Global Journal of Environmental Sciences. Vol. 4, No. 2, 2005,: 151 - 154
ISSN 1596-6194. Copyright Bachudo Science Co. Ltd. Printed in Nigeria.

FITTING PROBABILITY DISTRIBUTIONS TO COMPONENT WATER QUALITY DATA FROM A TREATMENT PLANT

N. E. NWAIWU and APAGU BITRUS

(Received 2 August 2004; Revision accepted 1 November, 2004)

ABSTRACT

The treatment of water is carried out to make the available water meet the standards for its intended use. Such use may be for drinking and other household needs, industries,livestock rearing or fisheries etc. poor quality water is commonly treated to ensure potability. Potable water should be free from unpleasant tastes and odour, have a good appearance and be suitable for domestic as well as a wide range of industrial purposes. Continuous statistical distributions are usually applied to engineering situations. A goodness-of-fit test is usually necessary to determine the fitness of a distribution to specific data. The Kolmogorov-Smirnov test which is a widely used goodness-of-fit measure was used in the work. A total of four continuous distributions namely normal, log normal, gamma and weibuill were employed for data obtained from effluents of coagulation/clarifying process, (settled water), filtration (filtered water) as well as chlorination process (final water). For the settled water, colour and turbidity are lognormally distributed while pH is normally distributed; filtered water has colour, turbidity, free chlorine and total chlorine log normally distributed while pH has gamma distribution for best fit. The log normal distribution fitted all the parameters undertaken for the final water.

KEYWORDS: water, treatment plant, clarifier, filter, distributions.

INTRODUCTION

The treatment of water is required for a number of reasons of which the removal of germs and diseases is the most important. Potable water should be free from unpleasant tastes and odours, have a good appearance and be suitable for domestic as well as a wide range of industrial purposes. Basically the steps employed in water treatment include screening, plain sedimentation, flocculation with sedimentation, filtration, disinfection and miscellaneous tertiary treatment, where the need arises.

There are three main classes of treatment for potable water namely:
(i) the physical processes which depend essentially on physical properties of the impurity, for example, particle size, specific gravity and viscosity. (ii) Chemical processes which depend on the chemical properties of an impurity or which utilize the chemical properties of added reagents and (iii) biological processes. Physical processes include screening, sedimentation, filtration and gas transfer while the chemical processes include coagulation, precipitation and ion exchange. In addition common physical water quality parameters are colour, odour, turbidity, pH, electrical conductivity and temperature.

Continuous statistical distributions are usually applied to engineering situations (Benson, 1993; Benson and Charbeneau,1991; Bogardi, et al.,1989; Johnson, et al., 1990; Krápac, et al.1989). The choice of a distribution to represent any system or quality is verified using the available data. (Hahn and Shapiro, 1967).

METHODOLOGY

A total of four continuous distributions namely normal, lognormal, gamma and weibull were used to fit the data, and the goodness-of-fit was determined using Kolmogorov-Smirnov (K-S) test. The data (1994-2000) were obtained from the Maiduguri water treatment plant which supplies water to the metropolis. Maiduguri is the capital of Borno State, a former capital of the defunct North eastern State, Nigeria as well as major stop-over town on the trans Saharan trade route.

The parameters for which the data were obtained include colour, turbidity, pH, temperature turbidity, free and total chlorine. The collected data are for the effluent from the clarifier called the settled water; effluent from the filter called the filtered water and the effluent from the storage tank after the disinfections called the final water. A goodness-of-fit test is usually necessary to determine the fitness of a distribution to a specific data.

Monthly values of recorded water quality parameters were used to obtain the values for the descriptive statistics tables. Cognizance was taken of the months in which there were no records made. The descriptive statistics were accomplished for the settled filtered and final water quality parameters. The statistical distribution was carried out as follows on Microsoft Excel Worksheet. Any spread sheet having distribution functions may also be used.

(1) Specified monthly data were sorted serially from minimum value to maximum value (column 1).

(2) The data were numbered serially from number 1 in ascending order noting the maximum serial number ,n, which is the same as the total number of data values available (column 2).

(3) The values in column 2 were divided by n, which is the same as the total number of data values available(column 3). Column 3 has a maximum value of 1.

(4) Specified distribution values were obtained for values in column 1 using the computer spreadsheet(column 4).

(5) Values in column 3 were subtracted from values in column 4 (column 5).

