VALIDATION OF LOG LOGISTIC DISTRIBUTION TO MODEL WATER DEMAND USING UK AND NORTH AMERICAN DATA

Seevali Surendran[1,2*], Kiran Tota-Maharaj[1] , Hua-Peng Chen[1]

[1]University of Greenwich, Faculty of Engineering & Science, Department of Civil Engineering, Medway Campus, Kent, ME4 4TB, United Kingdom

[2]Environment Agency, Kings Meadow house, Kings Meadow Road, Reading, RG1 8DQ, United Kingdom [1]seevali.surendran@environment-agency.gov.uk

[2]k.tota-maharaj@greenwich.ac.uk

[3]h.chen@greenwich.ac.uk

Abstract

There is a general assumption that potable water consumption is stochastic in nature with increasing in recognition of the need to allow for this when planning, designing or assessing the performance of water distribution systems. The stochastic nature of water demand is best addressed by fitting water demand into a suitable probability distribution. It is often assumed that variations in water demands follow a normal distribution without adequate justification. However, there had been insufficient studies previously to conclude the suitability of the appropriate probability distribution functions in modelling water demand. The purpose of this study is to validate an appropriate probability density function to apply in simulating water demand using real water consumption data. Daily water consumption data for five years obtained from a water company in Canada and four years for UK are analysed using normal, log normal, log logistic and Weibull distributions and a comparison on the applicability of each distribution was assessed. Statistical modelling was performed using MINITAB. The Anderson Darling (AD) statistic was used as the goodness of fit parameter in the analysis.

Key words

Water demand, stochastic nature, probability distribution function, log logistic distribution, Minitab, Anderson Darling

1. INTRODUCTION

A major unresolved problem in the water consumption modelling is the identification of an appropriate statistical distribution which best represents the water consumption pattern. The aim of this research is to study real water consumption data and to find out a standard distribution to use in water demand modelling to address the probabilistic nature of water demand. Water companies are facing challenges to increase or manage water demand to cater government housing targets, water

* Tel.: +4401189533147

E-mail address: seevali.surendran@environment-agency.gov.uk

efficiency targets, population projections and technological changes. The advantage of modelling real water consumption data is that it will permit forecasting the probability of occurrence in any demand value and assist planners in future projections.

Water demand varies with time of use, season and socio economic pattern of the consumers and hence defined as a continuous random variable. Therefore incorporating probabilistic nature of demands in modelling will lead to more realistic assessments of the performance of water distribution systems. However, few studies can be found in which the random variations of demands have been considered.

Goulter and Bouchart (1990), Xu and Goulter (1997, 1998, 1999) made an assumption that the demands have a normal distribution. Mays (1994) used randomly generated water consumption data using a range of distributions to study the sensitivity of the system's performance to changes in water consumption patterns. Khomsi et al. (1996) stated that the demand is behaving as having a normal distribution based on the Kolmogorov-Smirnov test. Surendran and Tanyimboh (2002), Tanyimboh and Surendran (2004) addressed the issue of the modelling of short term demand variations in a comprehensive way using UK water demand data and concluded, water demand data fit well in to log logistic distribution than a normal distribution.

AWWA Research foundation sponsored a study (Bowen et al.1993) in residential water demand use patterns in USA results, revealed that the demand data was not distributed normally. Several data transformations to improve the data analysis were investigated and it was found that the log transformation was only mildly effective in reducing the positive skewness of the frequency distributions of the data, making them more nearly normal.

2. METHODOLOGY

Water companies in UK and Canada were approached to obtain the daily water consumption data. To analyse the UK water consumption pattern, daily water consumption data for 4 years from April 2009 to April 2013 were obtained from a UK water company to use in this research. The particular Water Works system delivers water to approximately 6.7million households and businesses in UK. Data were collected using data loggers at 15 minutes interval by the water company.

To analyse the Canadian daily water consumption data, data for 3 demand zones were obtained. Canadian Water Company collected the data using data loggers at the water treatment plant by the water services division. The data were obtained from 01.01.09 to 31.12.14. This Water supply system delivers an average of 225 million litres of water to approximately 270,000 households and businesses across approximately 297 square kilometres (114 square miles) of the developed portion of Canada.

In this research a suitable statistical distribution was selected using a series of applications. Data were screened and sorted by plotting raw demand data against time. This provided a quick reference to check the accuracy of data. If the points were homogeneously distributed and there were no negative points, this meant that the data is almost accurate. Similarly, if there was any inconsistently in distribution, this would allow removing all abnormal data points.

