

## Normal Optimization Technique for Hydrological Data Sets under Fuzzy Environment

M. Sunitha<sup>1</sup>, B. Polaiiah<sup>2</sup>, S. K. Khadar Babu<sup>3\*</sup>, K. Pushpanjali<sup>4</sup> and M. V. Ramanaiah<sup>5</sup>

<sup>1</sup>Senior Bio-Statistical Officer, Sri Venkateswara Institute of Medical Sciences, Tirupathi

<sup>2,4</sup>Sri Krishnadevaraya University, Ananthapur, Andhra Pradesh, India

<sup>3\*</sup> VIT, Vellore, Tamil Nadu, India

<sup>5</sup>Sri Venkateswara University, Tirupathi-517502, Andhra Pradesh, India

\*Corresponding Author E-mail: [khadar.babu36@gmail.com](mailto:khadar.babu36@gmail.com)

Received: 16.10.2019 | Revised: 20.11.2019 | Accepted: 27.11.2019

### ABSTRACT

*At Present, the challenging aspect in research is mainly observed that to estimate the pre-defined well-versed parameters for a specified data sets. A Novel approach in research methodology is to study and analyse the parameters of the big data sets in nature. Now-a-days the data availability is in high dimensional in the sense of size aspect only the main source. The present research paper is proposed to estimate the pre-defined parameters by using the robust statistical techniques under the fuzzy environment only. Fuzzy statistical approach is only the source to analyse the different parameters for crisp data. In fuzzy approach, the data totally transformed under three different aspects like “happen, Most likely happen and rarely happen” cases for easy analyse and optimize the normal probability distribution and provide significant techniques in statistics.*

**Keywords:** Normal Probability Function, Fuzzy statistics, Crisp data analysis etc.,

### INTRODUCTION

Normal Probability distribution plays an important role in probability theory and statistical analysis. Now-a-days, it should be analysing different fields like medial and natural sciences.

In research study, we need to develop the different mathematical models for analysing the some specified parameters of the natural hazards like floods, storms, Volcano blasts etc. Actually for happy living in the society, bio-organ needs some standard living

comman facilities given by the nature is air, water and good atmosphere. These are an important components for happy and healthy living in the society. The one of the major aspects is how to transform the entire mathematical thinking and methods to express in terms human approach.

In Mathematical Sciences, Fuzzy is the best conventional language to explain the methods into the human thinkings. The fuzziness is the essential word language to express rich emotions.

**Cite this article:** Sunitha, M., Polaiiah, B., Khadar Babu, S.K., Pushpanjali, K., & Ramanaiah, M.V. (2019). Normal Optimization Technique for Hydrological Data Sets under Fuzzy Environment, *Ind. J. Pure App. Biosci.* 7(6), 264-269. doi: <http://dx.doi.org/10.18782/2582-2845.7865>

In mathematical statistics, “Happen”, “Most likely happen” or “rarely happen” are the vague statements and we can use the membership function to segregate the entire data into the conventional statements like the above and fix them into the fitting of the normal distribution.

Stein (1981), proposed how to use the sum of squared errors as loss and the estimation of the ways of normal in dependent random variables is advised, to consider for estimation of the parameters. Mikl (2007) applied normal distribution for statistics on the Stirling Permutation defined by Gessel and Stanley, he equidistributed statistics on these objects converge to a normal distribution (Sullivan et al., 1994). Even recently, CFF Karney published an algorithm for sampling from the normal distribution exactly, whose parameters are rational numbers, thereafter uniform random digit are copied into the representation directly (Khashei & Bijari, (2011). Ever since the built of fuzzy maths, it's application with calculation algorithm has been prospered (Song & Chissom, 1993; Srinivasan et al., 1994; Singh, 2007; Nayak et al., 2004; Rojas et al., 2004; Chen et al., 2014).

Traditional normal distribution can only calculate the determined indexes, such as “40 degree Celsius” or “120 pounds” or “45 years old”. But the uncertain concepts of data which can be represented by fuzzy number, like “appropriate temperature”, “normal weight” or “middle-aged”, also need to be count. So in this paper, we provide a model to calculate the fuzzy concepts probability by the integral of normal distribution probability density function.

### Review of Literature

Jeo Sullivan et al. (1994), introduced two methods, a first-order time-invariant fuzzy time series model and a first order time variant model. These are compared with autoregressive models, all of which are time invariant. Song and Chissom (1993), Based on the technology of Zadeh, studied about the fuzzy forecasting models with time variant and time invariant and also some special properties of explored in his article. Srinivasan et al.

