Journal of Applied Sciences Research, 5(8): 1067-1076, 2009

Using Regression Analysis to Estimate Water Quality Constituents in Bahr El Baqar Drain

Reham El-Korashey

Researcher, Central Laboratory for Environmental Quality Monitoring (CLEQM),
National Water Research Center (NWRC), Cairo, Egypt

Abstract: Regression analysis of water-quality data collected in 2004-06 was used to estimate concentrations for total nitrogen, ammonia, sodium, chloride, calcium, and biochemical oxygen demand. The explanatory variables examined for regression relations were monthly monitored properties of water —nitrate, electrical conductivity, chloride and chemical oxygen demand. For the conditions observed in 2004-06, nitrate was a significant explanatory variable for total nitrogen and ammonia estimated constituents, electrical conductivity was a significant explanatory variable for sodium and chloride estimated constituents, chloride was a significant explanatory variable for calcium estimated constituents, while chemical oxygen demand was a significant explanatory variable for biochemical oxygen demand estimated constituents. Water temperature pH, and dissolved oxygen were not statistically significant explanatory variables for any of the constituents in this study. The regression equations were evaluated using common measures of variability, including R2, or the proportion of variability in the estimated constituent explained by the regression equation. R2 values ranged from 68.0% for calcium concentration to 98.2% for total nitrogen concentration. By applying the developed regression equations on the fourth year of data collection 2007 to calculate the estimated concentrations and errors associated with these concentrations and calculating the median relative difference percentage (RDP) between measured constituent concentration and the constituent concentration estimated by the regression equations, the RPD values ranged from 0.82 for total nitrogen to 21.59 for chloride.

Key words: Regression analysis, water quality, MINTAB, Bahr El Baqar Drain

INTRODUCTION

In the past, to determine concentrations of pollutants in a stream, it was necessary to manually collect samples and send them to a laboratory for analysis. This procedure took time, and when human health is a concern, immediate information is necessary.

Estimation of water-quality constituents on the basis of surrogates provides several benefits. Although periodic water samples are collected manually and analyzed, the delay associated with laboratory analysis does not permit immediate identification of undesirable levels of constituents.

Regression equations can be used to estimate constituent concentrations. Constituent concentrations can be used by water-quality managers for comparison of current water-quality conditions to water-quality standards. Examination of stream flow and physical properties of water that act as surrogates for constituents of interest also helps optimize visits for the collection of water-quality samples.

MATERIALS AND METHODS

Site Characteristics: The Bahr El Baqar Drain system is located in the eastern part of the Nile delta and runs for some 170 km from Cairo to Lake Manzala. The agricultural area served by Bahr EI Baqar drain and its tributaries is about 317,000 ha. The BBD system is shown in Figure (1). It consists of a main drain that collects the effluent from two secondary drains: Bilbeis Drain and Qalyubya Drain. The two drains collect water from two other drains, respectively Gabal El Asfar and Shebeen. The drain system is frequently used to convey raw industrial and municipal wastewater. The main industrial area, which has an impact on the quality of industrial water in the Bahr El Baqar drain system, is Shobra EI Kheima. The industrial activities pursued in this area have not been changed over the last 10 years. These industrial activities include metal production, food processing, detergent and soap manufacturing, textile finishing, and Paper production[1,2,3]

In this study, Bahr El Baqar was shown as it's considered the most polluted of the drains which carries a mixture of treated and untreated waste water.

Corresponding Author: Reham El-Korashey, Researcher, Central Laboratory for Environmental Quality Monitoring (CLEQM), National Water Research Center (NWRC), Cairo, Egypt
E-mail: Reham_korashey@yahoo.com

Fig. 1: Bahr El Baqar Drain System

The drain is anoxic over its entire length [4]. It needs continous monitoring for wide range of parameters that's why develpoing special regression equations for this drain will help minimizing the monitoring time and cost.

Sample Collection and Analysis: Water-quality samples were collected manually by the Drainage Research Institute (DRI) according to methods described in *"Standard Method for Examination of Water and Wastewater"* [5]. During the collection of water-quality samples, physical properties of water were measured, including specific conductance, pH, temperature, turbidity, and dissolved oxygen.

The manual samples were collected monthly starting January 2004 through December 2006 and send to the Central Laboratory for Environmental Quality Monitoring (CLEQM) laboratory where it has been analyzed for total coliform, fecal coliform, biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), total volatile solids (TVS), nitrate (NO3), sulphate (SO4), chloride (Cl), ammonia (NH3), total phosphorus, heavy metals, alkalinity, turbidity and major cations.

