

DEVELOPMENT OF A MODEL FOR THE ESTIMATION OF WATER TREATMENT COST: A DATA MINING APPROACH

ABDULLAHI MOHAMMED EVUTI

ABSTRACT

A model based on data mining approach for the estimation of water treatment cost is presented. The model was developed using multivariable regression method based on the industrial data collected from Kaduna North water treatment plant for a period of one year. Comparison of the result from the simulation of the model and observed data shows a good prediction with a correlation coefficient of 0.9921.

Key words: water quality parameter, water treatment cost, data mining, modeling, multivariable regression, MATLAB, simulation, correlation coefficient.

1. INTRODUCTION

The need to know the monetary cost associated with water treatment is fundamental to a realistic planning approach for potable water supply. Traditionally, water in its natural state has been regarded as “free goods” of unlimited supply with zero cost at the point of supply. User pay for transfer costs relating to transport, treatment to meet quality requirement, and disposal of used water. Opportunity costs of water are generally ignored [1]. However, it is becoming increasingly obvious that water is not the “free goods” of classical economics, therefore it requires prices which reflect costs of provision and benefits in use [2]. For any given chemical process, the cost of production is needed to judge the viability of a project and to make choices between possible alternative processing schemes. These costs can be estimated from the flow sheet which gives the raw material and service requirement and the capital cost estimate [3]. However, water treatment cost varies with the amount of chemical used and depends also on the quality of raw water to be treated. The raw water quality has been found to depend on the water source, the season and the other human activities around the water source [4].

E-mail: evutimohd@yahoo.com

The costs of water supply include: (a) Operating cost such as personnel cost, energy cost, cost of treatment chemicals, cost of machinery and vehicles, maintenance of buildings and equipment, cost of external professional service like consultants, contractor and administrative overheads. (b) Financial cost such as depreciation cost and interest. (c) Taxes and various levies such as tax on profit, tax on assets, environmental levies, charges from government etc.

According to Olsson [5], water and wastewater systems are of ever-increasing complexity. The processes need to be better understood in order to achieve an improved final product quality while ensuring a safe and economic plant operation. Models can be developed to summarize these understandings. Modeling also serves as a complimentary approach to laboratory and pilot plant experimentation because it can simulate the dynamic behavior of a system. Millions of Naira is spent annually in Nigeria for treatment of water to meet the demand of her teeming population. The water industry is seeking ways to produce high quality water at a reduced cost. The operation of water treatment plants is significantly different from most manufacturing industrial operations because raw water sources are often subject to natural perturbations. Consequently, the water quality characteristics vary from period to period. This makes the

Abdullahi Mohammed Evuti is a lecturer in Department of Chemical Engineering, University of Abuja-Nigeria and currently a PhD student in the Department of Chemical Engineering, Universiti Teknologi, Malaysia, SKudai-Johor.

predictability of expenditure on chemicals and the overall water treatment cost challenging. An algorithm to precisely predict chemical dosages for optimum treatment using measured influent parameters does not exist at many water treatment plants [6].

Due to the advancement in information technology, data storage and retrieval in water treatment plant have become easier using supervisory control and data acquisition system (SCADA system). Therefore as a response to the general trend that the amount and complexity of available data are growing faster than the ability to analyze it, data mining and data driven modeling techniques have been developed and in fact data mining have been found to be a promising approach for modeling industrial applications [7]. Savic et al. [8] provides an excellent summary of the data mining and knowledge discovery techniques that have been used in the water industry.

Data mining is often regarded as one part of the broader problem of knowledge discovery. Knowledge Discovery in Databases (KDD) is defined as 'a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data' and data mining as 'exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns' [9, 10, 11]. Also, data mining is the application of specific algorithms to extract patterns from data [12]. In data-driven modeling, data characterizing a system are analyzed to look for connections between the system state variables without taking into account explicit knowledge of the physical behavior of the system. This approach is in contrast to physically based (or knowledge driven) modeling, where the aim is to describe the mechanistic behavior of the system [12].

The two main techniques of data mining are the classical data mining techniques and next generation data mining techniques. Classical data mining techniques include the nearest neighbor technique (K-NN), clustering and statistics while the next generation data mining techniques are decision and classification trees, artificial neural networks,

rule induction and genetic algorithms. The statistical technique such as multivariable regression is used for patterns discovering and for predictive models development [13]. The aim of the multiple regression is to predict the values of a continuous dependent variable Y from a set of continuous or binary independent variables (X_1, \dots, X_p) . A number of researchers have attempted to use the multi-variable regression (MVR) approach and artificial neural network (ANN) to predict the required coagulant doses in response to water quality changes [6, 14, 15, 16, 17, 18]. The U.S. Bureau of Reclamation (USBR) "Water Treatment Engineering and Research Group" [19] developed WaTER or "Water Treatment Estimation Routine" which is a model for estimating the cost of drinking water treatment. WaTER is an MS Excel program that is the basis for the Visual Basic program called "WTCost". However, reported works on the application of data mining to develop model to predict water treatment cost are scarce.