Following sorting out the data, the data were then analysed using MINITAB statistical package to fit into a probability distribution. Continuous distributions such as, normal, Log- normal, Weibull and log-logistic were applied to find a suitable distribution.

2.1 Analysing Data

Analysing univariate data (single column of data such as water demand) with a specific probability is one common application in modelling. Once data has been fitted in to any distribution, the goodness of fit method should be used to see how well the data will fit into the particular distribution. The appropriateness of the Log logistic distribution for water consumption data was assessed by comparison to the normal and lognormal distributions using the Anderson Darling goodness of fit parameter. The Anderson Darling (AD) statistical method is a measure of how far the plot points fall from the fitted line in a probability plot. A smaller AD value indicates that the distribution fits the data better. R Johnson (2000) stated that as a guide line, the large sample 5% point is 2.492 and the 1% point is 3.857 could be used to assess the data.

3. RESULTS AND DISCUSSION

3.1 The graphical method

There are various numerical and graphical methods used in estimating the parameters of a probability distribution. In this research, graphical methods were selected for the analysis along with the Maximum Likelihood method to draw the probability plots (Fig 1-8). The data were analysed using 95% confidence interval (5% significance level) and fitted to normal, lognormal, Weibull and log logistic distributions to establish the parameters of the particular distribution. The middle line in the probability plot shows the normal line and other two lines show the 95% confidence interval. Normal probability plots are useful in identifying distributions that actually fit to the normal distribution and the distributions which have tails heavier and lighter Montgomery and Runger (2002). The normal probability plots for the Canadian water consumption data shown on Figure 3. It shows that for Zone 2 and 3 data, the points on the left and right are above the normal line. This explains that the data has a heavy tailed distribution. Zone 2 data lies on the normal line. It was observed that by analysing water demand data, the water demand have a normal or heavy tail distribution.

