Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2019.24676

169 (2019) 173–180
November

Mathematical model to determine the required dosage of chlorine in the Bellavista rural water treatment plant in Peru

Fidel G. Aparicio Roque[a], Betsabe G. Blas Achic[b], Roger L. Moreno Jaimes[c,*], Juan M. Roque Gonzalez[d]

[a]Departamento de Ciencias Agrarias, Universidad Nacional Santiago Antúnez de Mayolo, Perú. Av. Centenario 200-Huaraz, Perú, Tel. +51945122204; email: fgar_58@hotmail.com
[b]Departamento de Estatistica, Universidade Federal de Pernambuco, Brasil. Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, 50670-901, Brasil, email: betsabe@de.ufpe.br
[c]Constructora y consultora Thiago's Morenos S.A.C.Psje. Señor Burgos S/N-Huaraz, Perú, Tel. +51956159493; email: rlmoreno2211@gmail.com
[d]Departamento de Ciencias Agrarias, Universidad Nacional Santiago Antúnez de Mayolo, Perú.Av. Centenario 200-Huaraz, Perú, Tel. +51995138836; email: jmrg5556@cip.org.pe

Received 25 June 2018; Accepted 8 July 2019

ABSTRACT

The goal of this work is to determine the quantity of chlorine required for water treatment at the Bellavista plant in Huaraz, Peru, by using a suitable regression model. It is shown that there exist linear relationship between chlorine and the covariates temperature, pH and coliform counts. We proposed the use of a multiple regression model with suitable transformations on both the dependent and independent variables and the model assumptions are verified by using diagnostic plots. The data set recorded from January to June 2016 in the Bellavista plant was used for the application section. The diagnostic plots for this data set show that the proposed model fits very well and it presents a high value of the coefficient of determination 0.9994.

Keywords: Regression model; Chlorine; Water treatment plant

1. Introduction

Water is essential for life and all people must have a satisfactory supply (sufficient, safe and accessible). Improving access to drinking water can provide tangible benefits for health. Every effort must be made to ensure that drinking water is as safe as possible. Safe drinking water (potable water), as defined in the Guidelines, does not cause any significant health risk when consumed over a lifetime [1].

The presence of *Escherichia coli* (*E. coli*) in water is a strong indication of recent contamination of wastewater or animal waste. It is important to keep in mind that *E. coli* and animal/human waste can enter our water in many different ways. For example, during rain and snow melt, the *E. coli* can be washed in rivers, streams, lakes or groundwater from the surface of the earth [2].

E.coli is a main bacterial indicator in drinking water and it is present in feces of warm-blooded animals at densities from 108 to 109 CFU/g of feces and comprises almost 95% of coliforms in feces [3].

Chlorination in water for human consumption is very important due to many waterborne diseases such as cholera,

---

* Corresponding author.

dysentery, among others; has declined almost completely in our population. The reaction of chlorine with the organic material in the water forms disinfection by-products, among them trihalomethanes (THMs), haloacetals, haloacetonitriles and haloketones, haloaldehydes, chlorinated phenols, etc. [4].

THMs were first detected in the water in the early 1970s and after many studies has been pointed out that there may be a relationship between the by-products of chlorination and adverse health effects by increasing the risk of bladder cancer and respiratory problems [5].

If chlorine is in the below required dose in treatment plants, water may retain microbiological agents or pathogenic microorganisms that threaten the health of the consumer due to aquatic diseases. If chlorine exceeds the required dose in treatment plants, it can be toxic to humans because it is an active ingredient and can react with different organic compounds, which increases the risk of producing THMs.

In [6] was proposed a regression model to determine the amounts of chlorine (Q) required for water treatment and it was considered predictive factors such as temperature (T), pH and coliform counts (C). The model discussed in [6] is:

$$Q = \beta_0 + \beta_1 \frac{1}{T} + \beta_2 \left(10^{-\mathrm{pH}}\right) + \beta_3 C + \varepsilon \qquad (1)$$

where ε is the error model with 0 mean and the unknown variance σ_2. β_0, β_1, β_2 and β_3 are unknown parameters.