Kaduna North water treatment plant, Nigeria has for many decades documented the raw water quality, the treatment plant performance as well as the quantity of treatment chemicals used. However, most of the information contained in this data that would assist the operator in further optimization of his or her plant remains unused. This study is aimed using statistical procedures (multivariable regression) to develop a model that can be used to estimate the water treatment cost by using established relationship between water quality parameters such as pH, temperature, alkalinity, turbidity, coliform counts and required chemical dosages. This model can be used as a predictive tool for accurate and reliable budget estimate.

2. METHODOLOGY

2.1 FACILITY AND DATA DESCRIPTION

In this research data from Kaduna North water treatment plant is used. The plant which has an installed capacity of 150 million litres of water per day consists of the following units: raw water intake house and balancing tank, mixing chamber, clarifier, filter and clear water tank. The pH, temperature,

alkalinity, turbidity, coliform counts, chemical (alum, chlorine, lime) dosage and financial data were collected from the daily records kept for a period of two years starting from May to April the following year. This translates to 365 data points for each variable which is really enormous. However, weekly averages of 52 weeks per year were used due to cumbersome nature of the data. The data for the first year was used to develop a prediction model using multivariable regression and MATLAB while that of the second year were used to test the prediction performance of the model derived.

2.2 MODELING PROCESS

This process employs a combination of formal model of a cost equation or a cost function from production economics which requires that we have costs for all inputs and existing statistical models for determining the quantities of alum, chlorine and lime required for water treatment. The cost of water treatment is the sum of operating cost, depreciation cost and taxes/levies. Expressed mathematically as:

$$WTC = OPC + DPC + TAX \quad (1)$$

Where WTC = Water treatment cost (₦)

OPC = Operating cost (₦)

TAX = Tax (₦)

DPC = Depreciation cost (₦)

But water production is tax free and depreciation cost are usually fixed percentage, while the operating cost is the sum of personnel cost, energy cost, cost of chemicals, maintenance costs and administrative overheads.

Expressed mathematically as:

$$OPC = PC + EC + CC + MC + AC \quad (2)$$

Where PC = Personnel cost (₦)

EC = Energy cost (₦)

CC = Cost of chemicals (₦)

MC = Maintenance cost (₦)

AC = Administrative cost (₦)

Assuming PC, EC, MC, and AC are fixed expressed as:

$$PC + EC + MC + AC = FC \quad (3)$$

Substituting (3) into (2) gives

$$OPC = FC + CC \quad (4)$$

The cost of chemicals and other inputs (CC) is the product of the quantity used and unit price and can be expressed mathematically as:

$$CC = Q_1P_1 + Q_2P_2 + Q_3P_3 \quad (5)$$

Where Q_1 = Quantity of alum (kg)

P_1 = Price of alum (₦)

Q_2 = Quantity of chlorine (kg)

P_2 = Price of chlorine (₦)

Q_3 = Quantity of lime (kg)

P_3 = Price of lime (₦)

The quantity of chemical (Q) is a function of the water quality parameters such as temperature (T), turbidity (t), alkalinity (Alk), bacterial (coliform) count (B) and pH. Expressed as [16, 17 18]:

$$Q_1 = f(\text{pH}, T, \text{Alk}, t) \quad (6)$$

$$Q_2 = f(\text{pH}, T, B) \quad (7)$$

$$Q_3 = f(\text{pH}) \quad (8)$$

Substituting equation (5) into (4) and then into equation (1) gives the final model equation for the water treatment cost as:

$$WTC = FC + Q_1P_1 + Q_2P_2 + Q_3P_3 + DPC \quad (9)$$

While the values of Q_1 , Q_2 , Q_3 will be obtained from model solutions to equations (6), (7) and (8) respectively. Introducing the model solutions to equations (6), (7) and (8) developed using least square regression method and MATLAB as reported in [16, 17,18] respectively,

$$Q_1 = -0.01276 - \frac{142694.7784}{T} + \frac{122908.4905}{ALK} + 2451.64(10^{\Delta pH}) + 17.041t \text{ (kg)} \quad (10)$$

$$Q_2 = 0.001836 + \frac{0.03836}{T} + 0.000128(10^{-pH}) + 13.6782B \text{ (kg)} \quad (11)$$

$$Q_3 = -1.5402 + 1735.539(10^{-\Delta pH}) \text{ (kg)} \quad (12)$$

And substituting them (equations (10), (11) and (12)) into equation (9) gives

$$WTC = FC + \left[-0.01276 - \frac{142694.7784}{T} + \frac{122908.4905}{ALK} + 2451.64(10^{\Delta pH}) + 17.041t \right] P_1 + [0.001836 + \frac{0.03836}{T} + 0.000128(10^{-pH}) + 13.6782B] P_2 + [-1.5402 + 1735.539(10^{-\Delta pH})] P_3 + DPC \quad (13)$$

