$
$$= 1 - \frac{(365-EF)^2}{4950} \qquad if\ EF \in [335,365]$$

It should be noted that the different techniques of discretization was applied to EF and CDI, BW, IR; in the first case ordinate is discretized whereas in the second case abscissa is discretized. The different type of method of discretization used for EF is due to the constraints as attributed by upper and lower CDF. The combined CDFs for the exposure frequency (EF) are given in figure 10. The two source of information is combined by Dempster-Shafer rule of combination (9).

The four random parameters gave rise to four dimensional boxes with 2^4 vertices. If monotonic properties of a specified function is not known then all 16 vertices has to evaluate to ascertain random set otherwise 4 would suffice. The extension principle for random set is used to map HI. The relevance of HI is important in risk analysis. It tells about the status of management control over processes and activities. The imprecision affecting parameters CDI, EF, IR and BW were mapped to HQ and consequently two cumulative plots are obtained. Figure 11 represents the CDF of dependent variable HQ. The upper CDF curve represents the plausibility and the lower CDF represents the belief of the HI. One can interpret these two curves with respect to a specified HI in this way that the probability of occurrence of HI <= 20, is bounded by two cumulative probabilities [lower CDF = 0.4, upper CDF = 0.59]. That is to say that belief of Probability of HI <= 20 is 0.46 and the corresponding plausibility is 0.62. It should be noted that belief and plausibility are the two extreme bounds of the intervals representing the epistemic uncertainty in that case when basic information are evaluated based on evidence. It can also be said that classical probability is bounded by belief and palsubilitiy signifying that belief and plausibility together are called as imprecise probabilities. The larger the variation one would have in data leads to larger uncertainty. The 50th percentile of the cdf plot could be used for administrative control in a system. This method provides the mechanism to deal with epistemic uncertainty involving domain knowledge. This method can be used to evaluate epistemic uncertainty in model parameters.

CONCLUSION

Random set theory based uncertainty quantification was discussed. Uncertainty evaluation of hazard index is presented. It is concluded that imprecise information on the parametric uncertainty can be handled with the random set theory. Parameters of the model considered are assumed as independent but in presence of correlation also the random set theory can be applied. Future work will be focused on the uncertainty evaluation of the model in presence of the correlated non-random parameters.

ACKNOWLEDGEMENTS

The authors sincerely acknowledge the support provided by Dr. P.K. Sarkar, Head, Health Physics Division, BARC.

REFERENCES

[1] Bardossy A, Bronstert A, and Merz B.. 1, 2 and 3 dimensional modeling of groundwater movement in the unsaturated soil matrix using a fuzzy approach. Advances in Water Resources, 18(4) (1995),237-251.

[2] Babu M. N. S, Somashekar R K, Kumar S A, Shivanna K et al. Concentration of uranium levels in groundwater International Journal Environmental Science Techology 2 (2008), 263-266.

[3] Dubois D, Prade H. Random sets and fuzzy interval analysis. Fuzzy sets and systems, 42(1) (1991), 87-101.

[4] Dou C, Woldt W, Bogardi I, and Dahab M. Steady-state groundwater flow simulation with imprecise parameters. Water Resources Research 31(11) (1995), 2709-2719.

[5] Dawoud E.A., Purucker S.T. Quantitative uncertainty analysis of superfund residential risk pathway model for soil and ground water, US DOE (1996).

[6] Freissinet C, Erlich M, and Vauclin M. A fuzzy logic-based approach to assess imprecision of soil water contamination modeling. Soil & Tillage Research 47(1998), 1-17.

[7] Gilman A P. Uranyl Nitrate: 28-Day and 91-Day Toxicity Studies in the Sprague-Dawley Rat. Villeneuve Toxicological Science 41(1998),117-128.

[8] Klir G J. Is there more to uncertainty than some probability theorists might have us believe?, International Journal of General Systems 15(1989), 347 – 378.

[9] Klir G J, Yuan B. Fuzzy sets and fuzzy logic, theory and application. Upper Saddle River, NJ:Prentice-Hall (1995).

