DESIGNING OF SAMPLING PROGRAMMES FOR INDUSTRIAL EFFLUENT MONITORING

F. Ntuli, PK. Kuipa* and E. Muzenda
Department of Chemical Engineering, University of Johannesburg, P.O. Box 17011, Doornfontein, Johannesburg, South Africa (corresponding author phone +27-11-559-6002; fax +27-11-559 6430; email: fntuli@uj.ac.za)
*Chinhoyi University of Technology; P Bag 7724, Chinhoyi, Zimbabwe

ABSTRACT

Monitoring of effluent discharges from industrial establishments discharging directly into municipality sewers is one of the major water pollution control activities conducted by municipalities. For largely industrialised municipalities the task can be quite expensive and not effective if sampling programmes are not properly designed. In most cases samples are randomly collected without proper knowledge of the discharge patterns of various industries. As a result the information obtained does not give a good reflection of the quality of effluent being discharged. These problems can be resolved by adapting a statistical approach to the design of sampling programmes. This approach is useful in determining the frequency of sampling, the number of samples needed to estimate the average concentration of target pollution indicator parameters and the magnitude of the uncertainty involved. The benefits and applications of this approach are demonstrated by a case study presented in this paper. It was found that the number of samples and cost of sample analysis can be greatly reduced by the use of systematic instead of random sampling. The statistical approach greatly improves the estimate of monthly means of pollution indicator parameters and is an effective approach for pollution control when coupled with the “polluter pays principle”.

Key words: sampling programme; effluent monitoring; sampling strategies; pollution control

1. Introduction

The success or failure of any monitoring scheme depends to a large extent on the design of the sampling programme. Although there is vast literature on chemical analysis of water and wastewater, and adequate information about sampling techniques, relatively little has been published on the third most important aspect of effluent monitoring, namely the design of the sampling programmes. Montgomery and Hart (1974) carried out a comprehensive review of the subject, and a later paper by Ellis and Lacey (1980) developed the subject further and made it easy to assimilate. They emphasised that the following elements should be clearly established before the design of a sampling programme, namely, the objectives of the sampling programme, the uncertainty tolerable in the results, what is known and can be assumed about the variability in the system and the resources available.

In order to curb and control the emission of pollutants to the environment many countries have put in place various pieces of legislation to regulate the quality and quantity of pollutant discharges.

The management and implementation of these pieces of legislation is normally handled by national or local government statutory bodies. Local municipalities are often tasked with pollution control activities within their areas of jurisdiction. Since most process industries are normally located within urban centres, most of them discharge their effluent into public owned treatment works (POTW) for subsequent treatment and disposal into natural water courses. Thus, it is the duty of the municipality to establish and enforce effluent discharge standards and to monitor the quality and quantity of effluent discharged. For this to be achieved sampling of the effluent stream is necessary and a well designed sampling programme is required to meet the objectives of sampling. Since sampling involves making inferences about the whole effluent stream from a minute fraction of the stream, assumptions of a statistical kind have to be made which depend on the objectives of the sampling programme.

The major sampling objectives for most municipalities are to monitor compliance with regulatory standards, to estimate mean concentrations and to assess the strength and biodegradability of industrial effluents to enable controlled treatment of these wastes. For these objectives to be effectively achieved at minimum cost, an effectively designed sampling programme is required. For most municipalities samples are randomly collected without proper knowledge of the discharge patterns of various industries. As a result the information obtained does not give a good reflection of the quality of effluent being discharged. This paper seeks to highlight the benefits of statistically designed sampling programme as an effective tool for effluent monitoring by municipalities.

The major objective of this study was to develop guidelines to assist municipalities in developing scientifically sound sampling programmes within the available cost and manpower constraints, in order to effectively monitor and control the discharge of industrial effluents into public sewers.

Effective pollution monitoring is essential in most developing countries as a result of the prevalent pollution of natural waters by effluent discharges from POTW. The rapid urban population growth in developing countries coupled with the increased industrialization has increased the complexity of pollution control. Thus, most public sewage treatment plants in developing countries have to deal with hydraulic loads above their design capacity and complex industrial waste which is not amenable to treatment by methods employed in most sewage treatment plants.

