Computational modeling and water analysis of parameter estimation by uncertainty quantification in water resources

R Judith Kiruba¹, G Geethalakshmi², R Ida Malarselvi³, R Irene Hepzibah⁴, A. Deepika¹, R Priscilla⁵ and C Ramachandra Raja⁶

¹Department of Computer Science, TBML College, Affiliated to Bharathidasan University, Porayar – 609 307, Nagapattinam, Tamil Nadu, India.
²Department of Mathematics, AVC College of Engineering, Affiliated to Bharathidasan University, Mannampandal, Mayiladuthurai –609 305, Tamil Nadu, India.
³Department of Physics, D G Government Arts College, Affiliated to Bharathidasan University, Mayiladuthurai – 609 001, Tamil Nadu, India.
⁴Department of Mathematics, TBML College, Affiliated to Bharathidasan University, Porayar – 609 307, Nagapattinam, Tamil Nadu, India.
⁵Department of Physics, A.D.M. College for Women (Autonomous), Affiliated to Bharathidasan University, Nagapattinam – 611 001, Tamil Nadu, India.
⁶Department of Physics, Government Arts College (Autonomous), Affiliated to Bharathidasan University, Kumbakonam – 612 001, Tamil Nadu, India.

Email Id: idamalarselvi@gmail.com

Abstract: Falling losses from aquatic disease is the main public condition goal in developed countries. The present study terms the application of Physical Sciences and Environmental Sciences techniques to solve the real-life problem associated with the society. Water samples are taken from a river, pond, bore well, and municipality at Kabristhalam in Kumbakonam Taluk, Tanjore District, Tamil Nadu, South India. The current work defines the expansion of mathematical typical to expect the interaction parameters which were used to determine the quantitative characteristics of the water. Analysis of Variance (ANOVA) was employed to identify the level of importance in interaction parameters on their characteristics. The effect of these parameters on interaction has been investigated using experimental designs. The new results have been correlated using the Fuzzy Evidence Theory. The salient aspect of the work is that a very simplified analytic output of the fuzzy model is achieved when all fuzzy sets in the fuzzy partition of the output space have the same power (the area under the membership function), and the determination of basic probability assignments associated with fuzzy Dempster-Shafer belief structure using fuzzy focal elements.

Keywords: Water resources modeling, Interaction parameters, Analysis of variance, Experimental design, Fuzzy evidence theory.
1. Introduction
The standards for drinking water quality are typically set by the government or by International Standards. These standards usually include minimum and maximum concentration of contaminants, depending on the intended purpose of water use. Visual inspection cannot determine if the water is of appropriate quality. Simple procedures such as boiling are the use of a household activated carbon filter that is not sufficient for treating all the possible contaminants that may be present in water from an unknown source. Even natural spring water - considered safe for all practical purposes in the 19th century – must now be tested before determining what kind of treatment, if any, is needed. Chemical and microbiological analysis, while expensive, is the only way to obtain the information necessary for deciding on the appropriate method of purification. According to a 2007 World Health Organization (WHO) report, 1.1 billion people lack access to an improved drinking water supply, 88 percent of the 4 billion annual cases of diarrheal disease are attributed to unsafe water and inadequate sanitation and hygiene, while 1.8 million people die from diarrheal cases are preventable through modifications to the environment, including access of safe water. Simple techniques for treating water at home, such as chlorination, filters, and solar disinfection and storing it in safe containers could save a huge number of lives each year [1-3]. In this paper, by using Sengupta and Pal's method of comparison of interval numbers and a new set of arithmetic operations for interval numbers, we propose a theory for the study of arithmetic operations on interval numbers by K. Ganesan et al, 2005 [4]. As there are no reports available in the literature concerning the study of fuzzy sets of water analysis, we report here some relevant reviewed work. Groundwater model parameters are calibrated using a limited number of direct and indirect observations of the material properties. A calibrated model often represents a simple parameterization with a small number of parameters and simplified models corresponding to a limited set of observations. This approach leads to a well-known notion that calibration is a non-unique inverse problem [Carrera and Neuman, 1986] [5], i.e., some different parameter sets can match observations to the same degree [Beven and Binley, 2001] [6]. Predictions utilizing calibrated models inevitably have uncertainty (i.e., a range of unknown errors). Recently, quantification of the uncertainty associated with model predictions has been highlighted to better understand different sources of uncertainty [e.g., Renard et al., 2010] [7] and provide decision-makers with increasingly accurate and precise predictive models [Keating et al., 2010] [8]. However, uncertainty analysis has not been routinely used in water resources modeling, and Pappenberger and Beven [2006] [9] recommend utilizing model predictive uncertainty methods in management practices for many decision policies [10]. A mathematical model was developed to predict the parameters which are used to determine the quality of the water. The water samples were analyzed by the parameters pH, temperature, Nephelometric Turbidity Units (NTU), conductivity (mg/l), Alkalinity (mg/l) and BOD (mg/l). The experimental results have been correlated using the Fuzzy Evidence Theory.