Several studies have demonstrated that the effect of disinfecting chlorinated water under normal conditions and its application was essential in emergency situations to prevent the spread of gastrointestinal diseases. Epidemiological studies continue to evaluate the risk of bladder cancer due to the chlorination of water against high risk due to the consumption of water contaminated with pathogenic microorganisms.

The problem of water for human consumption in Peru has its origins with the emergence of the cholera epidemic in the country towards the beginning of 1991. The alarming and widespread nature of this epidemic, a disease supposedly overcome and that corresponded to a medieval past, not only raised the concern of the competent authorities, it also provoked the concern of the most important multilateral international cooperation agencies such as the World Bank, the European Union and the Inter-American Development Bank, who developed various support programs to control the deficient health situation of the marginal urban areas, the main focus and expansion area of this biblical evil. The different diagnoses of national and international experts indicated at that time that the affected localities were facing an unfortunate situation of "chronic health vulnerability" [7].

Due that the Bellavista treatment plant has diverse conditions of climate, hydrogaphy, mountains, glaciers, etc., the model proposed in [6] cannot applied directly.

In section 2, we present a description of the Bellavista treatment plant and statistical descriptive analysis with the data set from January to June 2016 (Appendix A1, Tables A1–A3). Section 3 presents the proposed model for data sets coming from the Bellavista treatment plant. Finally, conclusions are given in section 4.

2. Material and methods

2.1. Case study

The study was carried out in the south east of the center of the city of Huaraz, province department of Ancash, Peru. This has an area of 4,500 m^2 and it is located at the UTM coordinates: latitude 9°32'24.48"S, 77°31'08.20"W, and average altitude of 3,155 MAMSL (Fig. 1).

The study area includes the operating system of the treatment plant of the Chavín Services Provider Company, supplied by the Paria River with headquarters in the town of Huaraz, built in 1996 to treat 170 LPS. The population and sample studied include the Bellavista treatment plant (Fig. 2), which benefits approximately 85,457 inhabitants [8].

The unit of analysis was the filtered water. Water sampling was conducted from January 1, 2016 to June 30, 2016.

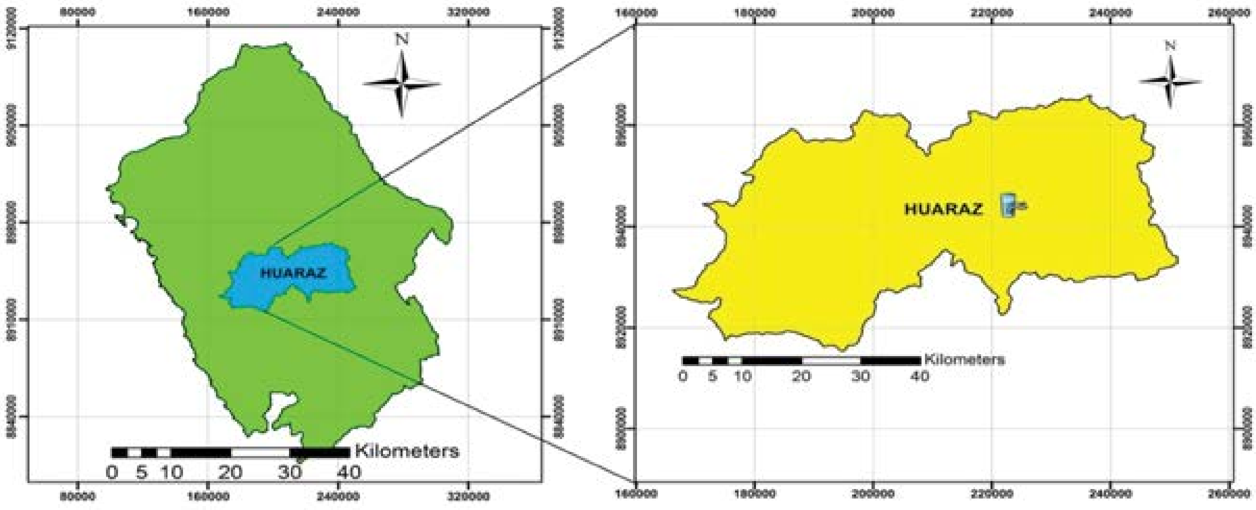

Fig. 1. Location of the studied area.

