



WATER QUALITY INFORMATION SYSTEM DEVELOPMENT: SAMPLE SIZE OPTIMIZATION

Anass El-Molla¹, M. A. Abdelall¹, M. Shaban² and Ahmed Habash²

¹ Irrigation Design and Hydraulics, Faculty of Engineering, Al-Azhar University.

² Drainage Research Institute, National Water Research Centre, Ministry of Water Resources and Irrigation, El-Kanater El-Khairia, PO 13621/5, Cairo, Egypt.

ملخص البحث :

يعد تكرار أخذ العينات من الأمور الهامة جدا للحكم علي مدي كفاءة شبكة مراقبة نوعية المياه. ويرتبط جزء كبير من تكاليف تشغيل شبكة رصد نوعية المياه ارتباطا مباشرا بتكرار أخذ العينات من قنوات المياه في دلتا نهر النيل. في هذا البحث تم أخذ العينات من مواقع المراقبة والرصد لنوعية المياه من قناة بحر شبين في وسط دلتا نهر النيل وقناة الإسماعيلية في شرق دلتا النيل. الهدف من التحليل الإحصائي هو تقييم وإعادة تصميم المعدل الزمني لأخذ العينات إلى الفترة الزمنية المثالية حول القيمة المتوسطة. ويساعد هذا النهج على تحديد تردد أخذ العينات الذي يسمح بتقدير قيمة المتوسط السنوي بدقة مقبولة. وتشير النتائج إلى أن النهج المقترح يمثل أداة مفيدة لدعم متخذي القرار من أجل الاختيار الأمثل لفترات تكرار أخذ قياسات وعينات نوعية المياه. وأظهرت النتائج أن أخذ 12 عينة سنويا من مواقع الرصد بقناة بحر شبين بوسط الدلتا تعطي خطأ في قيم الأكسجين الحيوي والأكسجين الكيماوي بمقدار 27% و 25% من النتائج الخاصة بقياسات عناصر نوعية المياه (الأكسجين الحيوي – الأكسجين الكيماوي). اما بالنسبة لترعة الإسماعيلية فإن تكرار أخذ 12 عينة سنويا من مواقع الرصد الخاصة بنوعية المياه تعطي خطأ في قيم الأكسجين الحيوي وقيم الأكسجين الكيماوي بمقدار 45% و 50% من النتائج الخاصة بقياسات عناصر نوعية المياه (الأكسجين الحيوي – الأكسجين الكيماوي)

ABSTRACT

Sampling frequency is a very important consideration to control the efficiency of a water quality monitoring network. A large portion of the costs of operating a monitoring network is related directly to the frequency of sampling applied to the water quality channels in the Nile delta. The sampling frequency was carried out for the monitoring locations in Bahr Shebeen canal in Middle Nile Delta and Ismaillia canal in Eastern Nile Delta. The statistical method of “Single Station, Single Variable” was used to assess and redesign of sampling frequency is based on the confidence interval around the mean value. The approach helps identifying the sampling frequency that allows estimating the annual mean value with an acceptable accuracy. Results indicate that the proposed approach represents a useful decision support tool for the optimized selection of water quality variables. The results showed that 12 samples per year at the monitoring site of Bahr Shebeen canal may allow the estimation of the BOD and COD mean values with 27% and 25% error, respectively. On the other hand, at the monitoring sites of Ismaillia canal may allow the estimation of the BOD and COD mean values with 45% and 50% error, respectively.

1. INTRODUCTION

About 97% of Egypt's water resources are from the Nile River (Dijkman 1993). The distribution of Egypt's share of the Nile water to its population is near the water poverty threshold and will fall well below this threshold in the years to come. The re-assessment of monitoring networks is the translation of the monitoring objectives to specify sampling sites, sampling frequency and the variables to be measured.

Both sampling frequency and sampling sites are influenced by the water quality variable being monitored, and therefore the selection of the specific variables of interest is intrinsic to the design and subsequent operation of a water quality monitoring network (Strobl and Robillard, 2008). Sampling frequency is a very important consideration in the design of a water quality monitoring network. A large portion of the costs of operating a monitoring network is related directly to the frequency of sampling. In addition, the reliability and utility of water quality data derived from a monitoring network are likewise related to the frequency of sampling. Water quality sampling frequency is the single aspect of network design which received most attention from researchers (Sanders et al. 1983). This doesn't mean that sampling frequency is more important than site selection or variable selection. Sampling frequency can be treated objectively using statistical analysis while other problems of network design tend to be more subjective especially in the initial design phases.

It is proposed that a minimum number of samples per year based on the monitoring objectives. For instance, a minimum of six samples per year is suggested in the case that the monitoring objective is to assess seasonality. In addition, uniform sampling is suggested (e.g. if 12 samples to be collected per year, it could be once a month; if six samples to be collected per year, it could be once every two months and so forth). If enough records are available per season, the approach can be applied to identify the number of samples required to estimate seasonal mean values (e.g. number of samples per season). The objective of this paper is to develop the water quality information system by enhancing the water sampling frequency for each water quality variable at each of the sampling sites, using the statistical method of "Single Station, Single Variable" for Bahr Shebeen and Ismaillia canals in the Nile Delta.