Regression Analysis: Simple linear regression examines the linear relationship between two continuous variables: one response (y) and one predictor (x). When the two variables are related, it is possible to predict a response value from a predictor value with better than chance accuracy.

Regression provides the line that "best" fits the data. This line can then be used to:

- examine how the response variable changes as the predictor variable changes

- predict the value of a response variable (y) for any predictor variable (x)

The method used to draw this "best line" is called the least-squares criterion. The least-squares criterion requires that the best-fitting regression line is the one with the smallest sum of the squared error terms (the distance of the points from the line).

The rationale and methodology for expressing water-quality constituent concentrations in terms of other surrogate constituents or physical properties in a regression equation are explained in Helsel and Hirsch [6]. Helsel and Hirsch also detail the computations for regression estimation and identify measures commonly used to evaluate regression equations, including mean square error, standard deviation, and coefficient of multiple determination R2.

Minitab statistical software is a comprehensive statistical and graphical analysis software package trusted by quality professionals and educators around the world. It is the leading package used in Six Sigma and other quality improvement projects, and is widely known for its comprehensive collection of methods, reliability, and unsurpassed ease-of-use. Minitab 14 software was used to develop regression equations between constituents in concern.

Regression analysis of water-quality data collected in 2004-2006 was used to estimate concentrations for total nitrogen, ammonia, sodium, chloride, calcium, and biochemical oxygen demand.

To test the developed regression equations from the first 3 years of data collection (2004-2006), the equations were applied to the fourth year of data collection 2007 to calculate the estimated concentrations and errors associated with these concentrations.

As an indicator of the ability of the regression relations to estimate constituent concentrations, the measured concentrations were compared to the concentrations estimated by the regression relations by calculating relative percentage difference (RPDs) using the following equation:

$$RPD = \left| \frac{E - M}{M} \times 100 \right|$$

Where; **E:** is the constituent concentration estimated from the regression equation **M:** is the measured constituent concentration

RESULTS AND DISCUSSIONS

For the conditions observed in 2004-2006, specific conductance was a significant explanatory variable for sodium and chloride. Nitrate was a significant explanatory variable for total nitrogen and ammonia. Chemical oxygen demand was a significant explanatory variable for biochemical oxygen demand. Chloride was a significant explanatory variable for calcium.

Regression relations between constituents of concern and surrogate properties were examined, and a regression equation was developed for each constituent using a surrogate variable. The regression equations and each equation’s associated with R2, adjusted R2, pearson correlation P and median relative percentage difference (RPD) are listed in table (1). A discussion of each constituent and the associated regression equations follows.

The regression equations were evaluated using common measures of variability, including *R*2, or the proportion of variability in the estimated constituent explained by the regression equation and pearson correlation *P*. *R*2 values ranged from 68.0 % for calcium to 98.2 % for total nitrogen concentration, while *P* values ranged from 0.825 for calcium to 0.991 for total nitrogen.

To test the regression equations developed from the first 3 years of data collection (2004-2006), the equations were applied to the fourth year of data collection (2007) to calculate estimated water quality constituent concentrations and errors associated with these concentrations and calculating the median relative percentage difference (RPD) between measured constituent concentration and the constituent concentration estimated by the regression equations where it ranged from 0.82 for total nitrogen to 21.59 for chloride.

Total Nitrogen: Nitrogen is of vital importance in plant and animal nutrition and occurs in water as nitrite or nitrate anions, ammonia cations, and in other forms such as cyanide. Total nitrogen concentration is related to land use and sources of nitrogen include precipitation, soil leeching, agricultural runoff, and waste disposal[7]. Nitrogen levels may fluctuate seasonally, with high concentrations occurring during high flow when there is drainage from cultivated fields or feedlots [8] and with an autumn increase due to the decomposition of leaves [7]. For Bahr El Baqar drainage system, nitrate concentrations were chosen as explanatory variable for estimating total nitrogen. The range of nitrate values was 1.1 to 39.4 mg/l. The regression equation was:

$TN = 0.0556 + 1.308\ NO_3$ Equation (1)