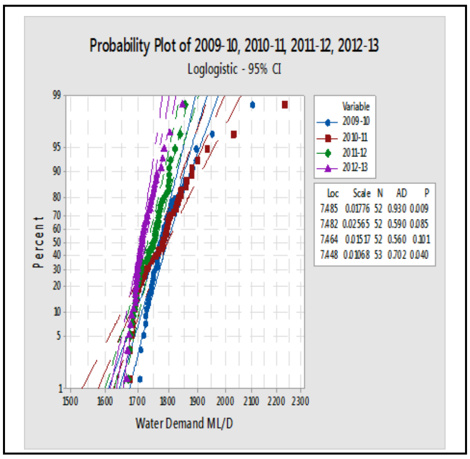

Fig. 1. Probability plots for Log logistic -UK

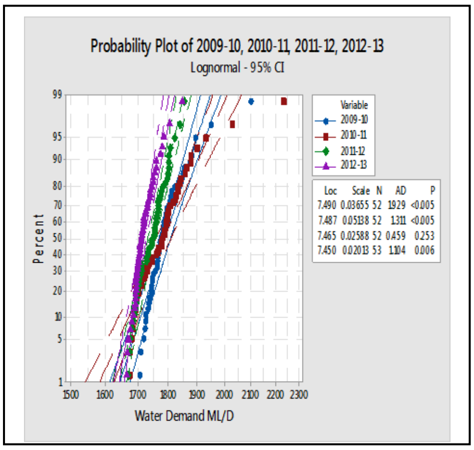

Fig. 2. Probability plots for Log normal-UK

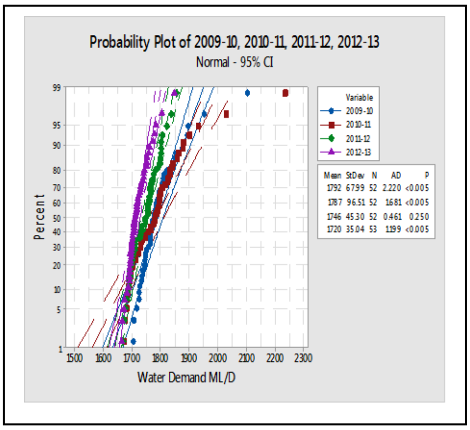

Fig. 3. Probability plots for Normal -UK

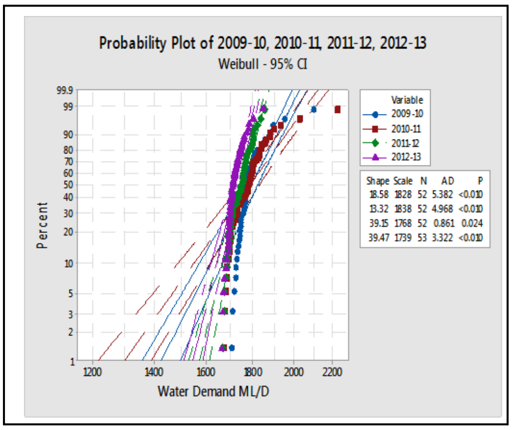

Fig. 4. Probability plots for Weibull-UK

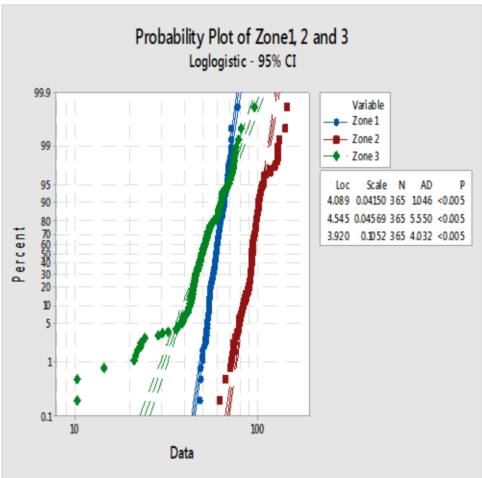

Fig. 5. Probability plots for Log logistic- Canada

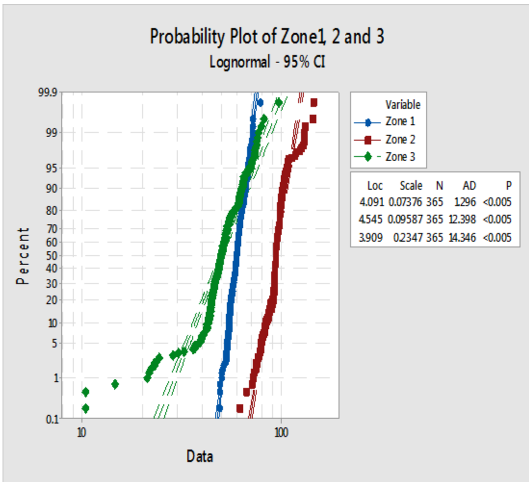

Fig. 6. Probability plots for Lognormal -Canada

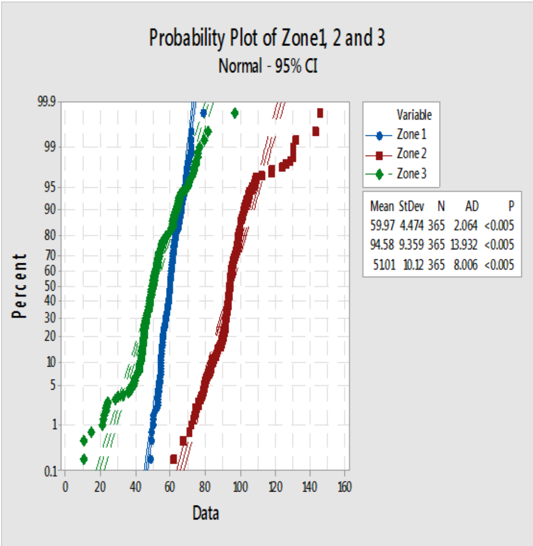

Fig. 7. Probability plots for Normal- Canada

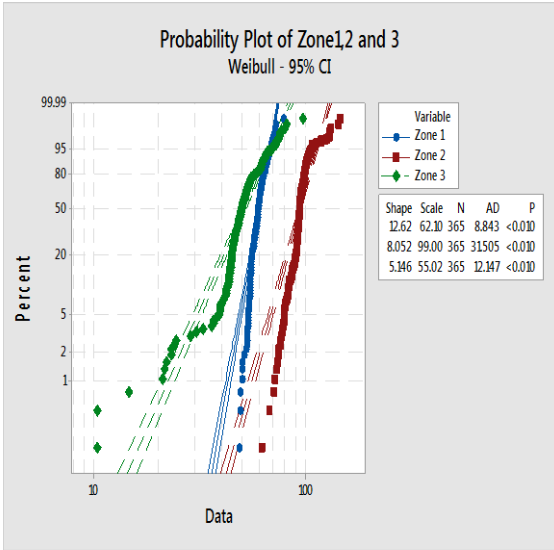

Fig. 8. Probability plots for Weibull-Canada

3.2 The goodness of fit

The goodness of fit used for this analysis is Anderson Darling test. AD values for normal, lognormal, Log logistic and Weibull distributions for UK and Canadian data are shown in Fig.9 to 12. It can be seen that Log logistic distribution has the lowest AD values when compared with the normal, Weibull and log-normal distribution.