2. Experiment
The samples remained collected and the various parameters pH, Temperature, Nephelometric Turbidity units, Conductivity, Alkalinity, and BOD were measured and the values are recorded. Analysis of
Variance (ANOVA) aims to find how much of the total variability is due to each factor and by comparing these contributory amounts of variations.

3. Materials and methods

The pH of the water sample was measured by a scale of 0-14 and a measure of hydrogen ions concentration and acidity in the sample. Suspended particles in water clarity were measured from Nephelometer Turbidity Units (NTU) with a range of 60 units and it is the main indicator of deferred sediment and erosion levels. The electrical conductivity of the sample was measured between the ranges 50 to 1500 micros/cm. Dissolved oxygen is measured in mg/L. The extreme solubility of oxygen in water at 1 atm pressure ranges from about 15 mg/L at 0°C to 8 mg/L at 30°C. The measure of alkalinity is the acid-neutralizing capacity of solutes in a water sample, reported in milliequivalents per liter, and it regulates the pH of a water body and also the metal contents. Water purification is the process of removing unwanted chemicals, biological impurities, suspended solids, and gases from polluted water. The goal is to produce water fit for a specific purpose most water is disinfected for human conception but water purification may also be designed for a variety of other purposes, including fulfilling the requirements of medical pharmacological chemical and industrial application. The standards for drinking water quality are typically set by the government or by international organizations. These standards typically include minimum and maximum concentration of contaminants, depending on the intended purpose of water use. Temperature is usually expressed in degrees Fahrenheit or Celsius and the maximum temperature is the highest water temperature at which the organisms will live for a few hours.

3. MATHEMATICAL MODELLING FOR WATER ANALYSIS THROUGH QUALITY PARAMETERS USING FUZZY EVIDENCE THEORY

Dempster-Shafer Theory of Evidence

We introduce some basic concepts and mechanisms of the Dempster-Shafer theory of evidence [11-14] which required for our procedure. The Dempster-Shafer theory is a formal framework for plausible reasoning providing methods to represent and combine weights of evidence. Let $\Theta$ be a finite set of mutually exclusive and exhaustive events or hypotheses about the problem domain, called the frame of discernment.

A Dempster-Shafer belief structure, information granule $m$, is a collection of non-null subsets of $\Theta$, $A_i$, $i = 1 \ldots n$, called focal elements, and a set of associated weights $m(A_i)$, called basic probability assignment (BPA). This BPA must be such that

$$m(A_i) \in [0, 1], \quad m(A_i) \neq 0, \quad \sum_i m(A_i) = 1 \quad (1)$$

When our knowledge is of the form of a Dempster-Shafer theory belief, because of the imprecision in the information, when attempting to try to find the probabilities associated with arbitrary subsets of $\Theta$ we can’t find exact probabilities but lower and upper probabilities.

Firstly one measure, Bel, is introduced to capture the relevant information. Let $B$ be a subset of $\Theta$, we define

$$\text{Bel}(B) = \sum_{A_i \subseteq B} m(A_i) \quad (2)$$

$$\text{Bel}(B) = \sum_{A_i \subseteq B} m(A_i) \quad (2)$$
Then we define $P_l(B) = 1 - \text{Bel}(B)$. \hspace{1cm} (3)

One advantage of Dempster-Shafer theory is its capability to express degrees of ignorance, that is the belief in an event and the belief in its opposite do not necessarily add up to one like in probability theory. A situation of total ignorance is characterized by $m(\Theta) = 1$.