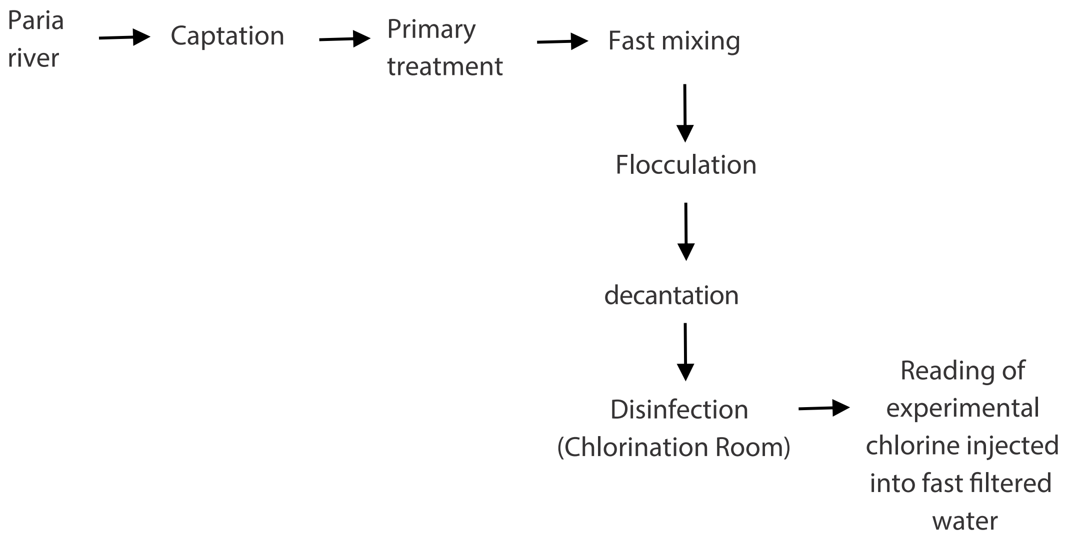

Fig. 2. Rapid filtration-type treatment plant for human consumption, Bellavista-Huaraz.

2.2. Data set from Bellavista treatment plant

The analysis of the quality of water for human consumption, takes into account statistical models to predict the physical and bacteriological treatment, such as alum, chlorine and lime, considering co-variables, for example, pH, temperature, turbidity, conductivity, coliform counts, iron, manganese, etc. Taking into account the cost of water treatment planning at the Bellavista plant, we recorded data on pH, temperature, coliform counts and chlorine (Appendix A1, Tables A1–A3).

The collection of the values of the independent variables, in this case represented by the microbiological (*E. Coli*) and physical (Temperature) parameters, was carried out following the specifications and the standard procedures which are present in the relevant protocols.

The standardized methods for the analysis of drinking water and wastewater, include international standards for the characterization of water quality: American Public Health Association (Method 9222) and American Water Works Association, Water Quality Permit Program, supplemented by technical guides or protocols of quality control [9].

In this case, the specific operating procedures with coding 03 (PECO) of the service provider company - Chavín SA, available in its operational unit laboratory where the present research work was carried out. The samples were obtained from raw water were collected in the water-source of the Bellavista treatment plant every day in the time period from 08:00 a.m. to 09:00 a.m., in the month of January to June 2016; and immediately after such collected samples were carried out to the chlorination laboratory, in which chlorine was added through the Regal 210 gas chlorinator.

Fig. 3 shows statistical description on this raw data set. The distribution of each variable is shown on the diagonal. The bivariate scatter plots with a fitted line are displayed on the bottom of the diagonal. The values of the correlation plus the significance level as stars are shown on the top of the diagonal. Each significance level is associated to a symbol as follow: the *p*-values 0 and 0.001 are related to the symbols "***" and "**", respectively. We observe that the distribution of chlorine, temperature and coliform are asymmetric and pH seems to follow a symmetric distribution. Chlorine presents a linear relationship with temperature and pH. Chlorine is high inversely correlated with temperature, but poorly correlated with pH.