Since most municipalities discharge their effluent directly to natural watercourses, they have to conform to national discharge standards set by individual Water Acts within their countries. Most of these legislative pieces impose penalties for non-compliance of discharged effluent, as a result most municipalities owe substantial amount of money to national Pollution Control Units and the debts are expected to continually increase as a result of continued non-compliance. Given this scenario most municipalities have introduced the "Trade Tariff Scheme", which is based on the "the polluter pays" principle as an effective means of transferring the cost of effluent handling and treatment to the producer of the effluent. This means that the polluter would bear the cost of preventing or minimising and perhaps remediation of any environmental damage he has caused. For this approach to be effective properly designed sampling programmes that allow accurate characterisation of effluent at a minimum cost need to be developed by municipalities.

2. Experimental

2.1 Methodology

The sampling programme was divided into two phases. The first phase lasted for two months and was designed to simulate random sampling data. Five snap samples were collected on each 8h-day work shift after a time interval of 1.5 h. Samples were collected from each industry after a one day time interval during the two months period, ensuring that sampling was done twice per each specific day of the week in each sampling month. The objective of this phase was to build an understanding of the variation of the effluent quality over time and thus provide data for the statistical design of the sampling programme.
The objective of the second phase was to establish pollution trends and monthly means of the major pollutant parameters. Four samples were systematically collected per month from each industry for a period of 3 months, based on the results obtained in the first phase.
Four industries were used as case studies, namely the tannery, textile finishing, sugar refining and sugar confectionery industrial sectors. The major pollution indicator parameters considered were the pH, total suspended solids (TSS) and chemical oxygen demand (COD).

2.1.1 Experimental methods and equipment

Samples were collected manually. The pH was measured on the spot and the TSS and COD were measured in the laboratory within 24 hours after sample collection. The samples were analysed using standard methods as outlined in the Standard Methods Handbook (AWWA, APHA and WEF 1998). Effluent pH was measured using a Hach pH meter (Model 51935-00). Chemical oxygen demand (COD) was determined by closed reflux method using a Hach COD reactor (Model 45600) followed by calorimetric determination of Cr^{3+} at a wavelength of 620 nm using a Hach spectrophotometer (Model DR 2010). TSS were determined by measuring the amount of light scattered by the solids at a wavelength of 810 nm, using a Hach spectrophotometer (Model DR 2010).
The amount of effluent discharged was measured using a Millitronics open channel monitor (OCM III, Model PL-505) in conjunction with a remote ultrasonic transducer (Model ST-25) with an in built temperature sensor.

3. Results and Discussions

3.1 Sampling strategies

Three sampling strategies of paramount importance in effluent monitoring, namely random sampling, random sampling with time weighting and systematic sampling were evaluated (Sanders et al. 1983). Random sampling is a method of sampling where the chances of obtaining different concentration values of a determinant are precisely those defined by the probability distribution of the determinant in question. Systematic sampling involves the collection of samples at predetermined intervals, often equally spaced in time. Random sampling with time weighting involves weighing each measured concentration by the time interval represented by that observation.

The variability of the system was estimated from the flow measurement results since this gives a better estimate of the system variability as compared to the water quality data. This is so because the pollution load has been shown to display a variation similar to that of the flow rate (Nicoll 1988). In order to confirm that the flow measurement data was normally distributed, normal probability plots were constructed and the *r* (correlation) and *p* values shown in Table 1 were obtained.

Table 1 The *r* and *p* values for flow data from each industry

Parameter	Tannery	Sugar refinery	Sugar confectionery	Textile
r	0.982	0.957	0.845	0.917
p	≈0.5	≈0.5	≈0.4	≈0.4

An excellent fit was obtained for data from the tannery and sugar refinery industries i.e. ($p \geq 0.5$ for excellent fit); while a good fit was obtained for data from the sugar confectionery and textile industries i.e. ($0.10 \leq p < 0.5$ for a good fit). Thus, the data can be assumed to closely approximate a normal distribution. The variability of the effluent load for each of the four industries is shown in Table 2.