Assume that $m_1$ and $m_2$ are two independent belief structures on a frame of discernment $\Theta$, with focal elements $A_i, i = 1, \ldots, n_1$, and $B_j, j = 1, \ldots, n_2$. Then the conjunction of $m_1$ and $m_2$ is another belief structure $m = m_1 \oplus m_2$ whose focal elements are all the subsets $F_k$ of $\Theta$, where $F_k = A_i \cap B_j$ and $F_k \neq \emptyset$. The basic probability numbers associated with each $F_k$ are defined as

$$m(F_k) = \frac{1}{1 - T} \left( m_1(A_i) \ast m_2(B_j) \right) \hspace{1cm} (4)$$

where

$$T = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \ast m_2(B_j)$$

Now the concept of the fuzzy Dempster-Shafer belief structure can be introduced. A fuzzy Dempster-Shafer belief structure is a Dempster-Shafer belief structure with fuzzy sets as focal elements [6]. When we combine two fuzzy Dempster-Shafer belief structures using a set operation $\nabla$, we simply use its fuzzy version.

Some basic concepts of D-S theory are briefly introduced as follows,

3.1.2 Frame of Discernment:
Let $\{\theta_1, \theta_2, \ldots, \theta_n\}$ be a set called frame of discernment, if it contains mutually exclusive and exhaustive events.

3.1.3 Basic Probability Assignments (BPA):
A function $m : 2^{\Theta} \rightarrow [0,1]$ is called a Basic Probability Assignment (BPA) on $\Theta$ if it satisfies the following three properties:

(i) $m(\emptyset) = 0$
(ii) $m(A) \geq 0$
(iii) $\sum_{A \subseteq \Theta} m(A) = 1$, $A \in 2^{\Theta}$ is called a focal element of being’ satisfies $m(A) \geq 0$

3.1.4 Belief Function:
From the BPA, a function $Bel(A) : 2^{\Theta} \rightarrow [0,1]$ is defined as $Bel(A) = \sum_{B \subseteq A} m(B)$ .

3.1.5 Plausibility Function:
From the BPA, a function $Pls(A) : 2^{\Theta} \rightarrow [0,1]$ is defined as $Pls(A) = \sum_{A \cap B = \emptyset} m(B)$ .

3.1.6 Dempster Rule of Combination:
Let $m_1$ & $m_2$ be two mass functions defined on the same frame of discernment, $\Theta$ and then a combined BPA can be obtained by using Dempster’s combination rule, the combined BPA
This is known as the Dempster rule of combination [9,10,12].

\[
A/B = [a_1, a_2] / [b_1, b_2] = [a_1, a_2] \left[ \frac{1}{b_2}, \frac{1}{b_1} \right]
\]

where

\[
k = \min \{(m(A) m(B)) - \alpha, \beta - m(A) m(B))
\]

3.1.7 Algebraic combination of fuzzy focal elements:

Let \( X_1 \) and \( X_2 \) be two variables whose values are represented by Dempster-Shafer structure with focal elements \( A_1, A_2, A_3, \ldots, A_n \) and \( B_1, B_2, B_3, \ldots, B_m \) corresponding Basic Probability Assignments (BPA) are as follows:

\[
m(A_i) = a_i \quad \text{and} \quad m(B_j) = b_j, \quad i = 1, 2, 3, \ldots, n \quad \text{and} \quad j = 1, 2, 3, \ldots, m \quad \text{respectively}.
\]

where \( \sum_{i=1}^n a_i = 1 \) and \( \sum_{j=1}^m b_j = 1 \)

Initially, we combine all the fuzzy focal elements using fuzzy arithmetic which will produce several fuzzy focal elements and thereafter the corresponding basic probability assignments of resulting fuzzy focal elements will be calculated as follows.

| Parameters          | River Water | Pond Water | Bore Water | Municipality Water |
|---------------------|-------------|------------|------------|--------------------|
|                     | FFE | BPA | FFE | BPA | FFE | BPA | FFE | BPA |
| PH                  | [6.8] | 6.7 | [6.8] | 6.8 | [6.8] | 6.5 | [6.8] | 6.6 |
| Temperature         | [31,33] | 32.8 | [31,33] | 32.9 | [31,33] | 32.5 | [31,33] | 32.6 |
| Nephelometric Turbidity(NTU) | [0.2] | 0.6 | [0.2] | 0.65 | [0.2] | 0.5 | [0.2] | 0.52 |
| Conductivity(mg/l)  | [0.6] | 0.52 | [0.6] | 0.54 | [0.6] | 0.416 | [0.6] | 0.423 |
| Alkalinity(mg/l)    | [0.1,0.3] | 0.3 | [0.1,0.3] | 0.294 | [0.1,0.3] | 0.256 | [0.1,0.3] | 0.258 |
| BOD(mg/l)           | [0.0,0.1] | 0.02 | [0.0,0.1] | 0.01 | [0.0,0.1] | 0 | [0.0,0.1] | 0 |