Where R2 for this equation is 98.2% with pearson correlation of 0.991

Table 1: Regression equations for estimates of total nitrogen, ammonia, sodium, chloride, calcium and biochemical oxygen demand in Bahr El Baqar Drain

Constituent	n	Equation	P	R2 %	R2 %(adj)	Median RPD
Total Nitrogen	405	TN=0.0556+1.308 NO3	0.991	98.2	98.2	0.82
Ammonia	405	NH3=-0.0118+0.295 NO3	0.988	97.7	97.7	2.75
Sodium	504	Na=-3.650+8.024 EC	0.969	93.8	93.8	18.45
Chloride	504	Cl=-4.224+7.531 EC	0.965	93.1	93.1	21.59
BOD	485	BOD=2.71+0.671 COD	0.962	92.6	92.6	10.04
Calcium	502	Ca=3.268+0.1563 Cl	0.825	68.0	67.9	16.35

[n, number of samples used to develop regression equation; R2, coefficient of multiple determination; P, pearson correlation; RPD, relative percentage difference; TN, total nitrogen; NO3, nitrate; NH3, ammonia; EC, electrical conductivity; Cl, chloride; Ca, calcium; BOD, biochemical oxygen demand; COD, chemical oxygen demand]

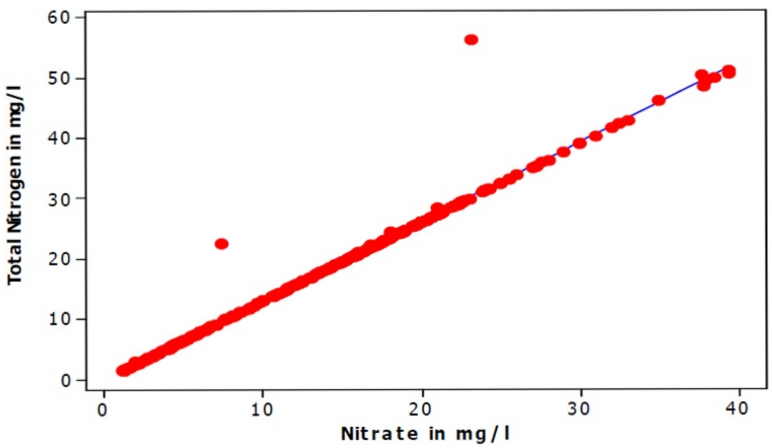

Fig. 2: Fitted line plot for total nitrogen regression equation

The measured total nitrogen concentrations on 2007 were compared to estimated total nitrogen concentrations calculated by equation (1) graphically and by calculating the RPD which was 0.82 (figure 3).

Ammonia: Nitrogen occurs in natural waters as nitrate (NO3), nitrite (NO2), ammonia (NH3), and organically bound nitrogen. As aquatic plants and animals die, bacteria break down large protein molecules containing nitrogen into ammonia. Ammonia is then oxidized by specialized bacteria to form nitrites and nitrates. Sewage is the main source of ammonia added by humans to rivers. The ammonia arises mostly from the hydrolysis of urea in urine, but additional ammonia is generated by the decomposition of other nitrogenous materials in sewage. In a flowing stream, the presence of ammonia in high concentrations indicates recent pollution. Sewage may be entering the water somewhere in the vicinity [14]. For Bahr El Baqar drainage system, nitrate concentrations were chosen as explanatory variable for estimating ammonia. The range of nitrate values was 1.1 to 39.4 mg/l. The regression equation was:

$NH_3 = - 0.01182 + 0.295\ NO_3$ Equation (2)

Where R2 for this equation is 97.7% with pearson correlation of 0.988

The measured ammonia concentrations on 2007 were compared to estimated ammonia concentrations calculated by equation (2) graphically and by calculating the RPD which was 2.75 (figure 5).

Sodium: Sodium in drinking water may be a concern for individuals on sodium restricted diets and high concentrations of sodium may make water unsuitable for irrigation [9]. Previous studies have shown a positive linear relation between sodium and specific conductance, as well as multiple regression relations that include pH and stream flow in addition to specific conductance[10,11,12]. For Bahr El Baqar drainage system, electrical conductivity was chosen as explanatory variable for estimating sodium. The range of electrical conductivity values was 0.41 to 6.92 mS/cm. The regression equation was:

$Na = - 3.650 + 8.024$ Field EC Equation (3)