2. STUDY AREA AND METHODOLOGY

2.1. Study area

Figure (1) presents the map of the two canals of interest, Bahr Shebeen and Ismaillia. Bahr Shebeen Canal represents a shallow and narrow irrigation canal. It extends for about 80 km into the Egyptian Delta, passing through three governorates and surrounded by two major cities, various villages, and cultivable lands. Ismaillia canal is 128 km long its depth is about 1-3 m and its width is about 30 - 70 m. A big part leads through sandy areas. As the water level is above the surrounding land there's is a significant amount of seepage from the canal. It is reported to increase the canal's capacity from 135 m³/s up to 440 m³/s (Asit K. et al.,1983). According to the year book of Egypt (Egypt year book, 1995) widening and deepening of Al Ismailia Canal and its subordinate ditches has been done in 1995 expected to serve around 832 thousand feddans (1 feddan = 0.4047 ha). The Drainage Research Institute (DRI) is the source of these measured water quality parameters at certain monitoring locations as shown in Figure (1). The locations are distributed upon the stream of Bahr Shebeen canal which includes (8 sites) in Middle of the Nile Delta and (5 sites) cover the stream of Ismaillia canal in the Eastern Nile Delta. The current monitoring frequency is 12 records per year for each water quality parameter between the years 2008 to 2016.

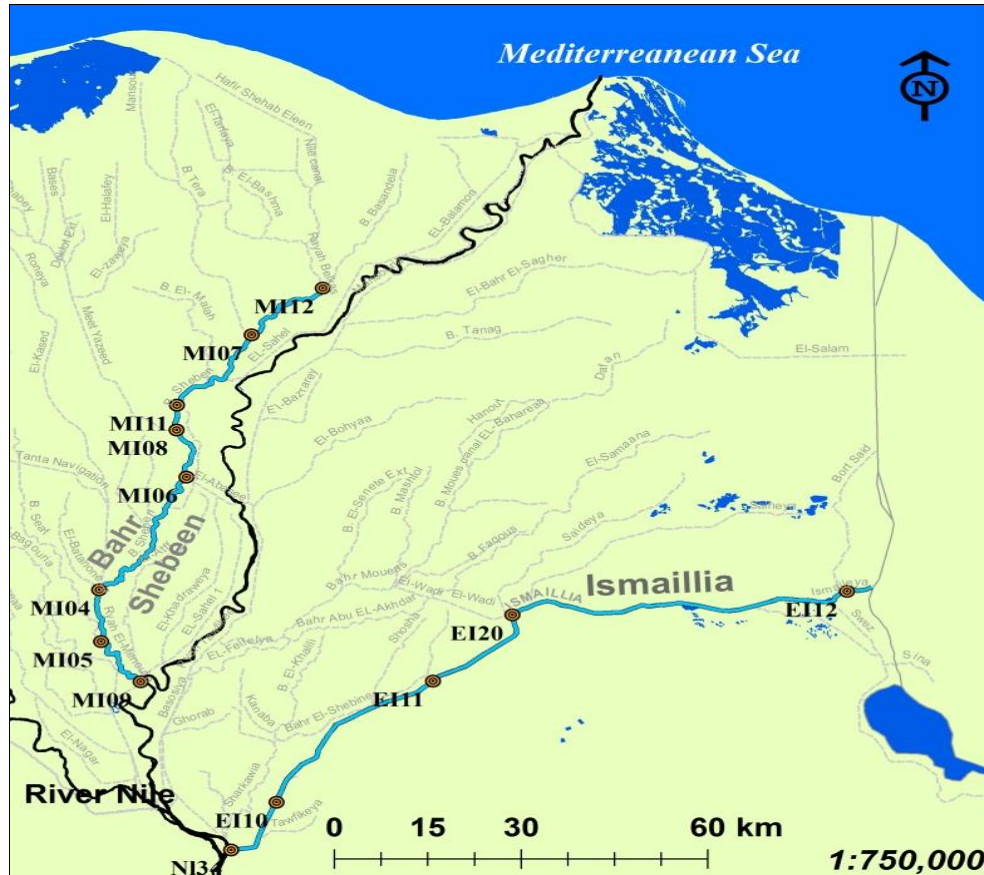


Figure (1): layout map of monitoring sites for Bahr Shebeen and Ismailia canals

2.2. Methodology

The statistical formulas were used to perform the sampling frequency analysis for selected water quality parameters of Iron (Fe mg/l), Chemical Oxygen Demand (COD mg/l), Biological Oxygen Demand (BOD mg/l), Dissolved Oxygen (DO mg/l), Nitrate (NO₃ mg/l), Negative Logarithm of Hydrogen Ion Concentration (pH), Total Dissolved Solids (TDS mg/l) and Ammonia (NH₄ mg/l).

This approach was proposed by Sanders and Adrian (1978) and Sanders et al. (1983). It assumes that the monitoring objectives are the determination of ambient water quality status and an assessment of annual means. The confidence interval approach allows selecting a sampling frequency that yields an estimate of the sample mean (\bar{x}) within a prescribed degree of accuracy (i.e. confidence limits) (Khalil et al., 2014). For the purpose of this study, the water sampling frequencies were estimated using the "Single Station-Multiple Variables Approach" described in Sanders et al. (1983). Since several water quality parameters are important, a separate sampling frequency were computed for each parameter and a simple average sampling across parameters were computed and used for the monitoring purposes.