The result obtained quality of water are discussed detailed and the quality of water and characterize by using methods are pH is measured by HANNA digital pH meter, dissolved
oxygen is measured by BOD bottles method, Alkalinity is measured by titration method and the conductivity is measured by digital mobile conductivity probe method. The experimental results have been correlated using fuzzy evidence theory and good agreement was obtained from bore water is the best among the experimental data and theoretical procedure.

4. ANOVA

To determine the quantitative formation characteristics of the composite, the Analysis of Variance (ANOVA) was active to identify the level of importance in interaction parameters on their performance characteristics. The effect of these parameters on interaction has been examined using experimental designs and the variation of this interaction parameter with the mole fraction of composite has been discussed in terms of molecular interaction. A new software verification technique called Design Expert 7.0 was used to reveal the validity of the observed values developed and the results were discussed below in tables. It reveals that the linear model is the best-suggested model. So, for further analysis, this model was used. It also provides the ANOVA results for the Response Surface Linear Model of the input parameters and is commonly used to summarize the test for meaning on individual model coefficients.

| Parameters                          | River water | Pond water | Bore water | Municipality water |
|-------------------------------------|-------------|------------|------------|-------------------|
| pH                                  | 7.32        | 7.88       | 6.74       | 6.91              |
| Temperature                         | 32.9        | 32.7       | 32.8       | 32.8              |
| Nephelometric Turbidity Units (NTU) | 0.12        | 1.51       | 0.59       | 0.14              |
| Conductivity                        | 0.4511      | 0.5349     | 0.429      | 0.2983            |
| Alkalinity                          | 0.2665      | 0.2665     | 0.2665     | 0.1999            |
| BOD                                 | 0.001       | 0.0005     | 0          | 0                 |

**Table 2: Analysis results report**

**Table 3: The types of water using six types of parameters types of water**

| Parameters                          | A            | B            | C            | D            |
|-------------------------------------|--------------|--------------|--------------|--------------|
| pH                                  | River water  | Pond water   | Bore water   | Municipality water |
| Temperature                         | River water  | Pond water   | Bore water   | Municipality water |
| Nephelometric Turbidity Units (NTU) | River water  | Pond water   | Bore water   | Municipality water |
| Conductivity (mg/l)                 | River water  | Pond water   | Bore water   | Municipality water |
| Types water | Alkalinity (mg/l) | River water | Pond water | Bore water | Municipality water |
|-------------|------------------|-------------|------------|------------|-------------------|
| pH          | 7.32             | 7.8         | 6.74       | 6.91       | 28.85             |
| Temperature | 32.9             | 32.7        | 32.8       | 32.8       | 131.2             |
| Nephelometric Turbidity Units (NTU) | 0.12 | 1.5 | 0.59 | 0.14 | 2.36 |
| Conductivity | 0.4511 | 0.5349 | 0.429 | 0.2983 | 1.69 |
| Alkalinity | 0.2665 | 0.2665 | 0.2665 | 0.1999 | 0.99 |
| BOD (mg/l) | 0.001 | 0.0005 | 0.0015 | 0.000002 | 0.000012 |

Table: 4 Level of Significance $\alpha = 5\%$

| Source of variance | Sum of squares | Degrees of freedom | Mean square | Variance ratio | Tabulate d F |
|--------------------|----------------|--------------------|-------------|----------------|--------------|
| Between Rows       | $Q_1 = 0.98$   | $h-1 = 6-1 = 5$    | $M_1 = Q_1/h-1$ | $F_{CAL1} = M_1/M_3 = 1.46$ | $F(5,15) = 2.90$ |
| Between Columns    | $Q_2 = 0.98$   | $K=1=4-1=3$       | $M_2 = Q_2/K-1$ |                |              |
| Residual variation | $Q_3 = 2.01$   | $(h-1)(k-1) = (5)(3) = 15$ | $M_3 = Q_3/(H1)(k-1)$ | $F_{CAL2} = M_2/M_3 = 1.82$ | $F(3,15) = 3.29$ |

Table: 5 ANOVA

5. Algorithm
Step: 1 Start the Program
Step: 2 Declare and initialize the variable parameters pH, temperature, Nephelometric
Turbidity units (NTU), conductivity, Alkalinity, BOD

Step: 3 Input file containing metrological data and dimensions related to study the types of water using six types of parameters.
- River water
- Pond water
- Bore water
- Municipality water

Step: 4 Define the Null hypothesis Ho
- There is no significant difference between rows.
- There is no significant difference between columns.

