Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 798 – 802

Procedia Engineering

www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

Log logistic distribution to model water demand data

Seevali Surendran[1, 2]* and Kiran Tota-Maharaj[1]

[1]University of Greenwich, Faculty of Engineering & Science, Kent, ME4 4TB, United Kingdom.

[2]Environment Agency, Kings Meadow house, Kings Meadow Road, Reading, RG1 8DQ, United Kingdom

Abstract

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 find an appropriate probability density function to apply in simulating water demand using real water consumption data. Daily water consumption data for four years obtained from a water company in UK and 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.

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

1. Introduction

The aim of this research is to study real water consumption data and establish a standard statistical distribution to use in water demand modelling. Water demand data for 4 years were obtained from a water company in UK for this study.

* Corresponding author. Tel.: +4401189533147
E-mail address: seevali.surendran@environment-agency.gov.uk

doi:10.1016/j.proeng.2015.08.940

Water demand varies with time of use, season and socio economic pattern of the consumers. It has never been same at any particular time and hence defined as a continuous random variable. Therefore incorporating variation 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

Daily water consumption data for 4 years from a UK water company was obtained and analysed. The data was collected using data loggers at 15 minutes interval. The data was obtained from April 2009 to April 2013.

In this research model selection was done in a series of applications. Data were screened and sorted to select the model. Raw demand data was drawn against time. This provided a quick reference to check the accuracy of data. If the points were homogeneously distributed and there was no negative points, this meant that the data is almost accurate. Similarly if there was any inconsistently in distribution, this would allow us to remove all abnormal data points.

2.1 Data Analysis

Following the sorting out of data, the data was 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 distributions.

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 1-4). The data were analysed using 95% confidence interval (5% significant level). Once the data was fitted to normal, lognormal and log logistic distributions, parameters of the particular distribution such as location, shape and scale were essential to describe the distribution. Table 1 shows the parameter estimates for the obtained data.