Table 2 Standard deviation of the effluent flow rates for the selected industries

Industry (average flow, m^3/day)	Standard deviation			
	Within-day ($n = 24$)	Between-day ($n = 7$)	Overall ($n = 165$)	Within/overall
Tannery (15)	0.75	3.68	3.75	0.20
Sugar refinery (30)	2.28	29.37	29.46	0.08
Sugar confectionery (0.3)	0.03	0.38	0.38	0.08
Textile (312)	36.67	313	315	0.11

The within-day/overall standard deviation ratio reveals that greater benefits are expected if systematic sampling is used. The smaller the value of this ratio the more pronounced is the between-day variation and the greater is the benefit of systematic sampling over random sampling (Montgomery and Hart 1978). A comparison of the three sampling strategies by Ellis and Lacey (1980) showed that although random sampling provides an unbiased estimate of the mean, it requires a large number of samples and is often an expensive exercise. Systematic sampling on the other hand reduces the sampling effort, as only few samples are needed to estimate the mean concentrations. However, it requires a greater understanding of the process to ensure that the sampling does not run in step with some cyclic component of variation in the process. Thus, systematic sampling was considered the best sampling strategy for industrial effluents. Sampling after every 8 days was found to be satisfactory and the statistical benefits of systematic sampling were realized.

3.2 Statistical design of the sampling programme

The number of samples (*n*) needed to estimate the average concentration of target analytes within the study boundary is given by equation 1 (Keith et al. 1983);

$$n = \left(Z_{1-\alpha/2}\sigma_p / E\right)^2 \qquad (1)$$

where $Z_{1-\alpha/2}$ is the value of the standard normal variant, σ_p is the standard deviation for the sample population and *E* the tolerable uncertainty for the estimate of the mean for the characteristic value of interest (an arbitrary value of the amount of uncertainty one is willing to accept in the data). Substitution of the sample standard deviation (*SD*) in place of σ_p in equation 1 requires the use of the student "t" static ($t_{1-\alpha/2}$) instead of $Z_{1-\alpha/2}$ and a two-step procedure based on equations 2 and 3. In cases where one wants to determine the number of samples (*n*) required, where the tolerable error and an estimate of variability is known, the *Z* distribution (equation 2) is initially used to estimate *n*. The *n* value calculated using equation 2 is then used to obtain the *t*-value used in equation 3, and then *n* is recalculated (iterative) using equation 3 until a stable value of is obtained *n* is obtained.

First pass equation: $n = \left(Z_{1-\alpha/2}SD / E\right)^2$ (2)

Second pass equation: $n = \left(t_{1-\alpha/2}SD / E\right)^2$ (3)

The uncertainty in the estimate of the monthly mean for any pollution parameter for a given number of samples, at 95% confidence interval was calculated using equation 3 and is shown in Table 3.

Table 3 The uncertainty levels in the estimation of monthly mean levels at 95% confidence interval for a given number of samples

Industry	Standard uncertainty (per no. of samples)			
	2	3	4	5
Tannery	±2.38	±0.359	±0.149	±0.0833
Sugar refinery	±7.24	±1.09	±0.453	±0.253
Sugar confectionery	±0.0953	±0.0143	±0.00597	±0.00333
Textile	±116	±17.5	±7.29	±4.07

From Table 3 it can be seen that the uncertainty in the estimate of the monthly mean decreases by approximately 94% if four samples instead of two are collected every month to estimate mean levels. Collecting five samples results in an approximately 96% decrease in the uncertainty estimate of the monthly mean. Thus, collecting more than four samples does not result in a significant reduction in the magnitude of the standard uncertainty. Therefore, four samples per month were collected in second phase of the sampling programme in order to estimate the monthly means for the parameters monitored.

The relative standard uncertainty for monthly mean estimates is approximately 1% for the tannery, sugar refinery and sugar confectionery industries and 2% for the textile industry if four samples are collected per month. Five samples are required for the textile industry to have a standard uncertainty of 1%.

3.3 Comparison of historical and experimental data

The historical data represents the data collected by the municipality while the experimental data represents the data collected in the monitoring survey. The historical data was collected by random sampling over a period of five years. The historical and experimental results are shown in Table 4.

For the sugar refining industry no data is presented since the historical data was not available and the effluent treatment plant was upgraded in between the two experimental phases making it difficult to compare the data from both phases.