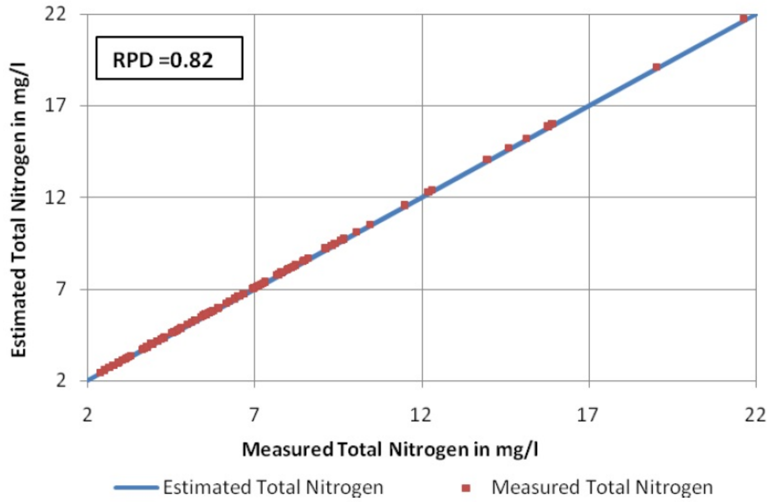

Fig. 3: Comparison of measured and estimated total nitrogen concentrations in Bahr El-Baqar Drain.

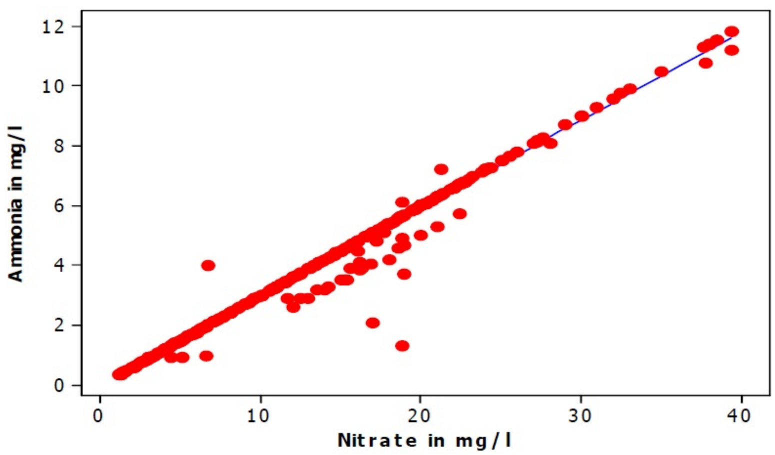

Fig. 4: Fitted line plot for ammonia regression equation

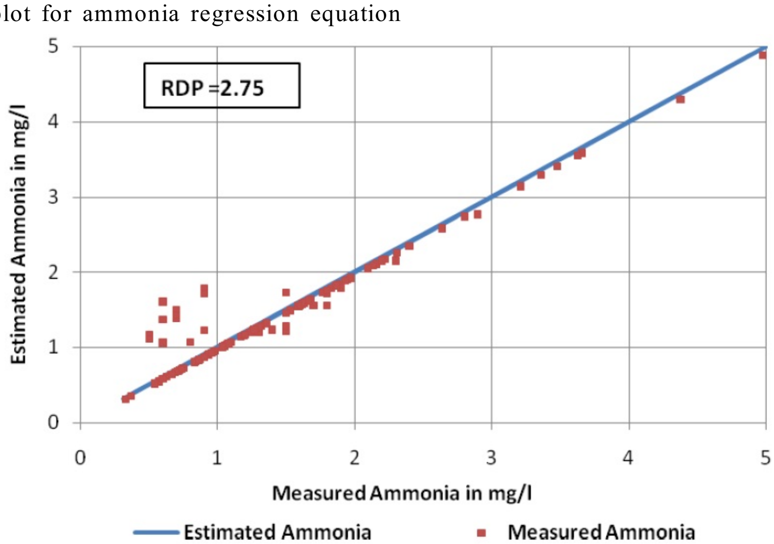

Fig. 5: Comparison of measured and estimated ammonia concentrations in Bahr El Baqar Drain.