Weighted average variances over the several water quality variables were computed for each station. However, this approach has its limitation when the water quality parameters measured have different order of magnitude. To overcome this problem, a normalization process was carried out to deal with percentages (from maximum values) instead using the original data. It has to be noted that this does not influence the original sampling variances. There are three main statistical approaches to evaluate the sampling frequency with respect to the mean value confidence interval as explained by Sanders et al. (1983) and Harmincioglu et al. (1999). These

approaches are “Single Station, Single Variable Approach”; “Single Station, Multiple Variables Approach”, and “Multiple Stations, Multiple Variables Approach”. For detailed assessment of sampling frequency for each water quality variable at each of the sampling sites, the “Single Station, Single Variable Approach” is recommended. A formula for sample size is based on the assumption that the water quality measures are stationary. Assuming that the population variance is known and observations are independent, the variance of the sample mean $\text{var}(\bar{x})$ may be computed from the following equation:

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n} \quad (1)$$

Where σ is the population variance; and n is the number of observations.

The n required to obtain a given degree of confidence can be derived from the following formula (Sanders et al., 1983):

$$\left[\bar{x} - z_{\alpha/2} \text{var}(\bar{x})^{\frac{1}{2}} \leq \mu \leq \bar{x} + z_{\alpha/2} \text{var}(\bar{x})^{\frac{1}{2}} \right] \quad (2)$$

Replacing $\text{var}(\bar{x})$ by σ^2/n , the n required to obtain a future estimate of the mean can be computed with a known level of confidence by the following equation (Sanders et al., 1983):

$$n \geq \left[\frac{z_{\alpha/2} \sigma}{\mu - \bar{x}} \right]^2 \quad (3)$$

Substituting the sample variance (s) instead of the population variance (σ) in the previous equation requires the Student’s “ t ” statistic instead of the z statistic. The difference between the true population mean and the sample mean ($\mu - \bar{x}$) is referred to as the error (E) (Harmincioglu et al., 1999). This substitution results in the following equation:

$$n \geq \left[\frac{t_{\alpha/2} s}{E} \right]^2 \quad (4)$$

Using Equation 4, an acceptable confidence level (t_{α}) and the error (E) must be specified to determine n . Since the t statistic is dependent on the degree of freedom, which is dependent on n , Equation 4 becomes an iteration problem. An initial n value is specified, from which a degree of freedom is calculated and used to find the t statistic. Equation 4 updates the value of n , and a new t statistic is found. These iterations are performed until the sample size converges to a relatively constant value. The relationship between the error (E) and the number of samples as described by Equation 4 becomes theoretically valid if the variance of the stationary component computed from different sampling intervals stabilizes after a certain sampling interval (Khalil and Ouarda, 2009). After this sampling interval, the variance becomes independent of the sampling interval and any change in the number of samples will only affect the expected error.

Tokgoz (1992) and Harmancioglu and Tokgoz (1999), illustrated that, in case of limited water quality records with large gaps there will be difficulties in meeting the underlying assumptions of the confidence interval approach. Khalil et al. (2014) recommended that the minimum size of records to apply this approach is 30 valid records. For water quality parameters that have a size of records less than 30, it is recommended to continue measure these water quality parameters, and sampling frequency assessment may take place when enough records are available. In practice, autocorrelation may be present, in this case the variance of \bar{x} is given in equation (5) by (Loftis and Ward, 1979):

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} \left[1 + \frac{2}{n} \sum_{k=1}^{n-1} (n-k) \rho_k \right] \quad (5)$$

Where σ is the population variance. If the $\text{Var}(\bar{x})$ given by Equation 5 is substituted into Equation 2 results in a quadratic equation that can be solved for n . k is the number of lags between samples and ρ_k is the autocorrelation coefficient for lag k . Gilbert (1987) obtained an approximate expression by ignoring the term in the quadratic equation that has n^2 in the denominator as in equations 6 and 7:

$$n = D \left[1 + 2 \sum_{k=1}^{n-1} \rho_k \right] \quad (6)$$

$$D = \left[\frac{(t_{\alpha/2}) s}{E} \right]^2 \quad (7)$$

In addition, normality tests were examined using Kolomogrov-Smirnov goodness of fit test. Autocorrelation, assessment of temporal trend, and data distributions will be examined in a preliminary analysis step.

3. Results and Discussion

As described in the methodology section, half of the confidence interval around the mean value is considered as the error in estimating the population mean. In this assessment, several error percentages were considered, ranging from 10% to 60% of the sample mean at proposed number of samples ($n=12$) for each monitoring site. The sampling frequency assessment curves were illustrated for only two sampling sites for each canal. The first sampling site is located at the beginning of the canal stream and the second is at the end of canal stream.