2.3. Regression model for chlorine

This section discusses the requirement of chlorine in function of variables such as pH, temperature, iron and manganese ions, and microbial content (fecal bacteria-*E. coli* counts).

On our analysis for the Bellavista water plant, we consider the iron and manganese content is negligible. Also, we consider the covariates pH, temperature (T) and coliform count (C) to determine chlorine (Q) [6]. Our statistical model is defined as

$$Q = F_\theta\left(T, C, \text{pH}\right) + \varepsilon \tag{2}$$

where Q is the response variable (kg/24 h), the covariates are C (NMP/100 ml), T (°C) and pH. F_θ (.) is a linear function on a parameter vector θ and ε is the error model.

2.3.1. Effect of temperature

In section 2.2, the descriptive analysis we observe that chlorine value decreases linearly as temperature value increases.

2.3.2. Effect of coliform count

Coliform bacteria are present in the environment and feces of animals and humans. The commonest group is *E. coli*, if it is found in a water system it indicates recent fecal contamination, which may pose an immediate health risk to anyone who consumes the water. In the descriptive analysis we observe that the requirement of chlorine for disinfection increases with the coliform count and it is not linear.

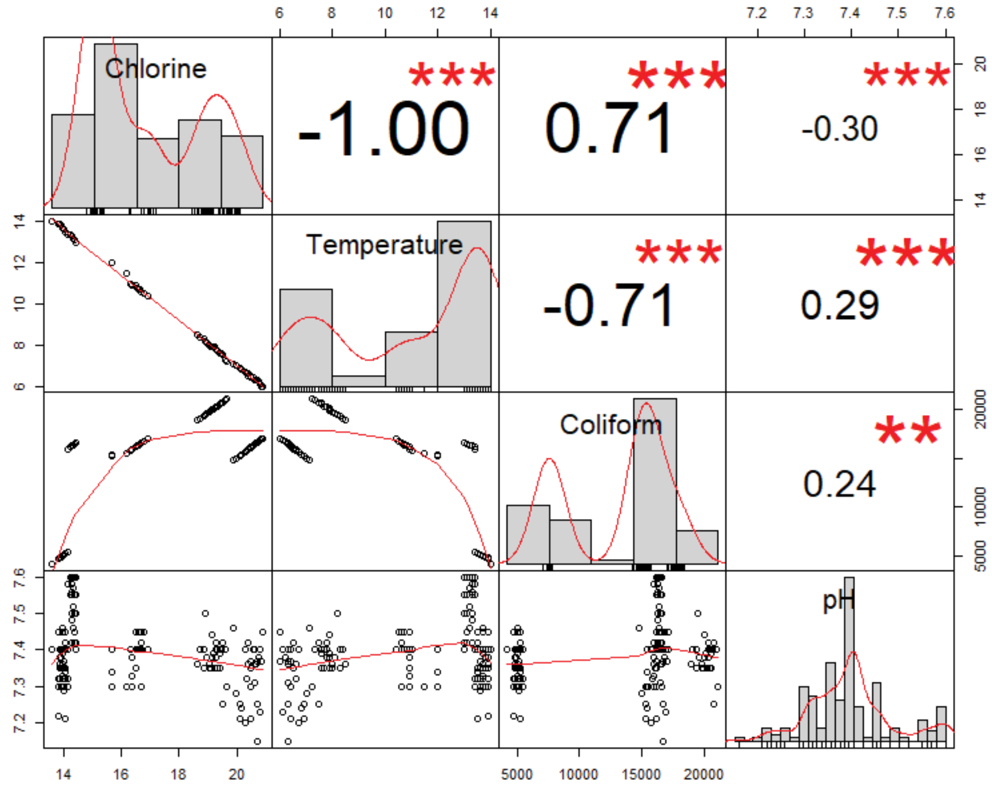

Fig. 3. Statistical description on raw data set from Bellavista plant.

2.3.3. Effect of pH

pH is an indicator of the acid or alkaline condition of water. It is mostly a result of natural geological conditions at the site and the type of minerals found in the local rock. The pH can also be affected by acid rain. The pH scale ranges from 0 to 14, with 7 being neutral. pH less than 7 is acidic while pH greater than 7 is alkaline (basic). Normal rainfall has a pH of about 5.6, slightly acidic due to carbon dioxide gas from the atmosphere.