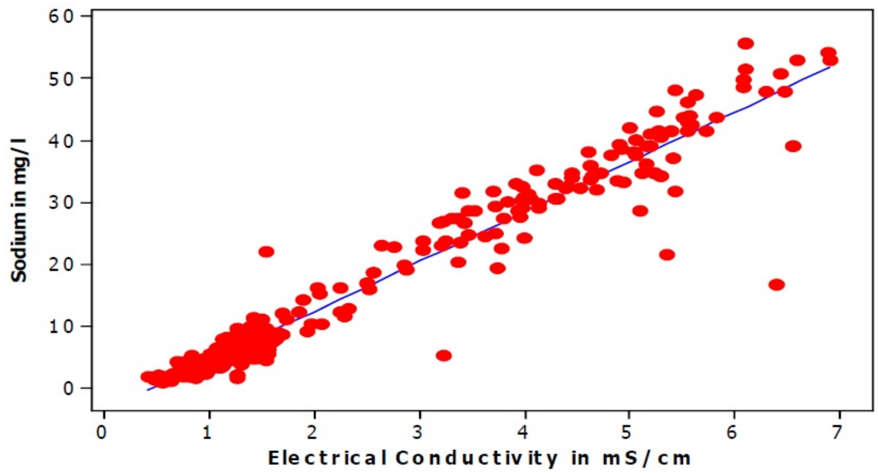

Fig. 6: Fitted line plot for sodium regression equation

Where R2 for this equation is 93.8% with pearson correlation of 0.969

The measured sodium concentrations on 2007 were compared to estimated sodium concentrations calculated by equation (3) graphically and by calculating the RPD which was 18.5 (figure 7).

Chloride: Chloride is present in virtually all rocks and is present in the tissues of all plants and animals. In surface water, concentrations of chloride range from less than 1 to more than 280,000 mg/L, the extreme value representative of the Dead Sea [13], and in most streams the concentration is less than that of sulfate [8]. Chloride is generally conservative (unchanging) in water; therefore, its circulation in the hydrologic cycle is mostly through physical processes [8]. Chloride in streams may originate from the weathering of rocks, ground-water inflow, precipitation, and municipal and industrial effluent [13,8]. The presence of too much chloride can have undesirable effects in drinking water; consumers may find the taste objectionable, and pipes in hot water systems may corrode [9]. Chloride is a negatively charged ionic species that makes water conductive. Therefore, as chloride concentrations increase, the specific conductance increases. Chloride concentrations may be higher in the winter when a higher percentage of stream flow is from ground-water discharge, and sudden increases in chloride may indicate pollution in the stream [8]. For Bahr El Baqar drainage system, electrical conductivity was chosen as explanatory variable for estimating chloride. The range of electrical conductivity values was 0.41 to 6.92 mS/cm. The regression equation was:

Cl = - 4.224 + 7.531 Field EC   Equation (4)

Where R2 for this equation is 93.1% with pearson correlation of 0.965

The measured chloride concentrations on 2007 were compared to estimated chloride concentrations calculated by equation (4) graphically and by calculating the RPD which was 21.6 (figure 9).

Calcium: Calcium contributes to hardness, is “a major component of the solutes in most natural waters,” and it is generally “the predominant cation in river water” [8]. Previous studies have shown a positive linear relation between calcium and specific conductance [10,11,12]. For Bahr El Baqar drainage system, Chloride was chosen as explanatory variable for estimating Calcium. The range of chloride values was 0.71 to 50.67 mg/l. The regression equation was:

Ca = 3.268 + 0.1563 Cl Equation (5)

Where R2 for this equation is 68.0% with pearson correlation of 0.825

The measured calcium concentrations on 2007 were compared to estimated calcium concentrations calculated by equation (5) graphically and by calculating the RPD which was 16.4 (figure 11).

Biochemical Oxygen Demand: Biochemical oxygen demand is a measure of the quantity of oxygen used by microorganisms (e.g., aerobic bacteria) in the oxidation of organic matter. Natural sources of organic matter include plant decay and leaf fall. However, plant growth and decay may be unnaturally accelerated when nutrients and sunlight are overly abundant due to human influence. Urban runoff carries pet wastes from streets and sidewalks; nutrients from lawn fertilizers; leaves, grass clippings, and paper from residential areas, which increase oxygen demand. Oxygen consumed in the decomposition process robs other aquatic organisms of the oxygen they need to live.