Table 4 Comparison of historical and experimental data

Parameter	Historical		Phase 1		Phase 2	
	Range	Mean	Range	Mean	Range	Mean
Sugar confectionery						
pH (pH units)	6.1–8.2	7.2	3.6–7.8	6.9	6.0–7.7	7.0
COD (mg/L)	263–49,000	12,892	5220–21,850	11,450	150–19,800	6,687
SS (mg/L)	40–3,430	959	12–632	132	12–383	59
Textile						
pH (pH units)	6.9–11.3	9.4	5.2–11.8	8.8	4.4–10.8	8.8
COD (mg/L)	39–3,658	1,468	3,700–10,200	5,849	280–59,130	7,122
SS (mg/L)	68–1,500	320	40–3,840	521	39–355	204
Tannery						
pH (pH units)	1.5–12.6	7.4	4.8–11.3	6.8	3.4–12.8	8.4
COD (mg/L)	1,100–39,984	12,862	2,100–41,400	20,826	3,800–57,900	35,267
SS (mg/L)	390–7,200	2,258	304–7,420	2,147	100–10,460	4,702

Comparison of the three sets of data (historical, phase 1 and 2) was done using the Kruskal-Wallis test with α = 0.05. The null hypothesis (H_o) was that the populations from which the data was drawn have the same mean, while the alternative hypothesis (H_A) was that one of the populations has a mean larger or smaller than at least one other population. From Table 5, K_w or K'_w is less than 5.99 (i.e. $\chi^2_{0.95,2} = 5.99$) only for the sugar confectionery pH; the textile pH and SS and the tannery COD. Thus, for these instances we can not reject H_o at α = 0.05.

Table 5 Computed Kruskal–Wallis statistic for the historical, phase 1 and 2 data

Industry	Calculated K_w or K'_w (for pH) $\chi^2_{0.95,2} = 5.99$		
	pH	COD	SS
Tannery	13.19	5.75	8.50
Sugar confectionery	3.77	6.69	22.13
Textile	1.73	14.58	2.13

Comparison of the data from phase 1 and 2 was performed using the Wilcoxon Rank Sum Test (one tailed α test) with $\alpha = 0.05$. The null hypothesis (H_o) was that the populations from which the two data sets were drawn have the same mean, while the alternative hypothesis (H_A) was that the measurements from phase 2 data exceed those from phase 1.

Table 6 shows that $Z_{rs} \geq 1.645$ (i.e. $Z_{0.95} = 1.645$) for the tannery pH and SS; and the textile COD thus for these cases we reject H_o and accept H_A that the parameters measured in phase 2 exceeded those in phase 1.

Table 6 Computed Z statistic for the phase 1 and 2 data

Industry	Calculated Z_{rs} $Z_{1-\alpha} = 1.645$		
	pH	COD	SS
Tannery	2.32	0.97	2.24
Sugar confectionery	0.3	1.58	0.78
Textile	0.4	2.13	1.42

From the above analysis it can be concluded that although systematic sampling (phase 2) reduces the sampling frequency and number of samples collected the results obtained were comparable with those using random sampling (phase 1) in a majority of cases. However, there was a low degree of agreement between the results obtained in the experimental sampling phases and those obtained in the historical phase. Mean levels of the parameters monitored in the historical phase were lower than those in the experimental phase for most of the instances. Values of the mean levels of the parameters monitored using systematic sampling (phase 2) were either comparable or higher than those obtained using random sampling (phase 1). This demonstrates that systematic sampling is effective in identifying peak pollutant levels although fewer samples are collected.

Pollution control activities require the collection of data that gives an accurate reflection of the effluent quality over shorter time period (ideally over a month). Given the resource and manpower limitations of most municipalities in developing countries, systematic sampling is a more effective method of pollution control as opposed to random sampling.

The successful implementation of systematic sampling requires the availability of adequate resources, which are not readily available in most municipalities. As a result, only a few parameters (pH, SS and COD) are monitored and the volume of discharge is not quantified. To address this resource limitation requires the implementation of the "polluter pays principle". This means that the industrial effluent dischargers will have a pay a monthly tariff to cover the cost of monitoring and treatment of their effluent based on the effluent quantity and quality. Previously most municipalities used a fixed tariff system in addition to penalties for non-compliance to sewer discharge limits. Since the cost of such tariffs and penalties is insignificant as compared to the cost of effluent treatment, most industries did not prioritise pollution control.

Implementation of the "polluter pays principle" will also require a shift from the use of concentration based standards to the use of pollutant loads as a means of pollution control. Pollutant load based standards deter industries from using effluent dilution as a means of meeting concentration based discharge limits and encourages industries to minimize the amount of effluent discharged or fresh water utilization. Implementation of this principle will ensure that industries have a financial incentive in moving away from pollution control to pollution prevention.