(6)The maximum absolute difference (irrespective of the sign) was called the D calculated.

(7) The Dcalculated values were compared to the Dmax values from Kolmogorov-Smirnov Tables obtained from textbooks (e.g. Miller and Freud, 1977).When the Dcalculated exceeded Dmax, it was concluded that the chosen continuous distribution did not fit the specified data. Otherwise the distribution was considered as fitting the data.

(8) Where more than one distribution satisfied the above condition in (7), the continuous distribution with the lower or lowest Dcalculated value was chosen as the distribution of best fit.

N. E. NWAIWU, Department of Civil and Water and Resources Engineering, University of Maiduguri, Maiduguri, Nigeria.
APAGU BITRUS, Department of Civil and Water and Resources Engineering, University of Maiduguri, Maiduguri, Nigeria.

Table 1: Descriptive statistics of the settled water quality parameters

Statistical Functions	Colour (pt/co)	Turbidity (NTU)	pH
Mean	35.764	7.723	6.48
Standard deviation	36.91	6.86	0.55
Variance	1362.35	47.43	0.31
Kurtosis	30.78	27.32	5.51
Skewness	4.71	4.16	-1.83
Minimum value	4.6	0.32	4.05
Maximum Value	314.61	58.9	7.29
Coefficient Of variation	103.2%	89.18%	8.54%
Number of observation	62	62	62

Table 2: Description Statistics of the filtered water quality parameters

Statistical Functions	Colour (pt/co)	Parameters Turbidity (NTU)	pH	Free Chlorine(mg/l)
Mean	30.00	6.16	6.57	0.38
Standard deviation	35.20	5.9114	0.46	0.48
Variance	1239.2	34.95	0.21	0.23
Skewness	4.21	3.44	1.171	4.675
Kurtosis	20.84	14.364	6.03	228.79
Minimum value	5.79	0.22	5.6	0.06
Maximum Value	229.51	37.18	8.66	3.47
Coefficient of variation	117.3%	96%	6.97%	124.4%
Number of observation	62	62	62	62

Table 3 : Description statistics of the final water quality parameter

Statistical Functions	Colour (pt/co)	Parameters Turbidity (NTU)	pH	Temperature ^0C	Electrical conductivity	Total Cl_2	Free Cl_2
Mean	20.91	3.997	7.01	26.37	0.18	1.56	0.41
Standard deviation	17.86	2.676	0.541	2.353	0.0887	8.014	0.5136
Variance	318.98	7.17	0.29	5.5364	0.008	64.2.	0.26
Kurtosis	4.48	7.63	9.20	-0.90	0.68	72.60	50.35
Skewness	2.16	2.49	2.49	0.0002	4.80	8.51	6.55
Minimum value	1	1.27	6.12	22.6	0	0.06	0.052
Maximum Value	79.83	16.44	9.4	29.74.	0.78	69	4.39
Coeff. Of variation	85.4%	66.95%	7.72%	8.92%	50.23%	515%	123.97%
No of observation	66	72	73	72	11	73	74

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 shows the descriptive statistics for the settle d water. The mean colour value is 35.76 pt/co(Platinum/ Cobalt) with a standard variation of 36.91 resulting in a standard variation of 103.2%. The curve is positively skewed and platykurtic with an excess coefficient of 27.78. The minimum observed values is 4.6pt/co while the maximum is 314.61 pt/co. The mean turbidity value obtained is 7.72 NTU (Table 2). The standard deviation is 6.89NTU and the coefficient of variation is 89.186%. There is a wide variation in the data collected for turbidity as well as for colour. The curve is positively skewed with kurtosis value of 27.32. The observed values range between 0.32 NTU and 58.9NTU. The wide variations in the observed data results from the wide seasonal variations that characterize the Sudan/Sahel Savannah Semi - arid region where the treatment plant is sited. The mean pH recorded for the clarifier effluent is 6.479 and the standard deviation is 0.5532 and the coefficient of variation is 8.54%. This indicates minimal variation. The curve is platykurtic and negatively skewed.

In Table 2 can be found the descriptive statistics of the filtered water quality parameters. Colour, turbidity, pH and free chlorine have mean values of 30 (pt/co), 6.16(NTU), 6.57 and 0.38 (mg/l) respectively. The curves for all the above parameters are positively skewed and platykurtic. The pH values have the least variation with a coefficient of variation(C.v.) of 6.97%. The free chlorine varies most width C.v. of 124.4%, colour width 117.3% and turbidity having 96% variation among observed data.