2.4. Modeling

The following multiple regression models is proposed based on the data set from the water treatment Bellavista plant to predict the chlorine dosage for water treatment.

$$Q^3 = \beta_0 + \beta_1\left(\frac{1}{T}\right) + \beta_2\left(\frac{1}{T}\right)^2 + \beta_3\left(10^{-\mathrm{pH}}\right) + \beta_4 C^2 + \beta_5 C^3 + \varepsilon \qquad (3)$$

where ε follows normal distribution with 0 mean and the unknown variance σ_2. β_0, β_1, β_2, β_3, β_4 and β_5 are unknown parameters.

Note that our original covariates T and pH were transformed into $1/T$ and $10^{-\mathrm{pH}}$ variables, respectively, as suggested in [6].

The basic regression model without the polynomial terms and the cubic chlorine variable in (2), that is, $Q = \beta_0 + \beta_1 (1/T) + \beta_2 (10^{-\mathrm{pH}}) + \beta_3 C + \varepsilon$, was studied and the model assumptions were checked by looking diagnostic plots. This model presented violations on the linearity, normality and homoscedasticity assumptions. However, based on this basic model we developed the model (3), for which the model assumptions were tested with plots and they were successfully verified as shown in the next section. The model parameters are estimated by using the maximum likelihood method and the RStudio package was used for computation. The fitted model is given in Table 1.

Table 1 presents the coefficient estimates of our proposed model (3), estimate of the standard errors and the *p*-values (Pr(>|*t*|)). We observe that all the coefficients are statistically significant. The adjusted *R*-squared is 0.9994, which indicates that our model predicts very well responses for new observations.

Table 1
Coefficient estimates of the proposed model, estimate of the standard errors, and *p*-values.

Variables	Coefficient estimate	Standard error	Pr(>|*t*|)
Intercept	–6.297736e+03	1.166e+02	<2e–16***
$1/T$	1.458787e+05	2.452e+03	<2e–16***
$(1/T)^2$	–3.242517e+05	1.015e+04	<2e–16***
$10^{-\mathrm{pH}}$	1.936401e+09	6.716e+08	0.004427**
C^2	6.180055e–07	2.045e–07	0.002882**
C^3	–3.481494e–11	1.034e–11	0.000929***

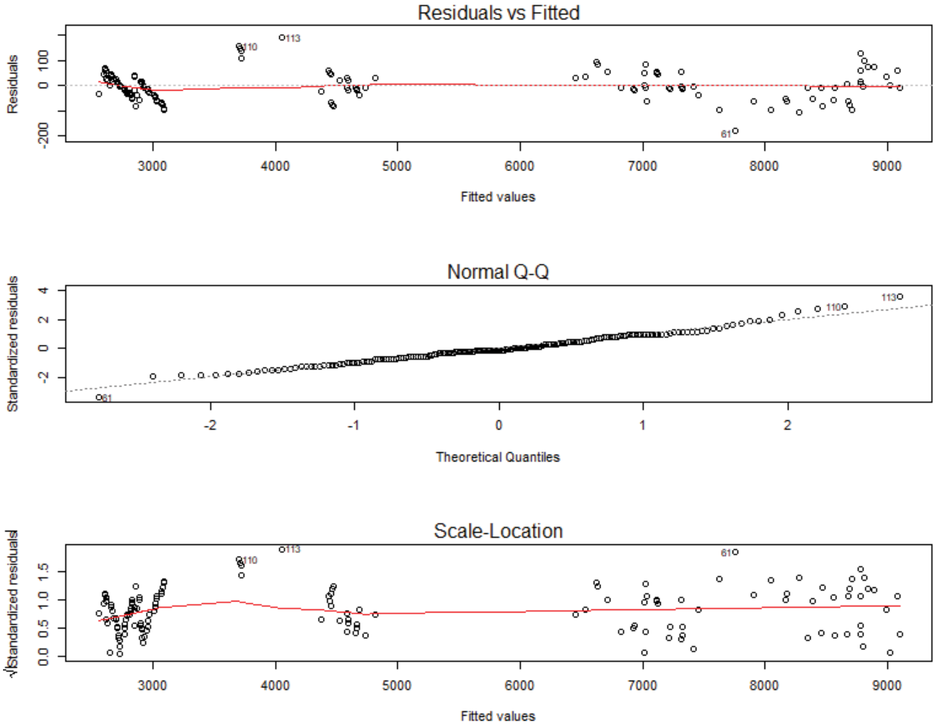

Fig. 4. Scatter plot for residuals vs. fitted values and normal *Q*–*Q* plot.

