Modeling of Iron (Fe) Concentration in Groundwater with Microbiological Quality Indicators of Drinking Water

1 M. E. Ubiaru, 2 U. Musa, 3 C. L. Eze, 4 S. S. Hati

1, 2 National Agency for Food and Drug Administration and Control, NAFDAC, Port-Harcourt, Nigeria
3 Institute of Geosciences and Space Technology (IGST), Rivers State University of Science and Technology Nigeria
4 Analytics Division, Stat-DataPro+, Abuja, Nigeria

ABSTRACT

Extensive attention is being drawn to modelling water quality indicators from historical, experimental or stochastic data due to the connection between poor water quality and human health. The present study utilized a Multiple Linear Regression (MLR) equation as a model function for Monte Carlo simulation of Fe concentration in ground water, using field data of microbiological water quality indicators (Total heterotrophic bacteria, Total Coliform Bacteria and Escherichia coli) as predictor parameters. Field data were collected and analysed using standard methodology. DiscoverSim® version 1.1 simulation and statistical add-in software for Microsoft Excel was used to perform the Monte Carlo simulation. Result of the model supports existing finding that microbial water quality indicators are strongly associated with the presence of Fe in ground water, and indicated that that up to 35% of ground water sourced for drinking water purposes within Lagos environs are expected to contain more than 10mg/l of Fe. This condition is a cause for concern as the microbial quality of such water sources is expected to increase as well. Further studies for model validation are recommended.

Keywords: Monte Carlo Simulation, Iron, THB, Total Coliform, E. coli

1. INTRODUCTION

Modeling water quality indicators from historical, experimental or stochastic data has received substantial attention due to the link between poor water quality and human health and environmental hazards. Modelling has generally been applied to water quality monitoring in terms of certain heavy metals (Umar et al., 2013), oxygen demand parameters (Stefan and Fang, 1994; Gwaky et al., 2013), wastewater treatment (Heikkinen et al., 2011), water resources and distribution (Maier and Dandy, 2000), and for forecasting water quality parameters in water treatment processes (Baxter et al., 1999; Maier, 2004). This is due to the fact that modelling is a cost effective means of elucidating certain risks associated with water quality and sustainability practice.

Iron (Fe) is a vital trace element for the majority of organisms; a very resourceful prosthetic constituent present in numerous major enzymes of key biological processes (Andrews et al., 2003). Fe is also considered an essential nutrient for bacterial growth, similar to other elements, such as nitrogen, phosphorus and carbon. For instance, it is a component of all heme enzymes, which comprise cyttochromes and hydroperoxidase. It is known to inhibit microbial development in aquatic environments (Church et al., 2000; Kirchman et al., 2000), especially with regards to activities in which bacteria are involved in mechanisms, like the production of sturdy iron chelators (siderophores) in severe Fe shortage (Guan, 2001).

Fe is an electron acceptor, in the absence of oxygen, for bacteria that can join organic matter oxidation to Fe(III) reduction (Lovley and Phillips, 1986). These Fe-reducing bacteria are consequently able to put on energy from the reduction of soluble or solid Fe species (Jorand et al., 2001; Nguema et al., 2002).

The benefits of the presence of Fe in drinking water system for bacteria, specifically the Escherichia coli strain SH 702, have been well documented by Appenzeller et al. (2005). Both suspended bacteria and biofilm-associated bacteria, in the presence of Fe corrosion products have been reported to support bacterial action in drinking water systems, resulting in increases of both suspended bacteria and biofilm-associated bacteria (ller et al., 2001; Niquette et al., 2000).

Foremost microbiological quality indicators of drinking water include Total Coliform Test, which in theory designate the presence of all coliform group bacteria (vegetative and fecal in origin) and E. coli, a coliform species found in the intestinal tract of warm-blooded animals. Its presence in drinking water is suggestive of clear pollution from human or animal waste. Although E. coli is generally not harmful, some strains (0157:H7) may be toxic. Examples of Total Coliforms are: Escherichia, Enterobacter, Klebsiella, Citrobacter and Fecal Coliforms are: Escherichia, Klebsiella, Citrobacter, but 60% to 90% of total coliforms are fecal coliforms and 90%+ of fecal coliforms are Escherichia (usually E. coli) (LeChevallier, 1990; LeChevallier et al, 1996).

Also, heterotrophic bacteria (THB), which broadly include primary and secondary bacterial pathogens, as are Coliforms (Escherichia, Klebsiella, Enterobacter, Citrobacter, Serratia) that can also be found in various water sources, such as public water distribution systems, boreholes, wells and rivers, well water pressure tanks, hot water heaters and on-the-shelf bottled water. The concern of THB levels in drinking water initially focused on its possibility for interference with the analysis of Total Coliform in water. Both Total Coliform and subsequently E. coli analyses indicate the potential presence of fecal contamination of the water. Where these
organisms interfere in the attempt to find such contamination, then it was thought to be of public health concern to be monitored for this purpose (Staley, 1985; Allen et al., 2004).

In this study, groundwater refers to waters from boreholes, and wells.

In Nigeria as elsewhere, microbial groundwater pollution by effluents is worsening (UNEP/UNESCO, 1983). Ademoroti (1996) reported that the disposal of domestic and industrial effluents via soak-away may impair groundwater quality unless there is an impermeable stratum between the disposal area and groundwater table Ademoroti (1996). This consideration is important in locating refuse/dumpsite for both industrial and municipal wastes. However, even in groundwater sources not adjacent to dumpsite, microbial quality status has been high (Ademoroti, 1996). Up to (2570-34.7x10^5 cfu/ml) of Coliforms have been reported in leachate samples at Olushosun, Lagos Nigeria (GOPA, 1997).