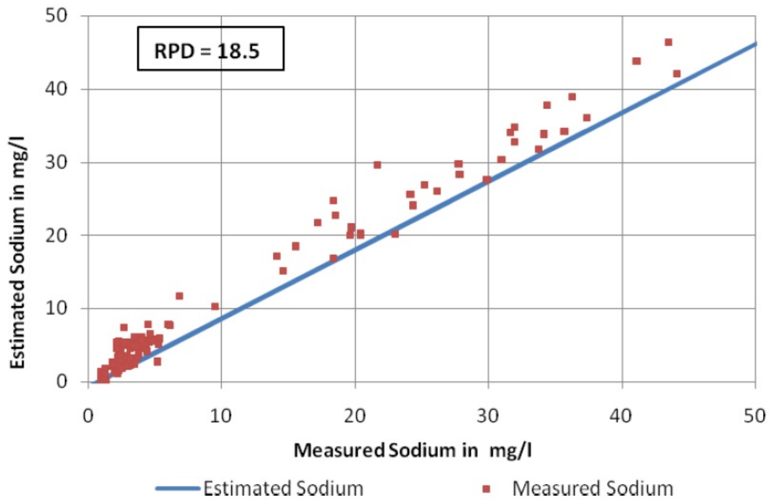

Fig. 7: Comparison of measured and estimated sodium concentrations in Bahr El Baqar Drain.

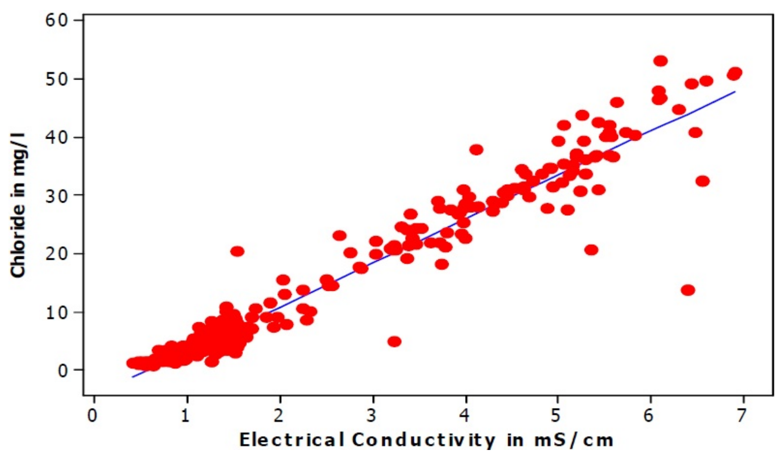

Fig. 8: Fitted line plot for chloride regression equation

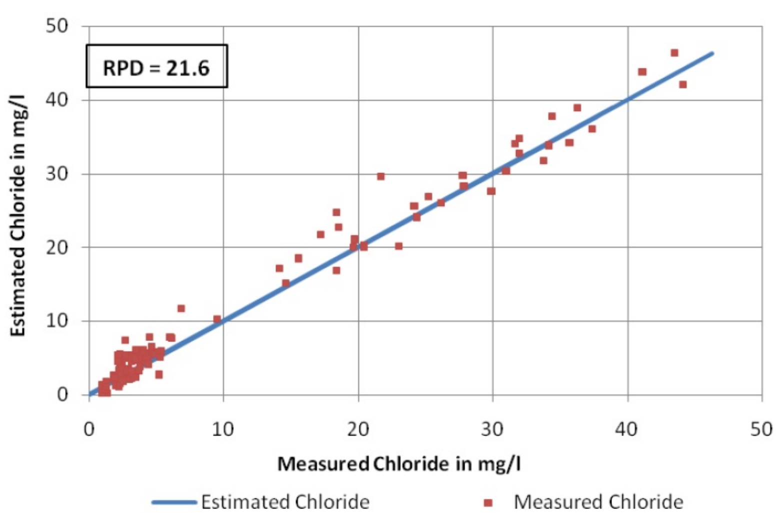

Fig. 9: Comparison of measured and estimated chloride concentrations in Bahr El Baqar Drain.