3.1. Bahr Shebeen Canal

The sampling frequency assessment at the MI09 sampling site (table 1), where the considered error percentages are plotted against the number of samples obtained for the 95% significant level. Figure (2) shows that 12 samples per year at monitoring site (MI09) may allow the estimation of the BOD and COD mean values with 40% and 42% error, respectively. Similarly, results for the NH_4 and TDS indicated that 12 samples per year may allow estimating the annual average with a 35% and 45% error, respectively (Figures 3-a and 3-b). On the other hand, the sampling frequency assessment at the MI12 sampling site as in Figure (4) and Table (1) explain that 12 samples per year may allow the estimation of the BOD and COD mean values with 37% error. In figures (5-a and 5-b), results for the NH_4 and TDS indicated that 12 samples per year may allow estimating the annual average with a 45% and 20% error, respectively. Sampling frequency calculated for all sampling sites covers Bahr Shebeen canal varied from 18 to 19 samples per year. The overall average was 19 samples per year with a standard deviation equals to 2.0 samples.

Table (1): Number of samples/year (n) versus expected error (%) at sites MI09 and MI12 on Bahr Shebeen canal

	WQ parameter	Site (MI09)				Site (MI12)			
		Error (e) %				Error (e) %			
		10	15	20	25	10	15	20	25
n (No. of Samples)	BOD	150	68	39	26	129	59	34	23
	COD	140	64	37	24	142	65	37	25
	TSS	211	95	55	36	147	67	39	26
	NO3	216	97	56	37	417	186	106	69
	NH4	574	256	145	94	408	182	104	67
	Fe	189	85	49	32	233	105	60	39
	PH	17	11	7	6	10	6	4	4
	TDS	85	39	23	16	46	22	13	10
	DO	16	12	8	6	13	10	8	6

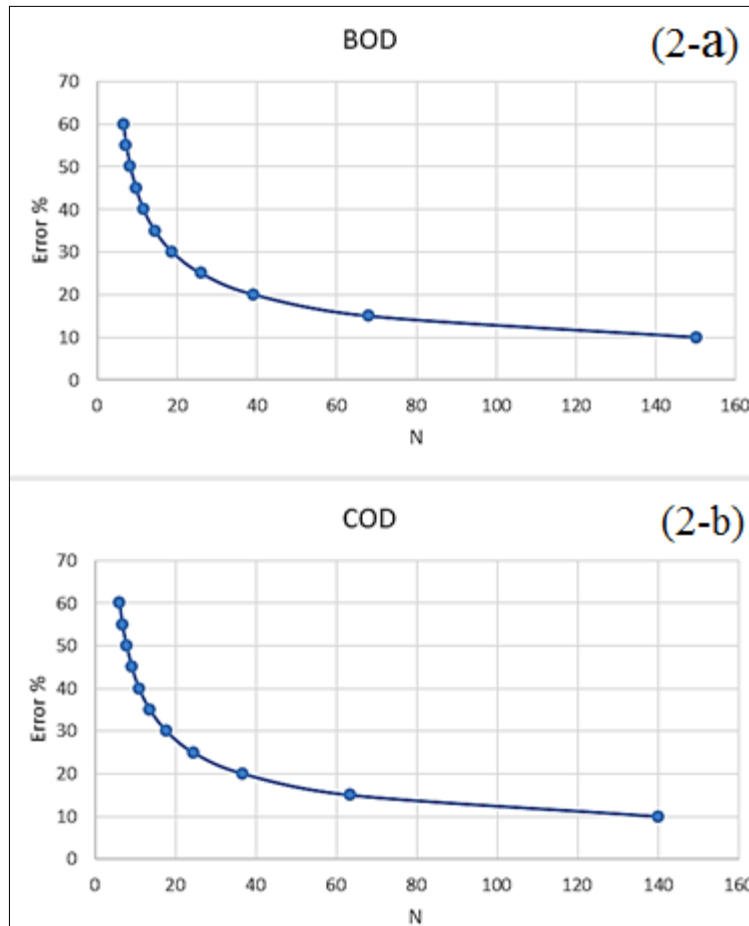


Figure (2): the curves (2-a) and (2-b) show the number of samples/year versus expected error for BOD and COD respectively at MI09 on Bahr Shebeen canal.