(1994), proposed new approach for electrical load forecasting using fuzzy logics and neural networks. In this article, the proposed model analysis comprises the load forecasts annually. (Ahmed et al. (2008). In this article, He proposed fuzzy metric as a trend predictor for forecasting TAIEX and enrollments' forecasting of the University of Alabama. Singh (2007), studied for the fuzzy models to forecast different type of data sets. A new method of fuzzy time series forecasting based on difference parameters. Further, the proposed method has also been implemented on a real life problem of crop production forecast of wheat crop and the results have been compared with other methods. Chen (2012), proposed a Least squares support vector fuzzy regression model for time series analysis. In this article, no of equations derived by using the least squares principle. It was developed by Legendre for making different equations using partial derivatives with respect to the identified parameters. Then we get the equations, that should equal to the number of parameters. This technique is useful to make equations to obtain the possible parameters by solving the derived equations. Nayak et al. (2004), derived different types of artificial neural networks and fuzzy logic approaches and they applied on different variety of problems. In this study, they combined neural networks and fuzzy logics and made a neuro-fuzzy computing techniques for various problems. This paper presents the application of an adaptive neuro fuzzy inference system (ANFIS) to hydrologic time series modeling, and is illustrated by an application to model the river flow of Baitarani River in Orissa state, India. Khashei et al. (2011), proposed different methods to generate the data by using ANN and ARIMA models for time series analysis. Rojas et al. (2002), proposed a RBF neural network model under radial basis for time series prediction.

### Study Area

The reservoir has a capacity of 7,321,000,000 cu ft (207,300,000 m<sup>3</sup>) with a full level of 119 ft (36 m). An area of 7,185 ha

(17,750 acres) of land is benefited by the left bank canal and 100 ha (250 acres) of land is

benefited by the right bank canal Thandrampat and Thiruvannamalai blocks.



Fig. 1

**METHODS AND DISCUSSIONS**

The traditional normal distribution formula is showed below:

$$f(x) = \frac{1}{S\sqrt{2\pi}} e^{(-\frac{(x-M)^2}{2S^2})}, -\alpha < x < \alpha$$

Where, M is the mean and S is the standard deviation. Then the standard normal variate is by transforming the x into Z is in the following formula.

$Z = \frac{x-M}{S}$ , where Z follows normal distribution with mean zero and variance unity. Fuzzy variable is a number it takes the values from 0 to 1, which is determined by using the following formula used by the Khadar Babu(2018) et al in his paper.

For membership function deviated the entire data converted into three internal and

that are fitted in at happen, most likely happen and rarely happen. It reduces the entire data into the three intervals for events only. For this process, we adopt the box-plot diagram to segregate the possible intervals. First calculate the all quartiles for the data collected from pre-specified study area and adopt the following terminology for membership function.

$$\begin{aligned} \mu_{A1}(X) &= (X_{min} - X_{m1}) \text{ if } min < X_k < X_{m1} \\ \mu_{A2}(X) &= (X_{m1} - X_{m3}) \text{ if } X_{m1} < X_k < X_{m3} \\ \mu_{A3}(X) &= (X_{m3} - X_{max}) \text{ if } X_{m3} < X_k < X_{max} \end{aligned}$$

The above membership function gives the values located between two intervals and get the following fuzzy statistical transformation

$$H(x) = \begin{cases} \frac{x-l}{m-l} = \frac{x-l}{\alpha}, & x \in (l, m) \\ 0, & \text{otherwise} \end{cases}$$

Standard normal form

$$H(Z) = \begin{cases} \frac{x-\beta}{m-l} = \frac{x-\beta}{\alpha}, & x \in (l, m), \text{ mean, } \beta \text{ and sd, } \alpha \\ 0, & \text{otherwise} \end{cases}$$

For the above literature, we get the following statistical analysis.

**Table 3.1: Fuzzy transformed data**

Data, $Y_i$	Standardised data, $Z_i$	Fuzzy standardised data, $Z_i$
79.175	2.746	1
78.25	2.319	0.8815
77.17	1.8201	0.7432
76.32	1.4273	0.6343
75.58	1.0854	0.5395
74.83	0.7388	0.4435
74.05	0.3783	0.3436
73.17	0.0283	0.2309
72.00	0.5690	0.0810
71.50	0.8000	0.0170
71.50	0.8000	0.0170
71.50	0.8000	0.0170
71.50	0.8000	0.0170
71.50	0.8000	0.0170
71.50	0.8000	0.0170
71.45	0.8231	0.0106
71.45	0.8231	0.0106
71.45	0.8231	0.0106
71.40	0.8462	0.0042
71.40	0.8462	0.0042
71.40	0.8462	0.0042
71.367	0.8615	0

**Table 3.2: Standard fuzzy normally distributed statistical Analysis**

Data bounds	Lower extreme flows	Occurances	Fuzzy transform	$\Phi(Z)$	Normal probabilities	Estimated Occurances
Below zero	$-\alpha$	0	0	0	0	0
0 – 0.0170	0	7	-0.6002	0.2676	0.2912	7
0.0170- 0.4435	0.0170	8	-0.5576	0.2912	0.4038	8
0.4435- 1	0.4435	6	0.5117	0.6950	0.2976	6
Above 1	1	0	-	0.9926	-	

The probability of happen scores should based on the proposed model is 0.2912 and its expected occurrences goes to 7. It is absolutely equal to the observed occurrences happen.