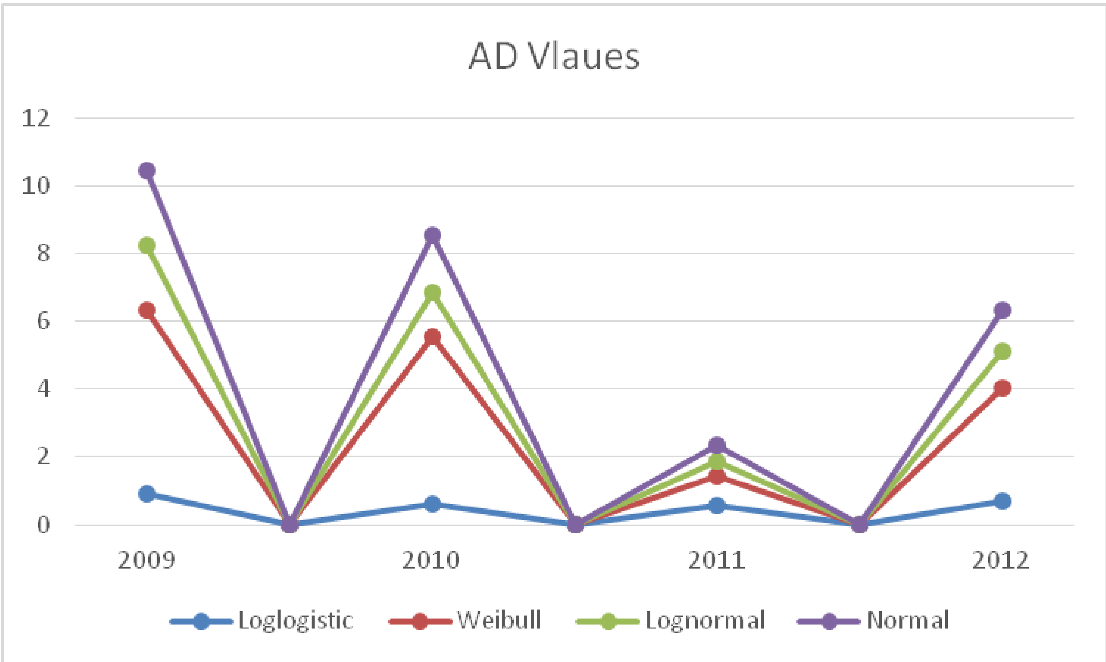

Figure 9:AD values -UK

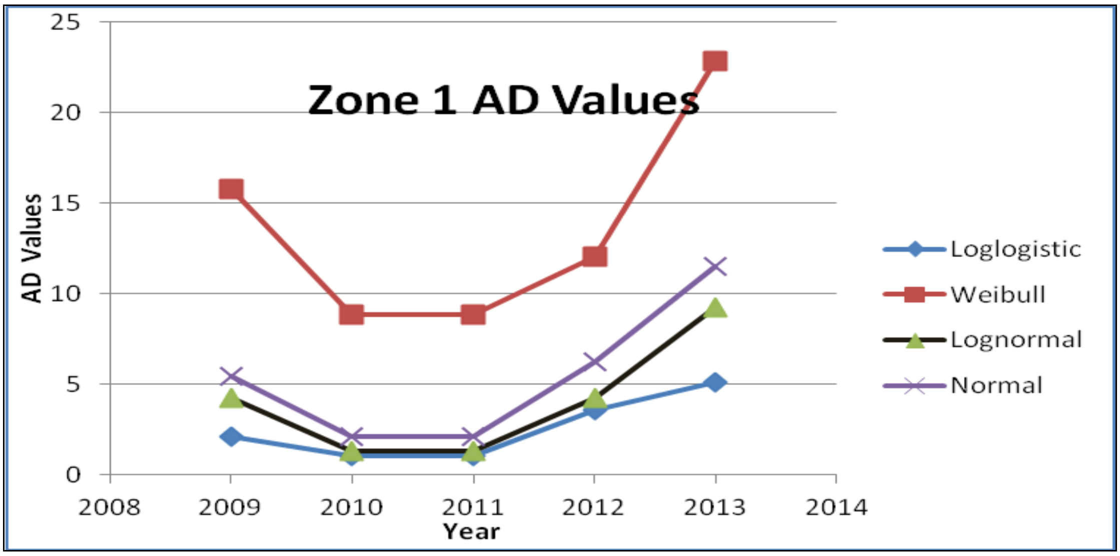

Fig. 10 AD values for Zone 1- Canada

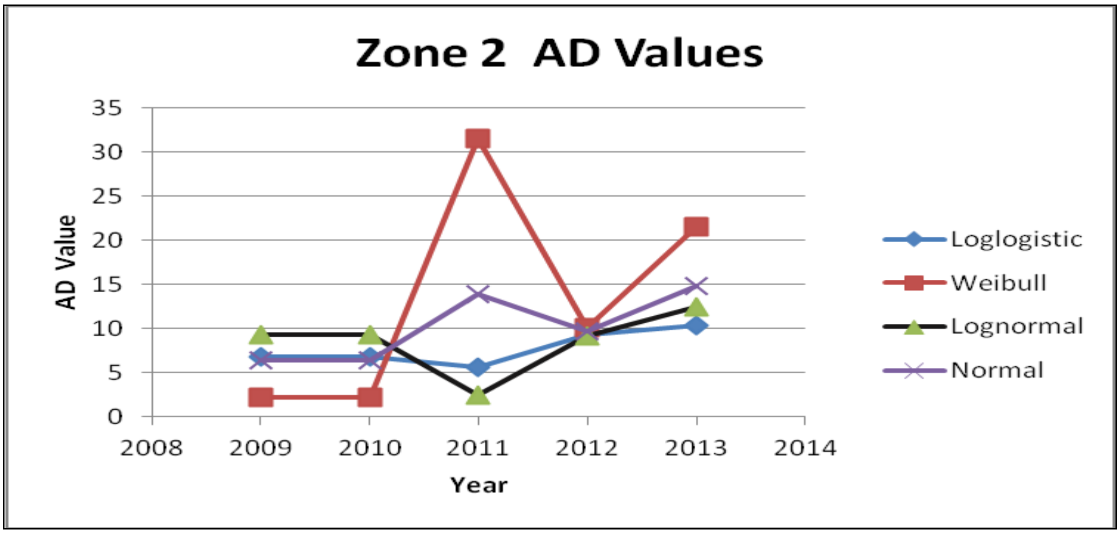

Fig: 11 AD values for Zone 2- Canada

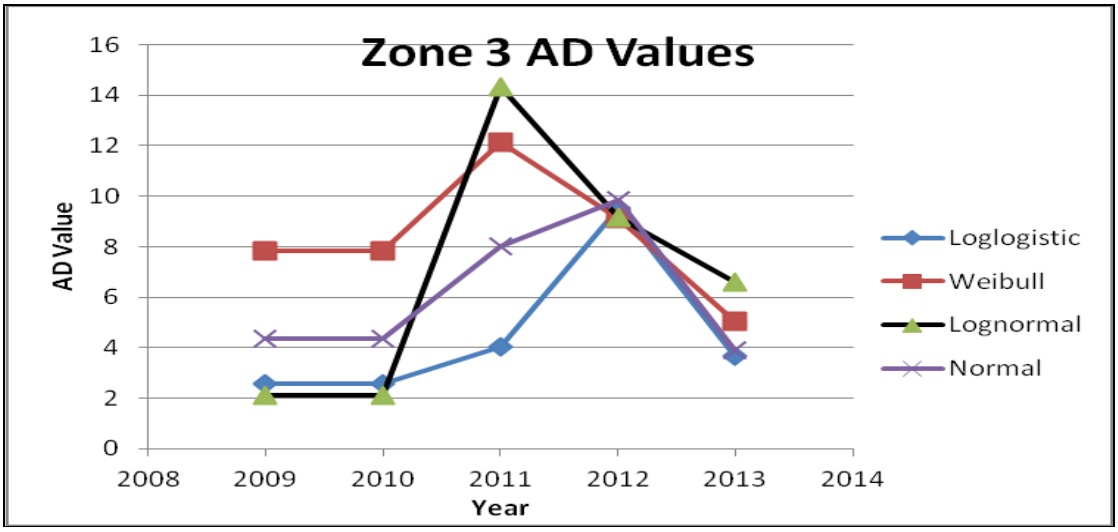

Fig: 12 AD values for Zone 3 - Canada

3.3 parameter estimates

The parameters of the particular distribution such as location, shape and scale are also essential to describe the distribution. The lowest location parameter obtained for log logistic is 7.448 and the highest figure is 7.485 for UK water consumption data. Similarly the lowest scale parameter is 0.0107 and the highest is 0.025 (Table 1).