Table 4: Statistical distribution of settled water quality parameters

Parameter	Distribution	Parameter	Dcalculated	Dmax (α= 0.05)	Best fitting
Colour	Normal	$\mu = x$ $\sigma = s$	0.19354		
	Log normal	μ = 3.3866 σ= 07678	0.087101	0.01727	*
	Gamma	α = 0.9389 β = 38.09	0.20101		
p.H	Normal	$\mu = x$ $\sigma = s$	0.08645	0.1727	*
	Log normal	μ= 1.8626 σ= 0.0927	0.10445		
	Gamma	α=137.19 β = 0.0472	0.09376		
Turbidity	Normal	$\mu = x$ $\sigma = s$	0.18307		
	Log normal	μ= 1.8364 σ= 0.8123	0.0929	0:1727	*
	Gamma	α=1.2578 β = 6.141	0.1579		

$\mu = x$= mean, $\sigma = s$= standard deviation.

Table 5: Statistical distribution of filtered water quality parameters

Parameter	Distribution	Parameter	Dcalculated	Dmax (α= 0.05)	Best fitting
Colour	Normal	$\mu = x$ $\sigma = s$	0.2650	0. 1727	
	Log normal	μ= 3.0986 σ= 0.6905	0.1134		*
	Gamma	α= 0.7263 β = 0.7263	0.2963		
p.H	Normal	$\mu = x$; $\sigma = s$	0.0972	0.1727	
	Log normal	μ= 1.8795 σ= 0.0682	0.0921		
	Gamma	α=204.11 β = 0.032	0.0812		*
Turbidity	Normal	$\mu = x$; $\sigma = s$	0.2603	0.1727	
	Log normal	μ= 1.535 σ= 0.7643	0.1129		*
	Gamma	α=1.08396 β= 5.6779	0.1690		
Free chlorine	Normal	$\mu = x$; $\sigma = s$	0.2327	0.1727	
	Log normal	μ= -1.3619 σ= 0.857	0.1118		
	Gamma	α=0.6459; β= 0.5937	0.2347		*
Total chlorine	Normal	$\mu = x$; $\sigma = s$	0.4903	0.1727	
	Log normal	μ= -0.5966 σ= 0.8649	0.1565		*
	Gamma	α=0.03768 β =41.284	0.784		

Table 6: Statistical distribution of final water quality parameter

Parameter	Distribution	Parameter		Dcalculated	Dmax (α= 0.05)	Best fitting
Colour	Normal	μ = x	σ =s	0.2126	0.167	
	Log normal	μ = 2.7618	σ =0.7635	0.0784		*
	Gamma	α=1.3707	β =15.25	0.1229		
Turbidity	Normal	μ = x	σ =s	0.2167	0.160	
	Log normal	μ = 1.2350	σ =0.5163	0.1483		*
	Gamma	α=2.2296	β =1.7927	0.1578		
Temperature	Normal	μ = x	σ =s	0.1915	0.160	
	Log normal	μ = 3.2687	σ =0.0897	0.1568		*
	Gamma	α=125.6	β =0.2100	0.1600		
p.H	Normal	μ = x	σ =s	0.2040	0.1895***	
	Log normal	μ = 1.9444	σ =0.0716	0.1813		*
	Gamma	α=168.025	β =0.0417	0.1967		
Free chlorine	Normal	μ = x	σ =s	0.2363	0.158	
	Log normal	μ = -1.1574	σ =0.6803	0.0534		*
	Gamma	α=0.6504	β =0.6370	0.2842		
Total chlorine	Normal	μ = x	σ =s	0.4903	0.1592	
	Log normal	μ = -0.5966	σ =0.8650	0.1565		*
	Gamma	α=0.0377	β =41.284	0.784		

*** at 90% confidence interval

The description statistics of the final water quality parameters is shown in Table 3. Colour has a mean value of 20.91 pt/co a standard deviation of 17.86 and a coefficient variation of 85.4%.The mean turbidity value is 4.0 NTU and a C.v. of 66.95%. The total chlorine has the highest C.v. value of 515%. This is expected as the amount of chlorine added to the water should depend on the bacterial load present. The free chlorine in the water depends on the amount of total chlorine added and the bacterial load present. The free chlorine has a C.v. of 123.97%, having a mean value of 0.41 and standard deviation, 0.51. The pH has mean value 7.01 and the

lowest standard deviation of 7.22%. The variation in the temperature values is 8.92%. This means that although there are wide variations in the atmospheric temperature the water temperatures does not take a direct cue from the atmosphere. The electrical conductivity has only eleven measurements recorded within the time stipulated. The mean value is 0.18, the standard deviation 0.09 and coefficient of variation of 50.23%.