The probability of most likely happen is 0.4038 and the rare happens is 0.2976. Its corresponding expected occurrences are 8 and 6 respectively. These three cases also conformed to reality. For all these three cases, the results will be changed if we take the differences by taking the S, 2S and 3S limits. S is the standard deviation of the present data set. In place of quartile data, we can consider 3S approach for finding the probabilities using the normal probability distribution.

### CONCLUSIONS

In this paper, we design a normal probability density function under the fuzzy environment. So that, the distribution is very closer to the real life situations. Here, three distributed events are “happen”, “most likely happen” and “rarely happen” were provided and the results for distribution is also logical. In some situations, we find the probabilities of events, apply different view points for comparison of their advantages and disadvantages. The present model is applicable for big data analytics and also useful to compress the cohort into a small data set, which can be applied very wide range of applications in the field of artificial intelligence.

### REFERENCES

- Chen, Y. (2012). Least Squares Support Vector Fuzzy Regression, *Energy Procedia* 17, 711-716
- Sullivan, J., William, H., & Woodall, (1994). A comparison of fuzzy forecasting and Markov Modelling, *Fuzzy sets and systems*, 64(3), 279-293.
- Khashei, M., & Bijari, M. (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied soft computing*, 11(2), 2664-2675.
- Song, Q., & Chissom, B.S. (1993). Fuzzy timeseries and its models, *Fuzzy sets and systems*, 54(3), 269-277.
- Srinivasan, D., Liew, A.C., & Chang, C.S. (1994), Forecasting daily load curves using a hybrid fuzzy-neural approach, *IEEE proceedings Generation, Transmission and Distribution*, 141(6), 561-567.
- Singh, S.R. (2007). A simple method of forecasting based on fuzzy time series, *Applied Mathematics and computation*, 186(1), 330-339.
- Nayak, P.C., Sudheer, K.P., Rangan, D.M., & Ramasastri, K.S. (2004). A neuro-fuzzy computing technique for modeling hydrological time series, *Journal of hydrology*, 291(1-2), 52-66.
- Rojas, I.H., Pomares, J.L., Bernier, J., Ortega, B., Pino, F.J., Pelayo, & Prieto, A. (2002). Time series analysis using normalized PG-RBF network with regression weights, *Neurocomputing*, 42(1-4), 267-285.
- Chen, J., Yu, J., & Zhang, Y. (2014). Multivariate video analysis and Gaussian Process Regression model based soft sensor for online estimation and prediction of nickel pellet size distributions, *Computers and chemical Engineering* 64, 13-23.
- Bertone, E., Kelvin, o'Halloran, Rodney, A., Stewart, Guilherme, F.de oliveira, (2017). Medium-term storage volume prediction for optimum reservoir management: A hybrid data-driven approach, *Journal of cleaner production*, 154(15), 353-365.
- Huang, S.C., & Wu, T.K. (2008). Combining Wavelet-based feature extractions with relevance vector machines for stock index forecasting. *Expert Systems* 25, 133-149.
- Prathumratana, L., Sthiannopkao, S., & Kim, K.W. (2008). The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River, *Environment International*, 34, 860-866.
- Nearing, G.S., Gupta, H.V., & Crow, W.T. (2014). Information loss in approximately Bayesian estimation

- Sunitha et al.** *Ind. J. Pure App. Biosci.* (2019) 7(6), 264-269 ISSN: 2582 – 2845
- techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity. *Journal of Hydrology*, 507(12), 163-173.
- Ruiz & Benefa (2012). Modelling facial expressions dynamics with Gaussian process regression. *Frontiers in Artificial intelligence and Applications*. 248, 91-100.
- Khadar Babu, S.K., Karthikeyan, K., Ramanaih, M.V., & Ramanah, D. (2011). Prediction of rain-fall flow time series using auto-regressive models, Pelagia research library, *Advances in applied science research*, 2(2), 128-133.
- Khadar Babu, S.K., Sunitha, M., & Ramanaih, M.V. (2017). Mathematical Modelling of RMSE Approach on Agricultural and Financial Data Sets, *International journal of Pure and Applied Bio-sciences*, 5(6), 942-947.