For Canadian data the lowest location parameter obtained for log logistic is 3.92 and the highest figure is 4.562. Similarly the lowest scale parameter is 0.0296 and the highest is 0.389 (Table 2).

Parameters for Normal distribution are mean and standard deviation, though Lognormal, Log logistic and Weibull distributions, they are location, shape or Scale parameters. These parameters will allow the distribution to have a flexibility and effectiveness in modeling applications. Scale parameters allow a distribution to take on a variety of shapes depending on the value of the shape parameter. The effect of the location parameter is to simply shift the graph to left or right on the horizontal axis. The scale parameter describes the stretching capacity of the probability distribution function. If the scale parameter is greater than 1 then it will stretch the probability distribution function.

Table 1. Shape and Scale parameters for Log logistic distribution for UK data

	Location	Scale
2009 data	7.485	0.01776
2010 data	7.482	0.0256
2011 data	7.464	0.01517
2012 data	7.448	0.01068

Table 2. Shape and Scale parameters for Log logistic distribution for Canadian data

	Zone 1	Zone 2	Zone 3

	Location	Scale	Location	Scale	Location	Scale
2009 data	4.089	0.0415	4.545	0.0457	3.920	0.105
2010 data	4.089	0.0415	4.545	0.0457	3.920	0.105
2011 data	4.187	0.055	4.446	0.063	4.308	0.142
2012 data	4.098	0.389	4.361	0.041	4.152	0.092
2013 data	4.115	0.0296	4.562	0.054	3.992	0.088

4. CONCLUSION

The AD values obtained for the Weibull distribution has higher values than other 3 distributions and it is not suitable to model water demand data. The study shows that out of four distribution patterns studied, the log-logistic seems to have the lowest AD values and it was the most suitable distribution pattern to standardise when modelling water demand. However, normal and the log-normal distribution also have marginally acceptable AD values.

ACKNOWLEDGEMENTS

The author wishes to express the gratitude to the Environment Agency for their support, and UK and Canadian water companies to providing data for this study.

REFERENCES

[1] Bowen P T , Harp J F, Baxter W J and Shull R D, (1993). ‘Residential water use patterns’, American Water Works Association - Research Foundation, USA.

[2] Goulter I C and Boulchart F, (1990). ‘Reliability constrained pipe network model’, Journal of hydraulic engineering, ASCE, 116, (2), 211- 227.

[3] Johnson R A (2000). Preobability and statistics fro Engineers, Prentice Hall, London.

[4] Khomsi D, Walters G A, Thorly A R D and Ouazar D, (1996). ‘Reliability tester for water distribution networks’, Journal of Computing in Civil Engineering, ASCE, 10, (1), 11-19.[4]

[5] Mays L W, (1994). Computer Modelling of Free Surface and Pressurised flows, Chaudray M H and Mays L W (editors), Kluwer Academic Publishers, Netherland, 485-517.

[6] Montogomary D C and Runger G C (2002), ‘Applied statistics and probability for Engineers’, 3rd edition, Printed in USA.

[7] Surendran S and Tanyimboh T T, (2004). 'A Peaking Factor Based Statistical Approach to the Incorporation of Variations in Demands in the Reliability Analysis of Water Distribution Systems', Journal of Computers and Structures or Advances in Engineering Software, ASCE, 723-728.

[8] Tanyimboh T T and Surendran S, (2002). Log-logistic Distribution Models for Water Demands, 4th International Conference on Engineering and Technology Civil-Comp press, Sterling.

9] Xu C and Goulter I C, (1997). 'A New model for reliability based optimal design of water distribution networks,' Managing Water: Coping with The 27th Congress of the International Association for Hydraulic Research, ASCE, 423-428.

[10] Xu C and Goulter I C, (1998). 'Probabilistic model for distribution reliability', Journal of Water Resources Planning and Management, ASCE, 124, (4), 218-228.

[11] Xu C and Goulter I C, (1999). 'Reliability based optimal design of water distribution networks', Journal of Water Resources Planning and Management, ASCE, 125, (6), 352-362.