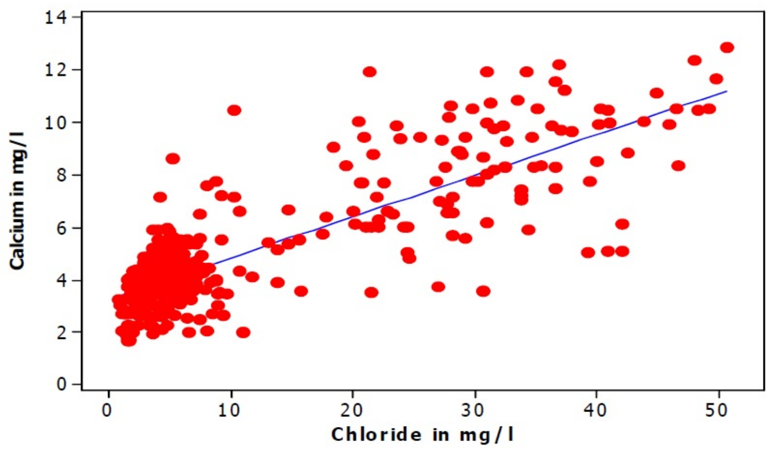

Fig. 10: Fitted line plot for calcium regression equation

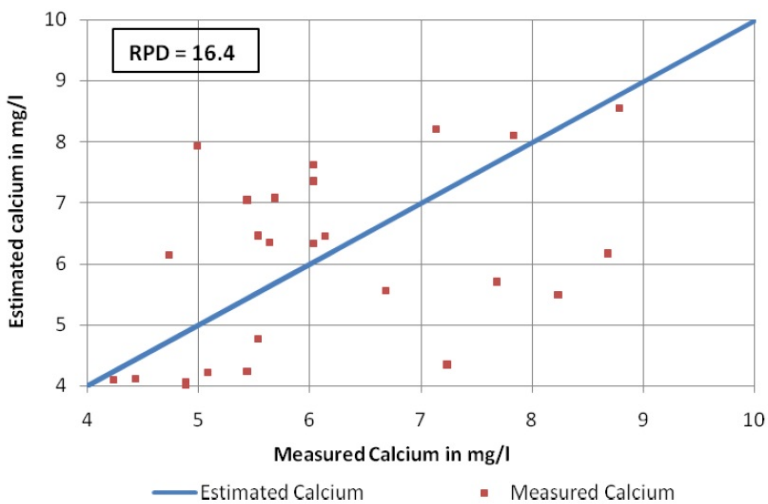

Fig. 11: Comparison of measured and estimated calcium concentrations in Bahr El Baqar Drain.

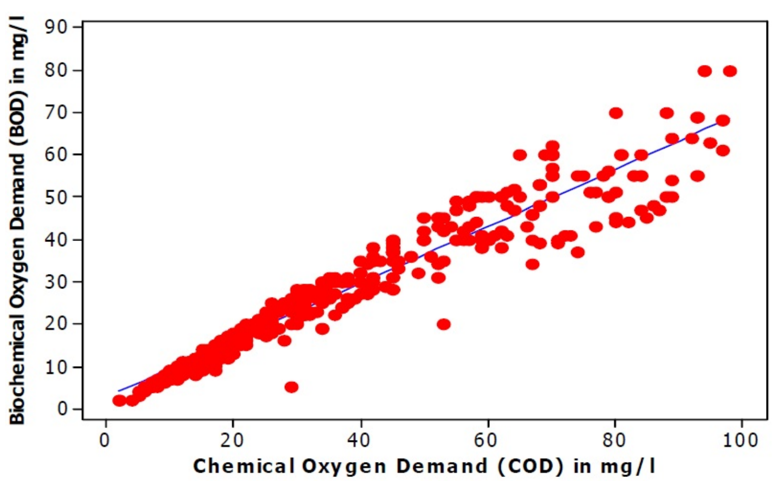

Fig. 12: Fitted line plot for BOD regression equation

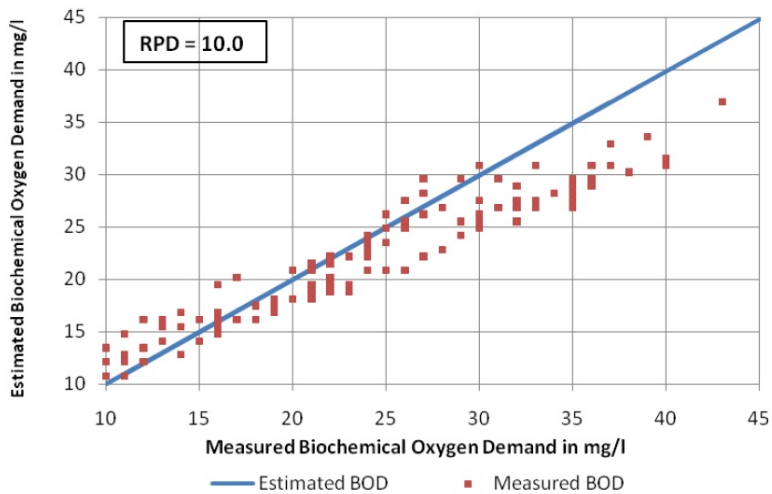

Fig. 13: Comparison of measured and estimated BOD concentrations in Bahr El Baqar Drain