[10] Labieniec P, Dzombak D, and Siegrist R. Evaluation of uncertainty in a site-specific risk assessment. Journal of Environmental Engineering 123(3) (1997), 234-243.

[11] Oberguggenberger M, Felin W. Reliability bounds through random set: Non parametric methods and geotechnical application. Computers and structures 86(10) (2008),1093-1101.

[12] Prado P, Draper D, Saltelli S, Pereira A, Mendes B, Eguilior S, Cheal R, Tarantola S. Conceptual and computational tools to tackle the long-term risk from nuclear waste disposal in the geosphere. Office for Official Publications of the European Communities, Luxemburg. European Commission Report EUR 19113 EN (1999).

[13] Sentz K, Ferson S. Combination of evidence in Dempster-Shafer theory, Albuquerque, NM. Sandia National Laboratories Technical Report SAND 0835 (2002).

[14] Tonon F, Bernardini. A. Concept of random set as applied to the design of structures and analysis of expert opinion for aircraft crash. Chaos solutions and fractals 10(11) (1999),1855-1868.

[15] Walley P. Statistical reasoning with imprecise probabilities. London Chapman & Hall (1991) A K Ranade, M. Pandey and D. Datta.

Table: I Value of the model input parameters is expressed in terms of an interval (units are as described in text)

Parameter	Minimum value	Maximum value	Mode value
CDI	10	1500	400
EF	180-190	350-360	330-340
IR	0.6	1.9	1.2
BW	64	84	75

Table: II Discretization of parameter CDI into 10 focal elements

Focal element A_k	Basic probability assignment M_{CDI}
[10,88]	0.01
[88,166]	0.03
[166,244]	0.05
[244,322]	0.07
[322,400]	0.09
[400,620]	0.26
[620,840]	0.21
[840,1060]	0.15
[1060,1280]	0.09
[1280,1500]	0.03

Table: III Discretization of Elements parameter IR into 10 focal

Focal element A_k	Basic probability assignment M_{IR}
[0.6,0.72]	0.018
[0.72,0.84]	0.055
[0.84,0.96]	0.092
[0.96,1.08]	0.129
[1.08,1.2]	0.166
[1.2,1.34]	0.194
[1.34,1.48]	0.151
[1.48,1.62]	0.108
[1.62,1.76]	0.064
[1.76,1.9]	0.022

Table: IV Discretization of parameter BW into 10 focal elements

Focal element A_k	**Basic probability assignment M_{BW}**
[64,66.2]	0.02
[66.2,68.4]	0.06
[68.4,70.6]	0.11
[70.6,72.8]	0.15
[72.8,75]	0.19
[75,76.8]	0.16
[76.8,78.6]	0.12
[78.6,80.4]	0.09
[80.4,82.2]	0.05
[82.2,84]	0.05

Table: V Focal element of parameter EF for ten discretization

j	Focal element A_{EF},1	Focal element A_{EF}, 2
1	[210,235]	[215,225]
2	[236,261]	[241,251]
3	[254,279]	[259,269]
4	[268,293]	[274,284]
5	[281,306]	[287,297]
6	[292,317]	[298,308]
7	[302,327]	[308,318]
8	[312,329]	[318,328]
9	[322,337]	[327,337]
10	[336,349]	[336,346]