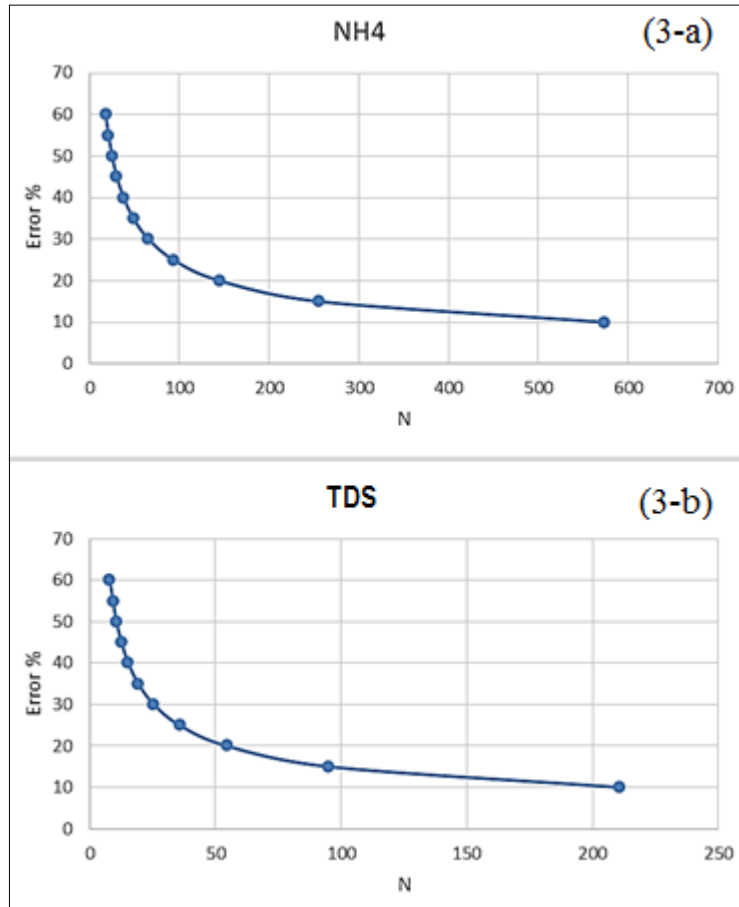


Figure (3): the curves (3-a) and (3-b) show the number of samples/year versus expected error for NH4 and TDS respectively at MI09 on Bahr Shebeen canal.

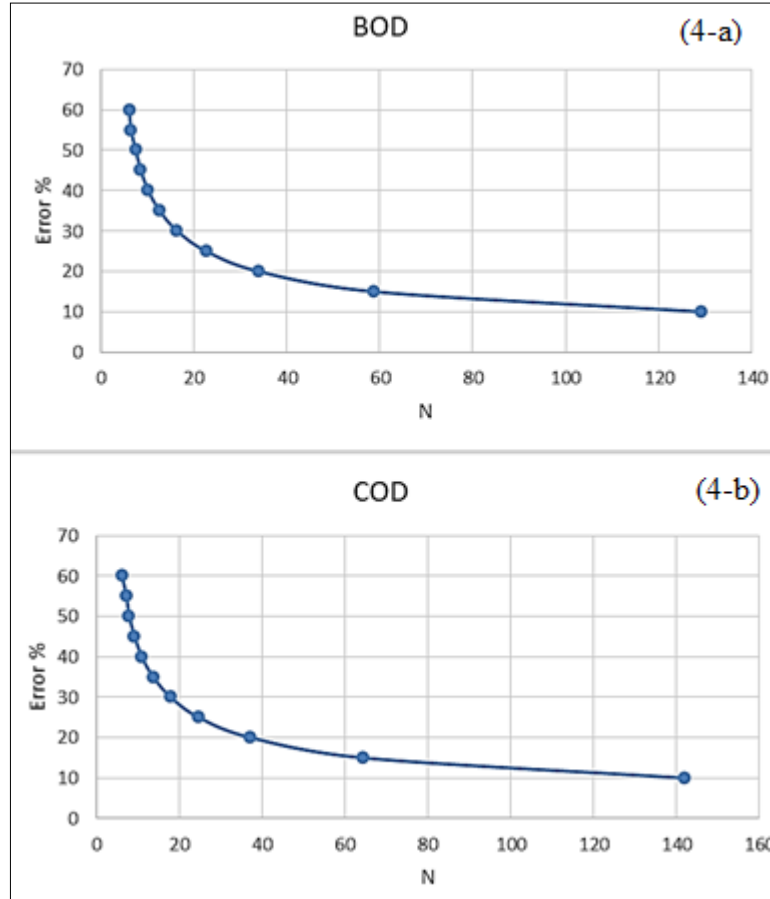


Figure (4): the curves (4-a) and (4-b) show the number of samples/year versus expected error for BOD and COD respectively at MI12 on Bahr Shebeen canal.

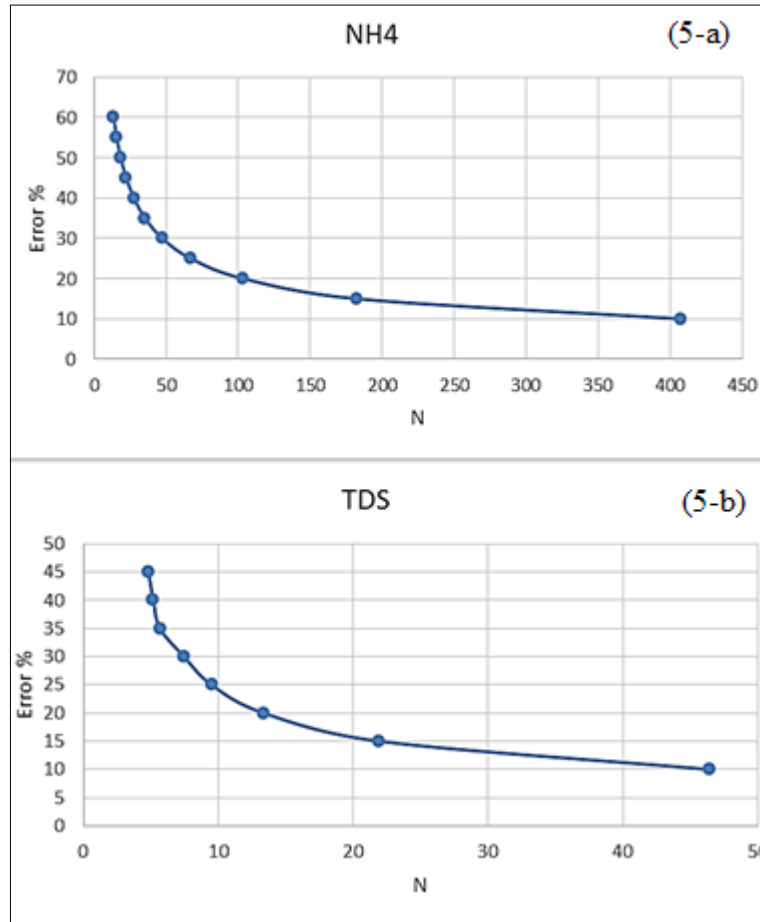


Figure (5): the curves (5-a) and (5-b) show the number of samples/year versus expected error for NH_4 and TDS respectively at MI12 on Bahr Shebeen canal.