3. RESULTS AND DISCUSSION

Because water treatment cost depends on various water quality parameters, statistical mining technique was used due to its suitability in predictive model development. The water treatment cost model was developed by a

combination of the various models earlier developed for the determining the quantities of alum, chlorine and lime which were based on least square multivariable regression analysis and MATLAB solutions to the resultant model equations.

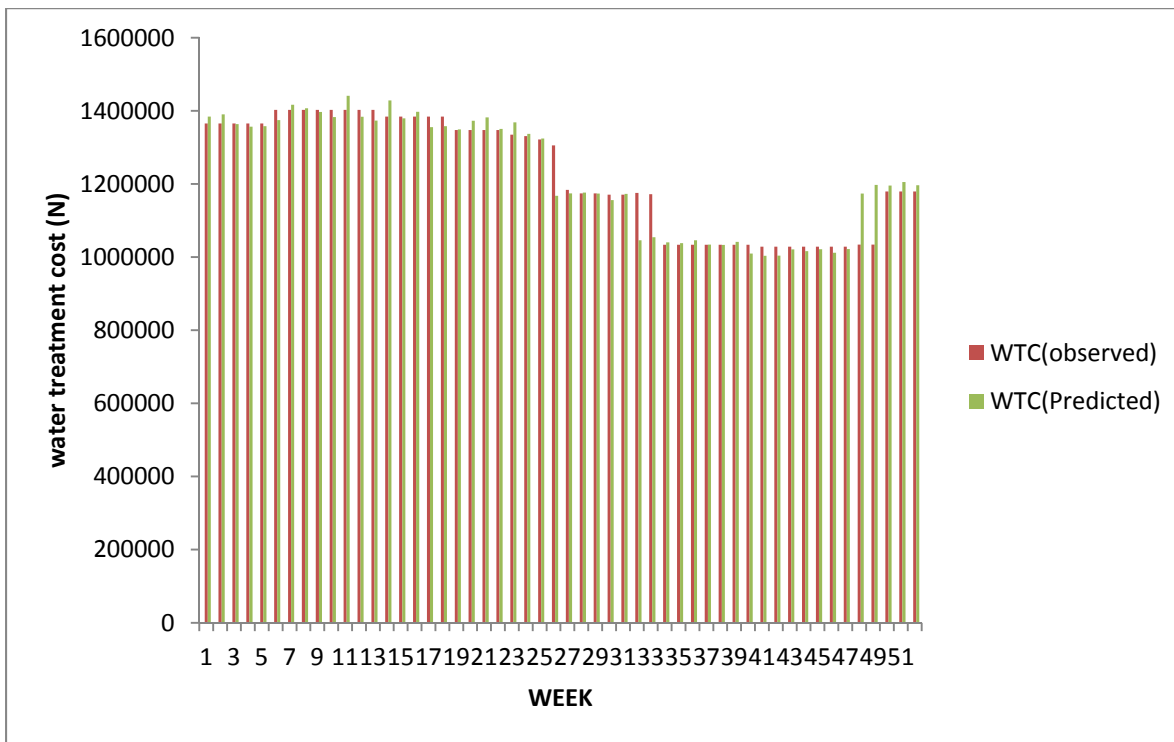


Figure 1: Comparison of the observed and predicted water treatment costs

Figure 1 shows the observed and the predicted data for the water treatment cost for a period of 52 weeks beginning from May to April of the following year. The model gives a good prediction of the water treatment cost with a correlation coefficient of 0.9921. Dearmont et al. [20] developed a model to estimate the cost of treating surface water based on turbidity or sediment load in the water supply using empirical approach to explain the per unit chemical treatment cost in terms of the quality of the raw water supply. However, the resultant model shows some biases in the coefficients due to a lack of treatment of other input items.

It was also observed from the chart that water treatment cost reduces towards dry season in Nigeria beginning from week 34 when the raw water is purer because

lesser quantities of alum, chlorine and lime are required for the treatment during this period. This agrees with the result of Gosh and Banerjee [21]. The seasonal changes in water quality are attributed to natural and anthropogenic inputs. The quantity of alum required in rainy season is usually higher because the water is more turbid due to water run off which adds to the main water body. They carry along with them sand, silt and other dissolved solids. Turbidity therefore is an important indicator because high turbidity interferes with chlorination and make the water unfit for human consumption. This is reflected in Figure 1 by a rise in water treatment cost from week 46 (April) when the rainy season begins in Nigeria.