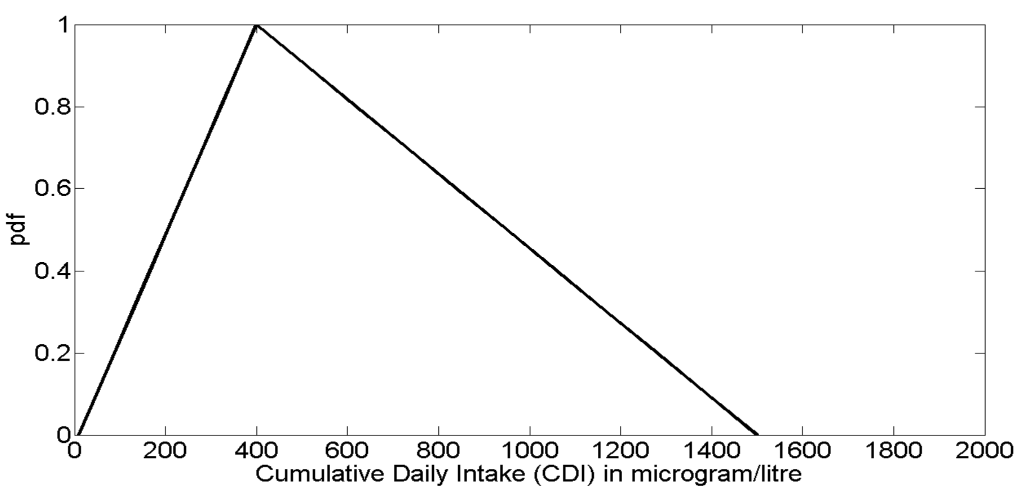


Figure: 1

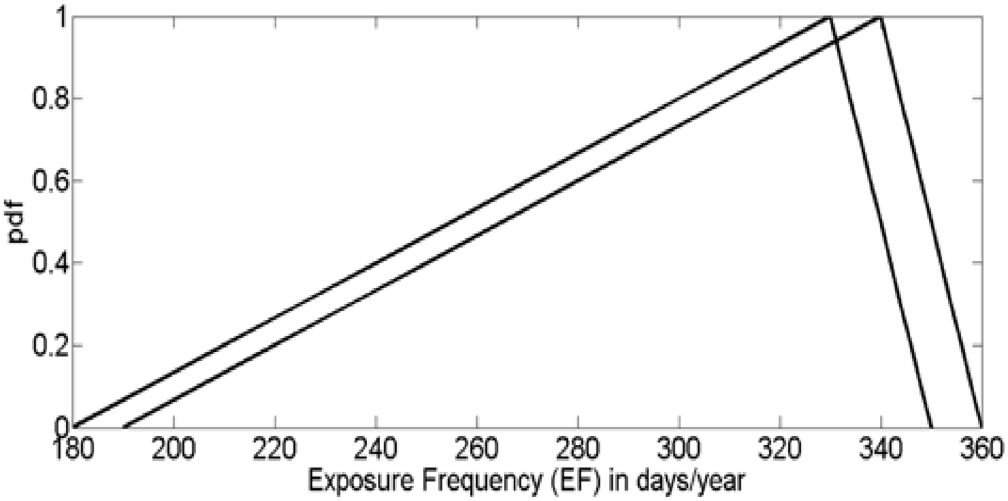