2.5. Diagnostic plots

The proposed model (3) makes some key assumptions: (a) there must be a linear relationship between the outcome variable (Q^3) and the independent variables. (b) Our model assumes that the residuals are normally distributed. (c) It is assumed homoscedasticity; this assumption states that the variance of error terms are similar across the values of the independent variables. In Fig. 4, the scatter plot residuals vs. fitted shows a horizontal line, without distinct patterns, that means there exists a linear relationship. The normal *Q*–*Q* plot shows that the residuals are normally distributed. In the plot scale-location, we observe a horizontal line with equally spread points, which is a good indication of homoscedasticity. With these plots we showed that our proposed model is suitable.

3. Conclusions

Our proposed model in Section 2 is quite different from the model discussed in [6]. We believe this is because the plants are located with diverse conditions of climate, hydrogaphy, mountains, glaciers, etc. However, both plants have very significant temperature, pH and coliform count predictors. In section 2, it is shown that the response variable chlorine (Q) is not symmetric, as well as the temperature and coliform count. By doing suitable transformation in these variables was possible to achieve the requirement to fit a linear model and the assumption are verified through the diagnostic plots. The proposed model (3) fits very well the data set from Bellavista plant. Table 1 shows that all the predictors are highly significant and the adjusted *R*-squared is 0.9994.

Symbols

Q — Chlorine, kg/24 h
T — Temperature, °C
C — Coliform count, NMP/100 ml
pH — Hydrogen potential

References

[1] Guías para la calidad del agua potable, Publicaciones de la Organización Mundial de la Salud, Tercera Edición 2004, p. 11.

[2] J.F. Griffith, S.B. Weisberg, C.D. McGee, Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples, J. Water Health., 4 (2003) 1.

[3] R.E. Martinez, L.B. Villalobos y L.K. Castillo, Escherichia coli Diarrogenicas procedentes de aguas marinas recreacionales caracterizadas por reacción en cadena de la polimerasa, Red de revistas científicas de America Latina y el Caribe, España y Portugal., 3 (2015) 248–254.

[4] C.H. Sánchez, L. González, R. Armendáriz, J.C. Mesa, N.B.C. El-Mousati, A.H. de la Torre, Trihalometanos en aguas de consumo humano, Revista de Toxicología, 2 (2011) 109–114.

[5] C. Freire, R. Soler, M.F. Fernández, C.M. Villanueva, J.O. Grimalt, N. y Olea, Valores de trihalometanos en agua de consumo de la provincia de Granada, España, Gaceta Sanitaria, 6 (2008) 520–526.

[6] A.E. Mohammed, B. Abdulkarim, Development of mathematical model for determining the quantity of chlorine required for water treatment, Department of Chemical Engineering, University of Abuja, J. Appl. Sci. Res., 8 (2010) 1002–1007.

[7] J. López, Sistema de Abastecimiento de Agua Potable en Nuevo Pachacútec, publicaciones centro de investigación social y educación popular, Alternativa 2003, p. 5.

[8] Instituto Nacional de Estadística e Informática: Censos Nacionales 2007, XI de población y VI de vivienda, Available at: http://censos.inei.gob.pe/cpv2007/tabulados/ (Acceso 15 diciembre 2017).

[9] APHA, AWWA, WPCF, Métodos normalizados para el análisis de aguas potables y residuales, Ediciones Díaz de Santos, Madrid, 1992.