4. CONCLUSION

With the advancement in information technology, data storage and retrieval in water treatment plant have become easier. As a result, enormous data have been generated and accumulated over a long period of time. There is therefore a critical need for an automated approach to effectively and efficiently extract the knowledge hidden in these large volumes of raw data. In this research data mining techniques have been successfully used to develop a model for determining the water treatment cost. Analysis of the observed and predicted data gave a correlation coefficient of 0.9921.

REFERENCES

- [1]. J. Morris, "Water policy: Economic theory and political reality," E and Spon publishers, London, pp 229, 1996.
- [2]. J.L. Brockman and S.N. Kulshreshthan, "The residential demand for water in Saskatchewan: A problem of price specification with block rate schedule," International water resources agency, New York, USA. pp 477-486, 1988.
- [3]. J.H Perry, "Chemical Engineers Handbook, 5th edition," Mc Graw Hill international book company, London, pp 25-33, 1983.
- [4]. G. Nikoladze, D. Mintis, and A. Kastalsky, "Water Treatment for publics and Industrial Supply,"MIR Publishers, Moscow, pp 1-250, 1996.
- [5]. G. Olsson, "Research needs in water modeling," proceeding of the international congress on environmental modeling and software, Barcelona, Catalonia, pp 12-15, 2008.
- [6]. A. Mirsepassi, "Application of Intelligent System for Water Treatment Plant Operation," Iranian Journal of Environmental Health Science and Engineering, vol. 1, No.2, pp 51-57. 2004.
- [7]. A. Kusiak, and X. Wei, "Prediction of methane production in wastewater treatment facility: a data-mining approach," Annals of operation research, Springer, DOI 10.1007/s10479-011-1037-6, 2011.
- [8]. D.A. Savic, J.W. Davidson, and R.B. Davis, "Data mining and knowledge discovery for the water industry," Savic and Walters *Ed*, 1999.
- [9]. V. Kumar, and M. Joshi, "Tutorial on high-performance data mining, at University of Minnesota" 1999.
- [10]. R. Kadyan, N. Arora and P. Chhabra, "An overview of data mining," International journal of scientific and engineering research, vol.3, issue 11, pp 1-7, November, 2012.
- [11]. R. Mahammad Shafi and P. Srinivas, "Data mining: A tool for enhancing business process in banking sector," International journal of scientific and engineering research, vol.3, issue 12, pp 1-7, December, 2012.
- [12]. D.J. Durrenmatt, "Data mining and data-driven modeling approaches to support wastewater treatment plant operation," an unpublished PhD thesis submitted to ETH ZURICH, pp, 1-10, 2011.
- [13]. M. Carbureanu, "Pollution Level Analysis of a Wastewater Treatment Plant Emissary using Data Mining," BULETINUL Universităţii Petrol – Gaze din Ploieşti, Seria Matematică - Informatică – Fizică, vol. LXII, No. 1, pp 69-78, 2010 .
- [14]. P. Naidoo, and J.J. Van der walt, "Artificial neural networks as a chemical dosing budgeting tool: The Rand Water Case Study," www.ewisa.co.za, pp 1-14, 2011.
- [15]. C.W. Baxter, Q. Zhang, S.J. Stanley, R. Sharraf, R-R.T Tupas and H.L. Stark, "Drinking water quality and treatment: the use of artificial neural networks," Canadian Journal of civil engineering, vol. 28, suppl. 1, pp 26-35, 2001.
- [16]. M.E. Abdullahi, and J.O. Odigire, "Development of mathematical model for determining the quantity of alum required for water treatment", Journal of engineering technology and industrial application, vol. 2, No. 2, 2006.
- [17]. M.E. Abdullahi and B.I. Abdulkarim, "Development of mathematical model for determining the quantity of chlorine required for water treatment," Journal of applied sciences research, vol. 6, No. 8, pp 1002-1007, 2010.
- [18]. M.E. Abdullahi, D.F. Aloko, G.A. Baba, and J. Mohammed, "Predictive model for lime dosage in water treatment plant," International Journal of Scientific and Research Publications, vol. 2, Issue 12, pp 1-5, 2012.
- [19]. US Department of the Interior, Bureau of Reclamation, "Water treatment estimation routine (WaTER)," 2006.

[20]. A.R. Gosh and R. Banerjee, "Qualitative Evaluation of the Damodar River water flowing over the Coal mines and Industrial area," International journal of scientific and research publication, vol. 2, No. 10, pp 1-6, 2012.

[21]. D. Dearmont, B.A Mc Carl, and D.A. Tolman, Cost of water treatment due to diminished water quality: A case study in Texas, Draft of paper in water resources research, vol. 34, No. 4, pp 849-854, 1998.