Organisms that are more tolerant of lower dissolved oxygen levels may replace a diversity of more sensitive organisms [14]. Chemical oxygen demand (COD) was chosen as explanatory variable for estimating biochemical oxygen demand (BOD). The range of chemical oxygen demand values was 2 to 98 mg/l. The regression equation was:

BOD = 2.71 + 0.671 COD Equation (6)

Where R2 for this equation is 92.6% with pearson correlation of 0.962

The measured BOD concentrations on 2007 were compared to estimated BOD concentrations calculated by equation (6) graphically and by calculating the RPD which was 10.0 (figure 13).

From the previous data it can concluded that the use of regression equations to estimate constituent concentrations provides timely water-quality information to resource managers that are otherwise not available. The regression relations may be used to continuously estimate constituent concentrations in Bahr El Baqar Drainage System and these estimates may be used to continuously estimate concentration loads. The regression equations presented in this study are site specific and apply only to Bahr El Baqar Drainage System.

REFERENCES

1. African Development Bank, 2008. Gabal El Asfar Wastewater treatment plant environmental and social impact assessment, project number: P-EG-E00-001, pp: 3-4.
2. El Baz, A., 2003. Environmental System Analysis and Management, Ph.D. Thesis, Zagazig University, Egypt.
3. El Monayeri El O., S. Bayoumi and A. Khalifa, 2006. Enhancement of self-purification of streams using stepped aeration. Tenth International Water Technology Conference, IWTC10 2006, Alexandria, Egypt.
4. UNDP., 1997. Lake Manzala Engineered Wetland, United Nations Development Program, Project Number: EGY/93/G31.
5. American Public Health Association (APHA), 1998. Standard methods for the examination of water and wastewater 20th Ed., Washington D.C.
6. Helsel, D.R. and R.M. Hirsch, 1995. Statistical Methods in Water Resources, Studies in Environmental Sciences 49. Elsevier, Amsterdam.
7. Allan, J.D., 1995. Stream ecology-structure and function of running waters: London, Chapman & Hall, pp: 388.
8. Hem, J.D., 1985. Study and interpretation of the chemical characteristics of natural water, (3rd ed.): U.S. Geological Survey Water-Supply Paper 2254, pp: 264.
9. North Dakota Department of Health, 2006. How minerals affect water supplies: Miscellaneous publication, pp: 4.
10. Christensen, V.G., A.C. Ziegler, P.P. Rasmussen, and Jian, Xiaodong, 2003. Continuous real-time water-quality monitoring of Kansas streams, in Proceedings of 2003 Spring Specialty Conference on Agricultural Hydrology and Water Quality, May 12-14, 2003, Kansas City, MO.: Middleburg, VA., American Water Resources Association, AWRA Technical Publication Series No. TPS-03-1.
11. Rasmussen, T.J., A.C. Ziegler and P.P. Rasmussen, 2005. Estimation of constituent concentrations, densities, loads, and yields in lower Kansas River, northeast Kansas, using regression models and

continuous water-quality monitoring, January 2000 through December 2003: U.S. Geological Survey Scientific Investigations Report 2005-5165, pp: 126.
12. Christensen, V.G., J.L. Graham, C.R. Milligan, L.M. Pope and A.C. Ziegler, 2006. Water quality and relation to taste-and-odor compounds in the North Fork Ninnescah River and Cheney Reservoir, south-central Kansas, 1997-2003: U.S. Geological Survey Scientific Investigations Report 2006-5095, pp: 43.
13. Ryberg, K.R., 2006. Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in The Red River of the North, Fargo, north Dakota, 2003-05 U.S. Geological Survey Water Resources Investigations Report 2006-5241, pp: 35.
14. Ryberg, K.R., 2007. Continuous water-quality monitoring and regression analysis to estimate constituent concentrations and loads in the Sheyenne River, North Dakota, 1980–2006: U.S. Geological Survey Scientific Investigations Report 2007–5153, pp: 22.