Figure: 2

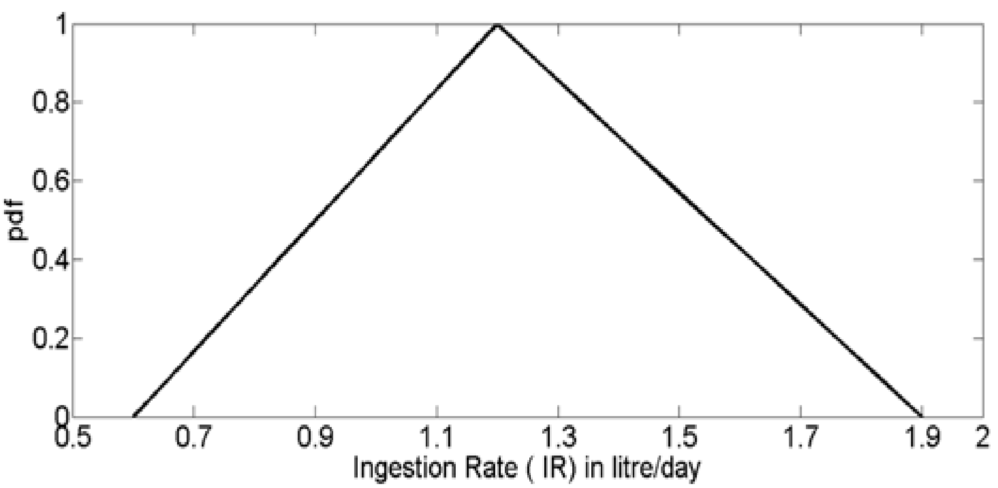
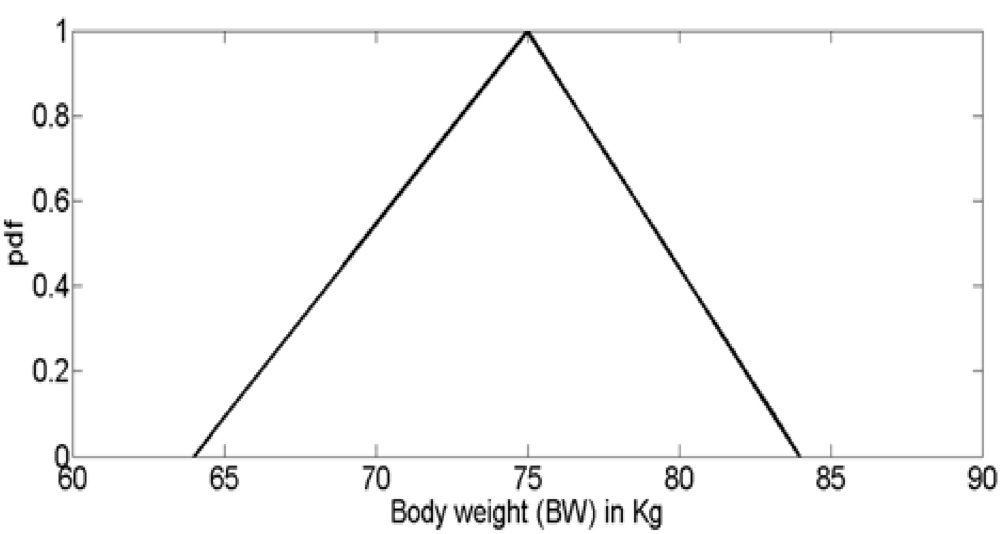