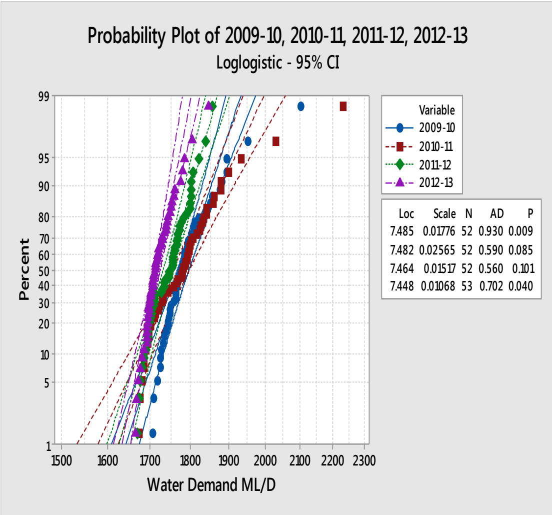

Fig. 1. Probability plots for Loglogistic

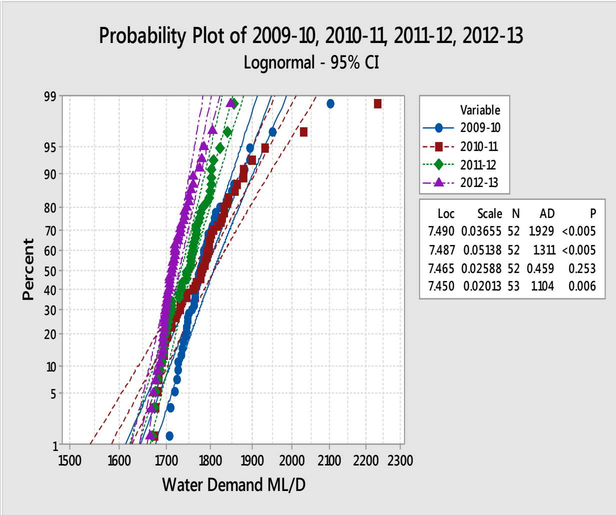

Fig. 2. Probability plots for Lognormal

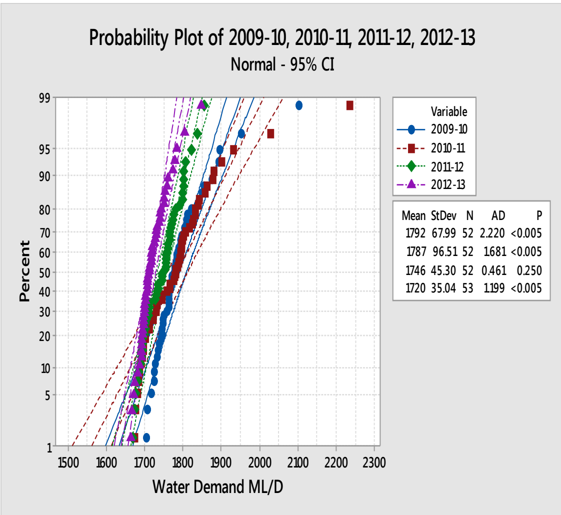

Fig. 3. Probability plots for Normal

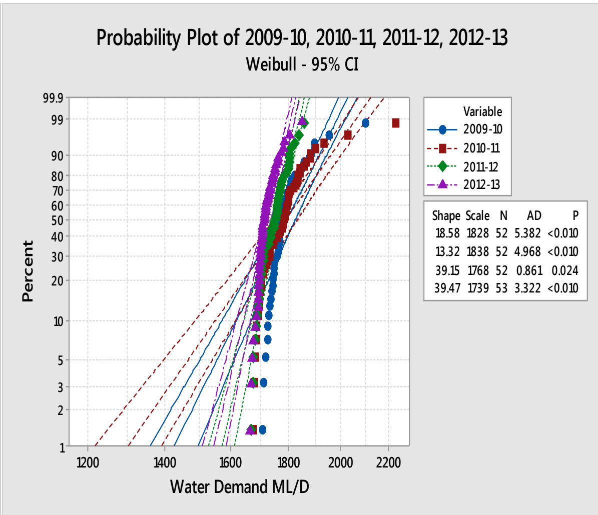

Fig. 4. Probability plots for Weibull

3. Results and discussion

Daily water consumption data from a UK were analysed. The Water Works system deliver water to approximately 6.7million households and businesses in UK.

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.

3.1 The goodness of fit method

The goodness of fit method used for this analysis is the Anderson Darling (AD) statistical method. 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 (1996) 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.

AD values for the data were obtained for normal, lognormal, Loglogistic and Weibull distributions and is shown in Fig.5. It can be seen that Loglogistic distribution has the lowest AD values when compared with the normal, Weibull and log-normal distribution.

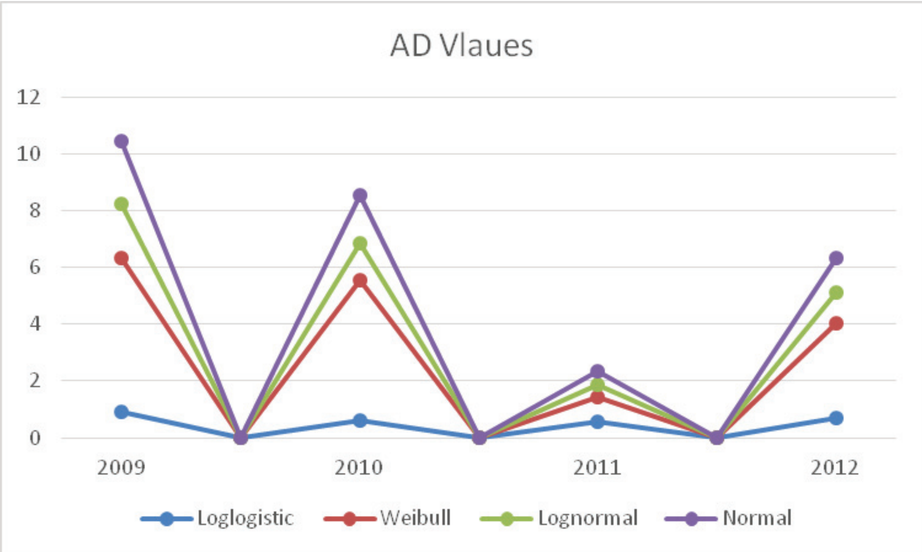

Figure 5: Graph representing loglogistic distribution has the lowest AD value

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. Similarly the lowest scale parameter is 0.0107 and the highest is 0.025.

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 shows the parameter estimates for the obtained data.

Table 1. Shape and Scale parameters for Loglogistic distribution obtained for 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

4. Conclusion

If water demand fits the normal distribution, then applying the mean value in designs would make no difference to the existing method. However, due to the positive extreme values in the demand data it is certain that the demand data would not fit the normal distribution.

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.

Also, findings from this study show that distribution patterns for UK is very similar to the studies completed in 2002 by Surendran and Tanyimboh. The general conclusion from these results is that normal, lognormal and loglogistic distributions would appear to be appropriate for the modelling of water demand, however log logistic distribution provided the best distribution out of the three and was used as a standard probability distribution model.

Acknowledgement

The authors are grateful for UK water company in providing the demand data and express the gratitude to the environment agency for their support in 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] 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]
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] Xu C and Goulter I C, (1998). 'Probabilistic model for distribution reliability', Journal of Water Resources Planning and Management, ASCE, 124, (4), 218-228.
[9] 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.