3.2. Ismaillia Canal

The sampling frequency assessment at the EI10 sampling site (table 2), where the considered error percentages are plotted against the number of samples obtained for the 95% significant level. Figure (6) shows that 12 samples per year at monitoring site EI10 may allow the estimation of the BOD and COD mean values with 35% and 32% error, respectively. Similarly, results for the NH_4 and TDS indicated that 12 samples per year may allow estimating the annual average with a 27% and 22% error, respectively (Figure 7). On the other hand, the sampling frequency assessment at the EI12 sampling site as in Figure (8) and Table (2) illustrate that 12 samples per year may allow the estimation of the BOD and COD mean values with 25% and 18% error, respectively. In figure (9), results for the NH_4 and TDS indicated that 12 samples per year may allow estimating the annual average with a 47% and 35% error, respectively. Sampling frequency calculated for all sampling sites covers Ismaillia canal varied from 13 to 20 samples per year. The overall average was 15 samples per year with a standard deviation equals to 3.0 samples.

Table (2): Number of samples/year (n) versus expected error (%) at sites EI10 and EI12 on Ismailia canal

n (No. of Samples)	WQ parameter	Site (EI10)				Site (EI12)			
		Error (e) %				Error (e) %			
		10	15	20	25	10	15	20	25
	BOD	124	56	33	22	63	29	18	12
	COD	107	49	28	19	37	18	11	8
	TSS	148	67	39	26	105	48	28	19
	NO3	463	207	117	76	169	77	44	29
	NH4	72	33	20	13	213	96	55	36
	Fe	409	183	104	67	181	82	47	31
	PH	6	4	3	2	6	4	3	3
	TDS	54	25	15	11	115	53	31	20
	DO	12	8	7	6	7	4	3	3

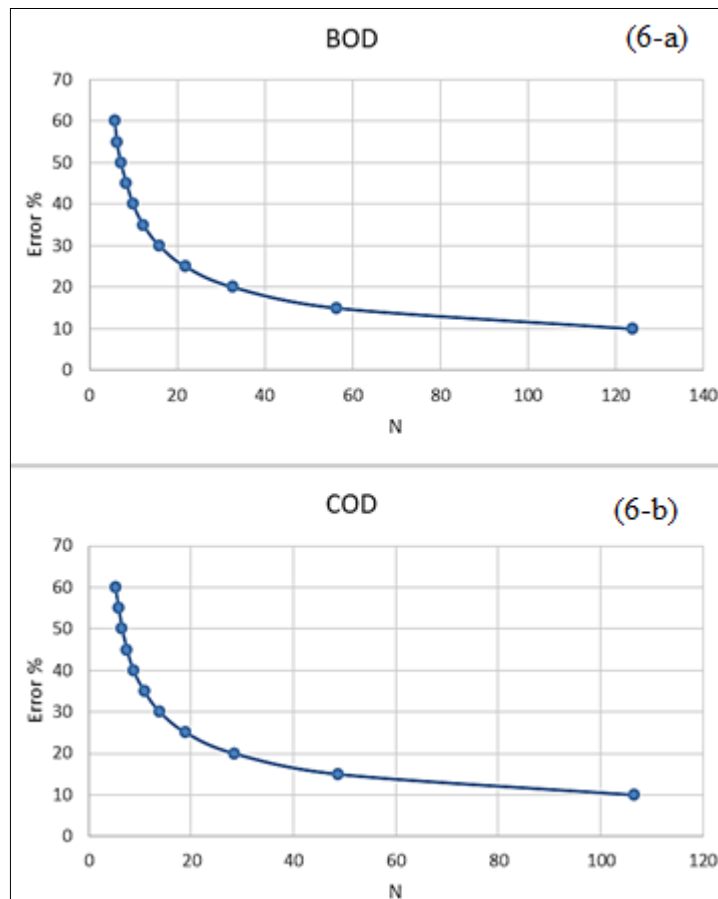


Figure (6): the curves (6-a) and (6-b) show the number of samples/year versus expected error for BOD and COD respectively at EI10 on Ismailia canal.