Figure: 3



z

Figure: 4

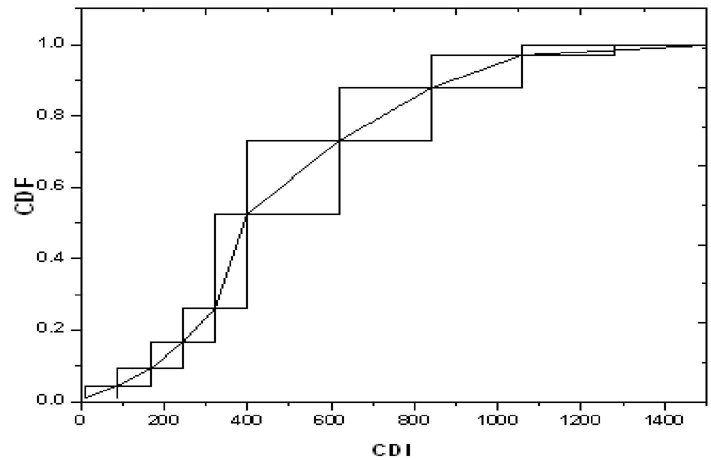


Figure: 5

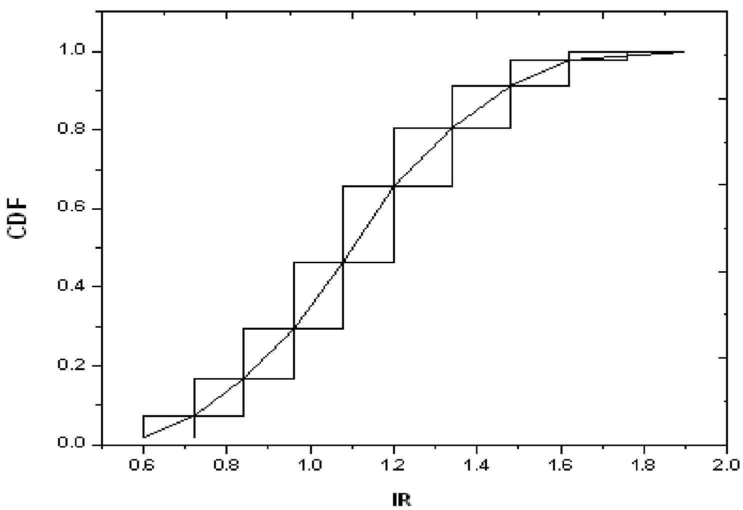


Figure: 6

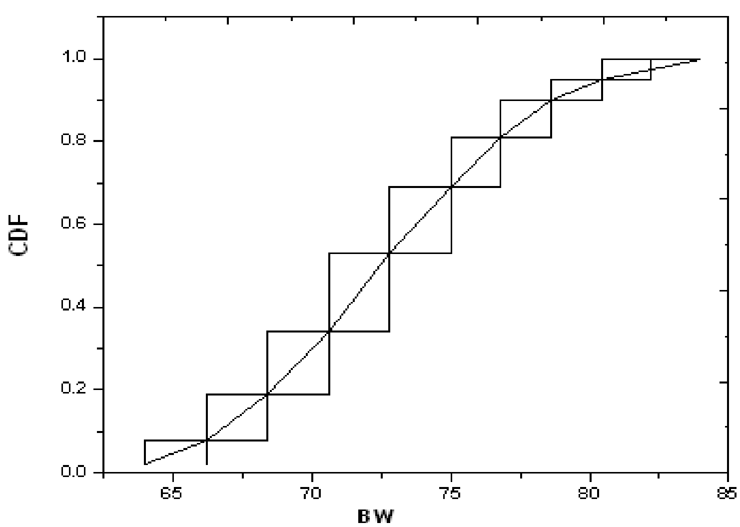


Figure: 7

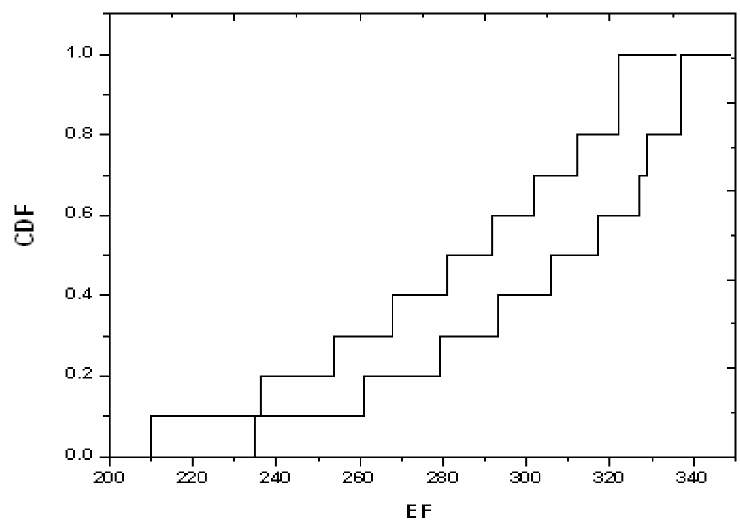


Figure: 8

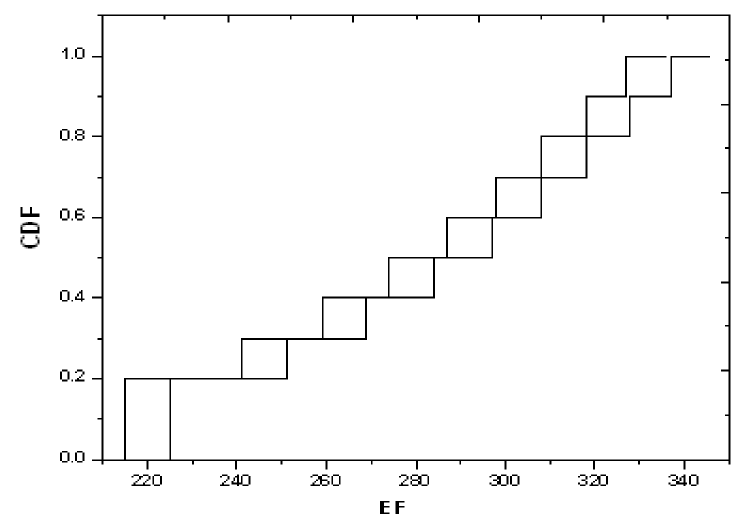


Figure: 9

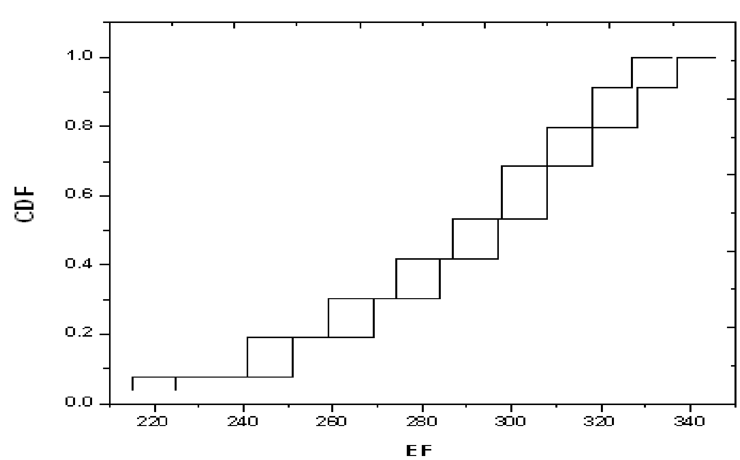


Figure: 10

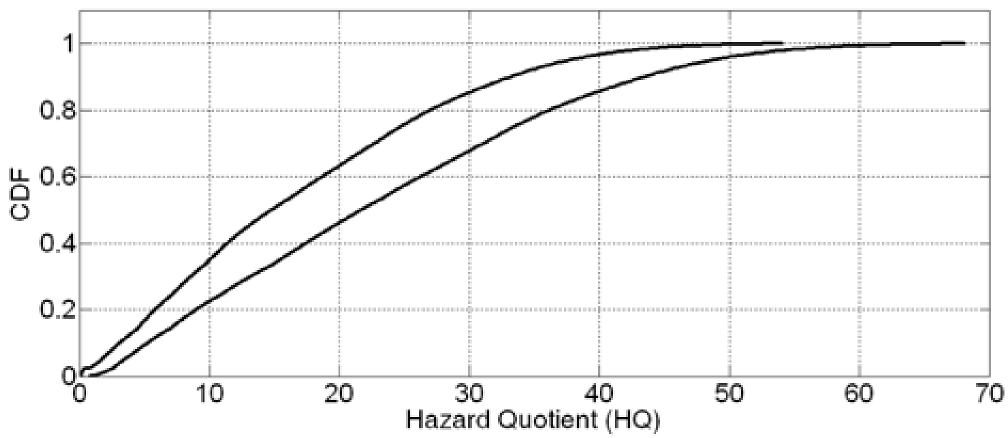


Figure: 11

Figure1: Probability distribution function of parameter Cumulative Daily Intake (CDI) in microgram/litre.

Figure 2: Probability distribution function of parameter Exposure Frequency (EF) in days/year.

Figure 3: Probability distribution function of parameter ingestion rate (IR) in litre/day.

Figure 4: Probability distribution function of parameter body weight (BW) in Kg.

Figure 5: Cumulative distribution function and upper and lower CDF of parameter CDI for ten discretization.

Figure 6: Cumulative distribution function and upper and lower CDF of parameter IR for ten discretization.

Figure 7: Cumulative distribution function and upper and lower CDF of parameter BW for ten discretization.

Figure 8: Upper and lower CDFs of parameter EF according to first source of information for ten discritization.

Figure 9: Upper and lower CDFs of parameter EF according to second source of information for ten discritization.

Figure 10: Upper and lower CDFs of parameter EF obtained by combining the information from the two sources of information.

Figure 11: Upper and lower CDFs for the response of Hazard Quotient.

****

AFFILLIATION(S): Computational Radiation Physics Section, Health Physics Division, Bhabha Atomic Research Centre, Mumbai-94, Telephone and fax number: 022 2559 0378, 022 2550 5151,

****