Appendix

A1. Data set

Table A1
Experimental data for quantity of chlorine model. Bellavista treatment plant period January–February 2016

	Month of January				Month of February			
No.	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)
1	14.29	16,300	13.30	7.56	19.87	14,800	7.10	7.46
2	14.24	16,200	13.40	7.60	19.96	15,000	7.00	7.28
3	14.42	16,600	13.00	7.60	20.14	15,400	6.80	7.24
4	14.29	16,300	13.30	7.50	20.09	15,300	6.90	7.25
5	14.42	16,600	13.00	7.55	20.28	15,700	6.70	7.20
6	14.33	16,400	13.20	7.60	20.32	15,800	6.60	7.32
7	14.29	16,300	13.30	7.56	20.73	16,700	6.30	7.15
8	14.24	16,200	13.40	7.60	20.09	15,300	6.90	7.21
9	14.42	16,600	13.00	7.60	20.50	16,200	6.40	7.21
10	14.24	16,200	13.40	7.55	20.64	16,500	6.30	7.31
11	14.38	16,500	13.10	7.60	20.50	16,200	6.40	7.37
12	14.33	16,400	13.20	7.60	20.55	16,300	6.40	7.40
13	14.15	16,000	13.40	7.58	20.28	15,700	6.70	7.40
14	14.29	16,300	13.30	7.60	20.46	16,100	6.50	7.35
15	14.38	16,500	13.10	7.60	20.77	16,800	6.20	7.38
16	14.29	16,300	13.30	7.58	20.73	16,700	6.30	7.37
17	14.33	16,400	13.20	7.60	20.68	16,600	6.30	7.36
18	14.29	16,300	13.30	7.55	20.37	15,900	6.60	7.36
19	14.38	16,500	13.10	7.55	20.64	16,500	6.30	7.40
20	14.33	16,400	13.20	7.46	20.55	16,300	6.40	7.40
21	14.38	16,500	13.10	7.50	20.73	16,700	6.30	7.25
22	14.29	16,300	13.30	7.48	20.82	16,900	6.10	7.23
23	14.33	16,400	13.20	7.52	20.41	16,000	6.50	7.42
24	14.33	16,400	13.20	7.57	20.87	17,000	6.00	7.37
25	14.38	16,500	13.10	7.45	20.64	16,500	6.30	7.36
26	14.42	16,600	13.00	7.40	20.82	16,900	6.10	7.40
27	14.29	16,300	13.30	7.42	20.91	17,100	6.00	7.45
28	14.33	16,400	13.20	7.46	20.73	16,700	6.30	7.25
29	14.38	16,500	13.10	7.42	20.50	16,200	6.40	7.30
30	14.42	16,600	13.00	7.42	–	–	–	–
31	14.42	16,600	13.00	7.50	–	–	–	–

Table A2
Experimental data for quantity of chlorine model. Bellavista treatment plant period March–April 2016

	Month of March				Month of April			
No.	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)
1	19.64	21,100	7.20	7.30	16.78	16,800	10.50	7.45
2	19.60	21,000	7.30	7.40	16.60	16,400	10.70	7.45
3	19.60	21,000	7.30	7.40	16.92	17,100	10.40	7.40
4	19.50	20,800	7.50	7.25	16.69	16,600	10.60	7.45
5	19.41	20,600	7.60	7.35	16.37	15,900	10.90	7.38
6	18.91	19,500	8.20	7.50	16.69	16,600	10.60	7.42
7	19.50	20,800	7.50	7.42	16.37	15,900	10.90	7.38
8	18.73	19,100	8.40	7.40	15.69	15,400	12.00	7.30
9	19.28	20,300	7.80	7.38	16.56	16,300	10.80	7.40
10	19.41	20,600	7.60	7.39	16.37	15,900	10.90	7.33
11	19.28	20,300	7.80	7.40	16.65	16,500	10.70	7.40
12	19.32	20,400	7.70	7.35	16.51	16,200	10.90	7.45
13	19.10	19,900	7.90	7.35	16.69	16,600	10.60	7.40
14	19.05	19,800	8.00	7.35	16.78	16,800	10.50	7.45
15	19.28	20,300	7.80	7.40	15.65	15,300	12.00	7.30
16	19.41	20,600	7.60	7.40	16.65	16,500	10.70	7.40
17	19.46	20,700	7.60	7.40	16.51	16,200	10.90	7.40
18	19.23	20,200	7.90	7.35	16.65	16,500	10.70	7.45
19	19.41	20,600	7.60	7.40	15.69	15,400	12.00	7.40
20	19.32	20,400	7.70	7.40	16.37	15,900	10.90	7.30
21	19.28	20,300	7.80	7.42	15.69	15,400	12.00	7.34
22	19.14	20,000	7.90	7.38	16.19	15,500	11.50	7.30
23	19.19	20,100	7.90	7.39	16.69	16,600	10.60	7.30
24	19.05	19,800	8.00	7.36	16.60	16,400	10.70	7.40
25	18.87	19,400	8.30	7.40	16.69	16,600	10.60	7.40
26	18.87	19,400	8.30	7.35	16.33	15,800	11.00	7.40
27	19.32	20,400	7.70	7.40	16.51	16,200	10.90	7.35
28	19.28	20,300	7.80	7.42	16.56	16,300	10.80	7.40
29	19.14	20,000	7.90	7.44	16.51	16,200	10.90	7.45
30	18.96	19,600	8.10	7.38	16.51	16,200	10.90	7.35
31	18.64	18,900	8.50	7.36	–	–	–	–