3.3.1 Financial benefits of systematic sampling

To demonstrate the financial incentives of systematic sampling, a tariff formula has been selected and used as a guideline. Most municipalities use a tariff formula based on volume of discharge and a particular effluent quality characteristic e.g. chemical oxygen demand (COD), permanganate value (PV), oxygen absorbed (OA) and settleable solids (SS). Of these parameters a tariff formula based on volume and COD and SS is considered the best guideline from an effluent treatment point of view. The major advantage of dichromate oxidation which yields the COD value as opposed to the permanganate oxidation yielding the PV value is that virtually all the organic compounds in the sample are oxidised because of its greater oxidising power as compared to other oxidants. The formula assumed for the purpose of these calculations is given by equation 4, where COD is in mg/L; *R* is the conveyance cost, assumed to be ZAR2/m^3; *T* is the treatment cost, assumed to be ZAR3/m^3 and the term $(COD - 2000/2000)$ and $(SS - 1000/1000)$ represents a penalty factor based on exceeding the COD limit of 2000 mg/L and SS limit of 1000 mg/L. The calculations are based on the average monthly COD and SS values obtained in both the historical and phase 2 sampling phases (Table 4) and a three month billing cycle.

$$ZAR/m^3 = R + T\left(\frac{1}{3} + \frac{COD}{1500} + \frac{COD - 2000}{2000} + \frac{SS - 1000}{1000}\right) \qquad (4)$$

For phase 2 results (multiple systematic sampling) the mean levels of the pollution indicator parameters were calculated using equation 5, where $\bar{x}$ is the mean level, *J* is the number of monthly systematic samples each of size n and $\overline{x_j}$ is monthly mean for each parameter. For random sampling (historical and phase 1) the arithmetic mean was used to calculate the monthly average value for each pollution indicator parameter.

$$\bar{x} = \frac{1}{J}\sum_{j=1}^{J}\bar{x}_j \qquad (5)$$

Comparison of the tariff fee charge obtained using the historical data (random sampling) and the phase 2 data (systematic sampling) is shown in Table 7. The results show that the tariff charge based on the historical results is far less than that obtained using the phase 2 results with the exception of the sugar confectionery industry.

Table 7 Comparison of tariff charges using the historical and phase 2 sampling data

Industry	Tariff charge (ZAR/m^3)	
	Historical	Phase 2
Tannery	48.79	106.36
Sugar confectionery	45.12	23.64
Textile	5.14	26.79

Thus, systematic sampling coupled with the "polluter pays principle" is an effective approach to pollution control and is capable of offsetting the cost of sampling, sample analysis, effluent conveyance and treatment while also acting as a financial incentive for industry to address pollution. Moreover, as illustrated in Table 7, it greatly reduces the sampling burden while ensuring that adequate information is gathered to ensure equitable billing.

4. Conclusions

This paper demonstrates the effectiveness of systematic sampling as a means of effective monitoring of industrial effluent discharges by regulatory agencies. When coupled with the "polluter pays principle" and pollutant load based discharge standards this method of industrial effluent monitoring is effective, economic and encourages a pollution prevention approach. This will improve the quality of treated effluent discharged into natural watercourses from POTW and ease stress on fresh water supplies by encouraging industries to explore ways of minimizing fresh water utilization.

References

APHA, WEF, AWWA (1998) Standard Methods for the Examination of Water and Wastewater. Clesceri, L.S., Greenberg, A.E., Eaton, A.D. (ed) 20th edn. American Public Health Association, Washington, DC, USA.

Ellis JC, Lacey, RF (1980) Sampling: Defining the task and planning the scheme. Water Pollution Control 452- 467.

Keith LH, Crummet W, Deegan J et al (1983) Principles of environmental analysis. Analytical Chemisty (United States) 55:14: 2210- 2218.

Montgomery HAC, Hart, IC (1974) The design of sampling programmes for rivers and effluents. Water Pollution Control (London) 73(1): 77- 101.

Nicoll EH (1988) Load variation. In: Small water pollution control works: Design and practice, Ellis Horwood, England, pp84 – 91.

Sanders TG, Ward RC, Loftis JC, Steele TD, Adrian DD, Yevjevich V (1983) Sampling frequency. In: Design of network systems for monitoring water quality, Water Resources Publication, USA: pp156.