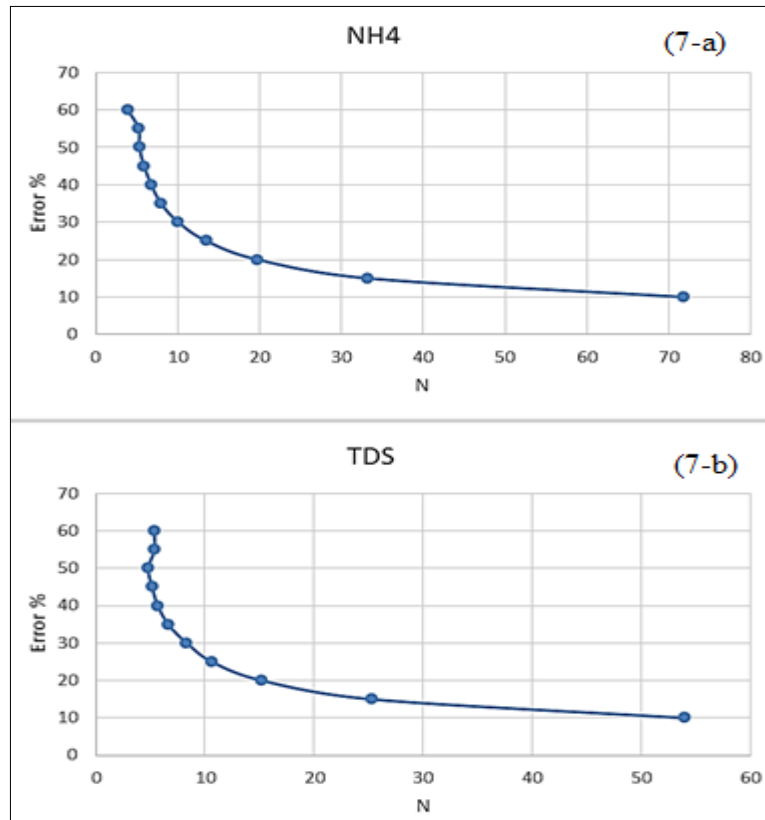


Figure (7): the curves (7-a) and (7-b) show the number of samples/year versus expected error for NH4 and TDS respectively at EI10 on Ismailia canal.

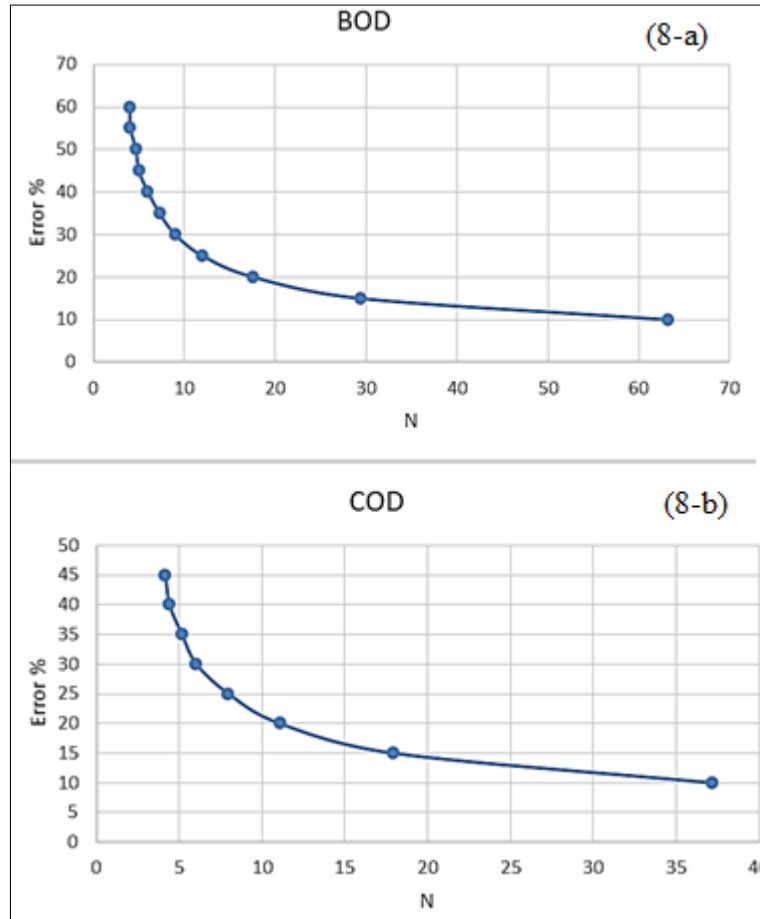


Figure (8): the curves (8-a) and (8-b) show the number of samples/year versus expected error for BOD and COD respectively at EI12 on Ismailia canal.