Table A3
Experimental data for quantity of chlorine model. Bellavista treatment plant period May–June 2016

No.	Month of May				Month of June			
	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)	Chlorine (Kg)	Fecal coliforms (NMP/100 ml)	Temperature (°C)	Hydrogen potential (pH)
1	14.15	5,400	13.40	7.40	13.97	5,000	13.70	7.32
2	14.06	5,200	13.50	7.32	14.06	5,200	13.50	7.40
3	14.15	5,400	13.40	7.32	13.97	5,000	13.70	7.35
4	14.06	5,200	13.50	7.45	14.06	5,200	13.50	7.35
5	14.02	5,100	13.60	7.40	14.02	5,100	13.60	7.30
6	14.15	5,400	13.40	7.40	13.93	4,900	13.80	7.38
7	14.02	5,100	13.60	7.35	14.02	5,100	13.60	7.29
8	14.15	5,400	13.40	7.30	13.83	4,700	13.90	7.45
9	14.06	5,200	13.50	7.21	13.97	5,000	13.70	7.32
10	13.88	4,800	13.90	7.32	13.93	4,900	13.80	7.40
11	14.02	5,100	13.60	7.40	13.93	4,900	13.80	7.30
12	13.61	4,200	14.00	7.40	14.02	5,100	13.60	7.32
13	14.02	5,100	13.60	7.40	13.97	5,000	13.70	7.40
14	14.06	5,200	13.50	7.42	13.97	5,000	13.70	7.39
15	13.88	4,800	13.90	7.35	14.06	5,200	13.50	7.45
16	13.83	4,700	13.90	7.30	13.97	5,000	13.70	7.40
17	13.88	4,800	13.90	7.32	14.06	5,200	13.50	7.35
18	13.61	4,200	14.00	7.40	13.97	5,000	13.70	7.45
19	14.15	5,400	13.40	7.40	14.02	5,100	13.60	7.37
20	14.15	5,400	13.40	7.40	13.88	4,800	13.90	7.38
21	13.83	4,700	13.90	7.22	13.97	5,000	13.70	7.46
22	13.88	4,800	13.90	7.32	13.88	4,800	13.90	7.40
23	13.97	5,000	13.70	7.30	13.93	4,900	13.80	7.30
24	13.83	4,700	13.90	7.32	13.97	5,000	13.70	7.35
25	13.97	5,000	13.70	7.33	13.93	4,900	13.80	7.35
26	13.88	4,800	13.90	7.40	13.88	4,800	13.90	7.32
27	13.88	4,800	13.90	7.36	14.06	5,200	13.50	7.34
28	14.06	5,200	13.50	7.42	13.97	5,000	13.70	7.36
29	14.02	5,100	13.60	7.30	14.06	5,200	13.50	7.40
30	13.93	4,900	13.80	7.40	13.88	4,800	13.90	7.32
31	13.97	5,000	13.70	7.30	–	–	–	–