Statistical Distributions

Settled water

Three distributions namely normal, log normal and gamma were used to fit the data collected for colour, pH and turbidity (see Table4). The log normal distribution, fitted the colour data best having a Dcalculated of 0.0871 which was lower than the critical table value Dmax value of 0.1727. The normal distribution, log normal and gamma distributions all fitted the pH data, being lower in Dcalculated values that the D max value 0.1727. Of the three, the normal distribution which had a Dcalculated of 0.0865, being the lowest of the Dcalculated values, fitted the data best. The log normal distribution fitted the turbidity data for the settled water best having a Dcalculated value of 0.0929. Therefore the data can be said to be log normally distributed. The log normal distribution is the model for a random variable whose logarithm follows a normal distribution with parameters μ and σ. In the log normal distribution μ and σ are scale and shape parameters respectively.

Filtered water

The data for colour of the filter effluent was log normally distributed having a Dcalculated 0.1134 which was less than Dmax of 0.1727 from the table of critical values with scale and shape parameters of 3.0986 and 0.6905 respectively. For the pH data, all of the three distributions used to fit the data namely Normal, log normal, and gamma, recorded Dcalculated values lower than the Dmax value(Table 5). Of these, the gamma distribution performed best having the lowest Dcalculated value of 0.0812. The gamma distribution was used to describe random variables bounded at one end. The curve was positively skewed and platykurtic nature with an excess coefficient of 0.029. The log normal distribution fitted the turbidity, free chlorine and total chlorine data best with respective Dcalculated values of 0.1129, 0.1179 and 0.1565.

Final water

The best fitting distribution for the following parameters of the final water studied which included colour, turbidity, temperature, free chlorine and total chlorine is the log normal distribution (see Table 6). For colour, both the log normal and gamma distributions fitted the data but the former was a better fit. This was also true for turbidity and temperature. None of the three distribution used could fit the pH data for the final water at 95% confidence interval. At 90% confidence interval the log normal distribution fitted the data with a Dmax of 0.18948 as against

CONCLUSION AND RECOMMENDATION

In order to effectively monitor the overall performance of a treatment plant, component data for individual processes are indispensable. In addition, the generation of confidence region using the distribution of best fit is essential as the existence of the confidence region will help the sanitary engineer effect changes in routine treatment methods to maintain consistent good water quality supply. The information contained in this paper can be used for the above purposes. Distribution of best fit is also a key parameter in reliability studies.

REFERENCES

Benson, C., 1993. Probability Distributions for hydraulic conductivity of compacted soil liners .Journal of Geotechnical Engineering. American Society of Civil Engineers.119 (3):471-486

Benson,C., and Charbeneau,R.,1991.Reliability Analysis for Time of Travel in Compacted Soil Liners. Proc. American Society of Civil Engineers, Geotech. Cong.1991, Boulder, Colorado, GSP, 28(1):456-467.

Bogardi,I., Kelly, W.E., and Nardossy, A.,1989. Reliability model for soil liners: Initial Design. Journal of Geotechnical Engineering. American Society of Civil Engineers. 115(5):658-669.

Hahn G. J. and Shapiro S. S., 1967. Statistical Models in Engineering. John Wiley and Sons Inc. New York 355 pp.

Johnson, G., Crumley, W.S., and Boutwell, G.P., 1990. Field Verification of Clay Liner Hydraulic Conductivity. Waste containment systems: Construction, Regulation and Performance. American Society of Civil Engineers ,226-245.

Kopac, I.G., Panno, S.V., Rehfeldt, K.R., Herzog, B.L., Hensel, B. R., and Cartwright, K. 1989. Hydraulic properties of an experimental soil liner: Preliminary result. Proc. 12th Ann. Madison Waste Conference. University of Wisconsin Extension, 395-411.

Miller, I. And Freund, J. E.,1977. Probability and Statistics for Engineers. Prentice-Hall, New Jersey.