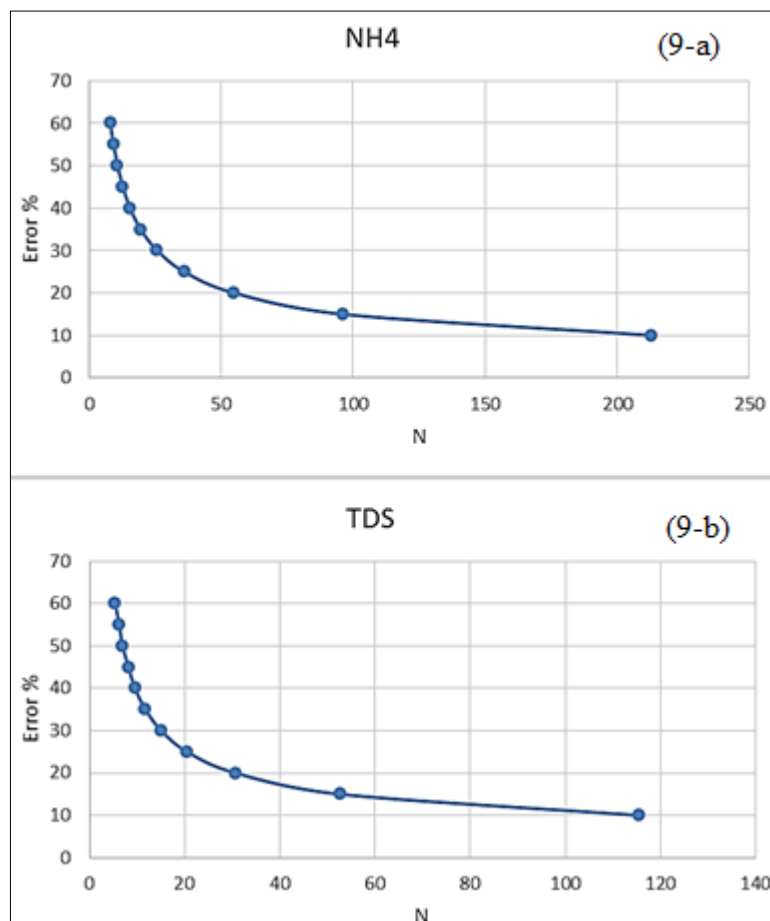


Figure (9): the curves (9-a) and (9-b) show the number of samples/year versus expected error for NH4 and TDS respectively at EI12 on Ismaillia canal

4. CONCLUSIONS

According to the results of Bahr Shebeen canal at the two sites of (MI09 and MI12), it seems that at number of samples ($n=12$) per year obtained for the 95% significant level of the WQ parameters (pH, TDS and DO) have an acceptable error percentage of (11% to 15%). Reversely, the WQ parameters (BOD, COD, TSS, NO_3 , NH_4 , Fe and TDS) need more than 12 samples per year at the same number of samples to shrink the error percentage in range (10% to 25%).

Regarding Ismaillia Canal at the two sites of (EI10 and EI12), the WQ parameters (BOD, COD, pH and DO) have an acceptable error percentage of (11% to 15%) obtained for the 95% significant level at number of samples ($n=12$) per year. On the other hand, the WQ parameters (TSS, NO_3 , NH_4 , Fe and TDS) require more than 12 samples per year to be in the acceptable error range (10% to 25%).

It can be concluded that the enhancing of sampling frequency for water quality monitoring is important to reduce the error percentage depending sampling intervals. However, we must also bear in mind that no sampling will be perfectly accurate (from a measurement perspective) as virtually all scientific measurements have an associated level of uncertainty. In this paper the “Single Station, Single Variable” approach was used to illustrate the relationship between the error and the number of samples as described which is theoretically valid if the variance of the stationary component computed from different sampling intervals stabilizes after a certain sampling interval.

REFERENCES

1. Abdel-Gawad S. T., Kandil H. M. and Sadek T. M., (2004). Water Scarcity Prospects in Egypt 2000–2050, in *Environmental Challenges in the Mediterranean 2000–2050*, ed. A. Marquina, Kluwer Academic Publishers, Dordrecht, , pp. 187–203.
2. Asit K. Biswas, Zuo Dakang, James E. Nickum, Liu Changming (1983), *Water Resources Series, Volume 3*, United Nations University.
3. Dijkman J. P. M., (1993). *Environmental Action Plan of Egypt, A Working Paper on Water Resources*, Directorate of General International Cooperation, Ministry of Foreign Affairs, The Netherlands.
4. *Egypt Year book (1995)*
(<http://www.sis.gov.eg/public/yearbook/YEARBOOK/CHAP07.HTM>)
5. Frenken K. (2005), *Irrigation in Africa in Figures, Aquastat Survey (2005)*, Food & Agriculture Org, Rome, Italy, p. 88.
6. Harmancioglu, N.B. and Tokgoz, S. (1999). Selection of sampling frequencies in water quality monitoring network design. *Journal of Water Pollution Control*, 5(1), 9 - 20.
7. Khalil, B. and T.B.M.J. Ouarda (2009). Statistical approaches used to assess and redesign surface water quality monitoring networks, *Journal of Environmental Monitoring*, 11, 1915 - 1929.
8. Khalil B., C. Ou, S. Proulx-McInnis, A. St-Hilaire and E. Zanacic (2014). Statistical Analyses of the adequacy of the surface water quality monitoring network in Saskatchewan. *Water Air and Soil Pollution*, 225 (10): 2128
9. MWRI, Ministry of Water Resources and Irrigation, (1997). *Review of Egypt's Water Policies, Strengthening the Planning Sector Project*, Ministry of Water Resources and Irrigation, Cairo, Egypt.
10. Sanders, T.G. and Adrian, D.D. (1978). Sampling Frequency for River Quality Monitoring. *Water Resources Research*, 14, 569 - 576.
11. Sanders, T.G., Ward, R.C., Loftis, J.C., Steele, T.D., Adrian, D.D. and Yevjevich, V. (1983). *Design of Networks for Monitoring Water Quality*. Water Resources Publications, Littleton, Colorado, 328 pp.
12. Strobl, R.O., Robillard, P.D., 2008. Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management* 87, 639–648.
13. Tokgoz, S. (1992). *Temporal design of water quality monitoring networks*, Master of Science thesis in Civil Engineering, DokuzEylul University, Graduate School of Natural and Applied Science, Izmir.