Step: 5 Set the level of significance $\alpha=5\%$ or $1\%$.

Step: 6 Calculate the total number of observations $N$.

Step: 7 Find the sums (Perform the calculations)

Sum of all observations is also calculated

\[ T = \sum' \sum' X_{ij} \]
\[ Q = \text{Total sum of squares} \]
\[ Q_1 = \text{Sum of squares of variation in rows} \]

Step: 8 The residual variation is calculated
\[ Q_3 = Q - Q_1 - Q_2 \]

Step: 9 Calculate the degrees of freedom
\[ \text{[h-1, k-1, (h-1)(k-1)]} \]
h – No of row
k – No of columns

Step: 10 Output the variables

Prepare the ANOVA table

Step: 11 End of the Program

\[ T=\sum' \sum' X_{ij} \]
\[ Q=\sum' \sum' X_{ij}^2 - T^2/N \]
\[ Q_1=\sum Ti^2/K - T^2/N \]
\[ Q_2=\sum Ti^2/K - T^2/K \]

6. Conclusion

Statistical analysis of variance (ANOVA) is used by the method of Random Block Design (RBD) to analyze the water samples. There is no major difference between the four samples (i.e Borewell, Pond, River, and Municipality). According to the pH level, the bore well water is pure in comparison to the above four samples. So, it is concluded that the best drinking water is bore well water. An algorithm was developed for the above model. The experimental results have been correlated using fuzzy evidence theory and good agreement was obtained from bore water is the best among the experimental data and theoretical procedure. In this work, we presented a fuzzy model which has more knowledge representing power. The salient aspect of this work is that we achieved a simplified and analytic fuzzy model when each focal element, a fuzzy subset in the fuzzy partition of the output space, has the same power.
7. Reference
[1] Combating Waterborne Diseases at the Household Level World Health Organization. 2007. Part 1. ISBN 978-92-4-159522-3.
[2] Water for Life: Making it Happen World Health Organization and UNICEF, 2005. ISBN 978-92-4-156293-5.
[3] Haze technical definition Archived August 22, 2015, at the Wayback Machine.
[4] K. Ganesan and P. Veeramani International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2005, Vol. 13, No. 06, pp. 619-631.
[5] Carrera, J. and Neuman, S. (1986). “Estimation of aquifer parameters under transient and steady-state conditions: 1. Maximum likelihood method incorporating prior information.” Water Resources Research WRERAO, Vol. 22, No. 2, pp. 199-210.
[6] Keith Beven, Jim Freer, Journal of Hydrology, 249 (2001) 11-29.
[7] Benjamin Renard, Dmitri Kavetski, George Kuczera, Mark Thyer, Stewart W. Franks, water resource research, 2010, vol. 46. Issue 5.
[8] Elizabeth H Keating, John Doherty, Jasper Vrugt, Qinjun Kang, Water Resources Research , 2010.
[9] F. Pappenberger, K. J. Beven, Water Resources Research , 2006, vol. 42. Issue 5.
[10] R. Irene Hepzibah, C. Ramachandra Raja, R. Ida Malarselvi, R. Priscilla, International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.51, 2015.
[11] Dempster, A.P.: A generalization of Bayesian inference. J. Royal Stat. Soc. (1968) 205–247.
[12] Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ (1976)
[13] Dempster, A.P.: Upper and lower probabilities induced by a multi-valued mapping. Ann. Mathematical Statistics 38 (1967) 325–339.
[14] Tang, Y., Sun S., Liu, Y.: Conditional Evidence Theory and its Application in Knowledge Discovery, Lecture Notes in Computer Science 3007 (2004), 500–505.
[15] Jiacheng Zheng and Yongchuan Tang, AI 2004: Advances in Artificial Intelligence, 816-827.