The present study attempts to utilize a Multiple Linear Regression (MLR) equation as a model function for Monte Carlo simulation of Fe concentration in ground water, using field data of microbiological indicators (Total heterotrophic bacteria, Total Coliform Bacteria and Escherichia coli) water quality indicators as predictor parameters.

The result of the simulation model is expected to reveal the extent of risk associated with drinking untreated ground water within Lagos and environs, in terms of Fe concentration and these microbiological water quality indicators.

2. EXPERIMENTALS

2.1 Groundwater Sampling

Groundwater samples were collected from three boreholes each of the two sampling locations: L1 (06° 35' 40N 003° 22' 46E) adjacent Olushosun dumpsite and L2 (06° 35' 25N 003° 22' 49E) Charity Oshodi, about 500m distance from one another, all in Lagos State Nigeria. Sampling locations were identified with a Global Positioning System (GPS 315 Magellan, Taiwan).

Sampling was carried out in replicates of three, and every other month commencing February 2007 to February 2008. Sample were collected in sterile polyethylene containers and analysed in the National Agency for Food Drug Administrations and Control (NAFDAC) laboratory Oshodi, Lagos.

Methodologies described in APHA (1992) were used in samples collections, preparation and analyses.

2.2 Determination of Fe in Groundwater

Fe was determined in groundwater samples by the standard methods of atomic absorption spectrometry (AAS) with spectrophotometer (PU 9100X, England). Prior to aspiration in the spectrophotometer, water samples were digested with 2M nitric acid, heated until clear solution was obtained. The samples readings were recorded against a standard curve prepared from standard solutions of Fe in mg/l.

2.3 Microbial Assay

2.3.1 Media Preparation

Astell Scientific autoclave (AAC046, England) was used for sterilization at 121°C for 15 minutes at 15psi gauze pressure. Broth culture media consisted of 9ml Mac-Conkey broth in McCartney bottles with inverted Durham tubes. All the glassware including McCartney bottle and conical flasks were sterilized in hot air oven (Binder FP400, Germany) at 160°C for 2h. Prepared media were maintained at 45°C in a water bath (Kottermann-MD-3165, Germany) before use.

2.3.2 Enumeration of Total Heterotrophic Count (THB) was carried by the pour-plate inoculation method described in APHA (1992). Sterile molten plate count agar medium was used with well labeled plates; each 1ml from 10^3-sample dilution was dispensed. Approximately 15ml of freshly prepared plate count agar (PCA) medium was poured onto each inoculated plate, rocked gently and allowed to solidify in the upright position. Inoculated plates were incubated in an incubator (Gallenkamp, 1H-150) in the reverse position at 37°C for 48-h. After incubation, the mean count of bacterial colonies in duplicate plates was recorded and the result expressed as:

\[
\text{THB} \left( \frac{\alpha \times 10^3 \text{ cfu} \text{ ml}^{-1}}{\text{Vol. of sample plated}} \right) = \frac{\text{No. of colonies} \times \text{df}}{\text{Vol. of sample plated}} \]  (1)

\(\alpha = \text{number of colonies counted;}
\)
\(\text{df = dilution factor}\)

2.3.3 Enumeration of Total Coliform Bacteria (TCB)

The Standard Most Probable Number (MPN) method using 3-tube dilution as outlined in APHA (1992) was employed for the enumeration of TCB. Mac-Conkey broth medium in McCartney bottles were set out in rows of 3s. Each tube was inoculated with 1ml of 10^3-sample dilution, incubated at 37°C for 48h. Positive results were indicated by the production of ‘acid and gas’, i.e. presence of gas in the inverted Durham tubes and change in colour of the medium from pink to orange. Results were computed from an MPN index table APHA (1992).

2.3.4 Enumeration of Escherichia coli (E. coli)

The method consisted of the use of a double-layered violet red bile agar (VRBA) medium. Inoculated plates were incubated at 44.5°C for 48-h. Presumptive colonies of E. coli (pinkish, spindle-shaped colonies) were purified and then confirmed after Gram’s staining, biochemical tests and results estimated from the reference table APHA (1992).
2.4 Simulation Modelling
DiscoverSim® version 1.1 (SigmaXL, 2013) simulation and statistical add-in software for Microsoft Excel was used to perform the Monte Carlo simulation of Fe concentration in groundwater in relation to microbial quality indicators.

The model function for Fe concentration was generated from a multiple linear regression (MLR) model expressing the relationship between the three input variables (THB, Coliform; E.coli) and the output variable (Fe), by fitting a linear equation to observed data samples. The MLR model equation is:

$$Fe = \beta_0 + \beta_1THB + \beta_2Coliform + \beta_3E.coli + \epsilon$$  

(2)

where $\beta_0$ is a constant; the intercept value, $\beta_{1,3}$ are the unidentified partial regression coefficients of each of the input variables, and $\epsilon$ consists of the uninhibited aspects and errors of the regression model. To minimize the sum of the squares of the vertical deviations from each data point, the least squares fitting was adopted, to express the best fits line for the observed data (Petri et al., 2012; Manly, 2009).

The simulation model was generated from 10000 iterations using the model function. The sequence in the Monte Carlo simulation using DiscoverSim® briefly consists of the following:

- Tests for normality and correlation analysis
- Selection of distribution fitting
- Input and output variables selection
- Run simulation and display results

3. RESULTS AND DISCUSSION

3.1 Field Data and Normality Test
Field data used for the simulation model is shown on Table 1. The average concentration of Fe (0.7±0.4 mg/l) in ground water samples from these locations is within the range reported in other locations. In anaerobic groundwater, Fe has been found in the range from 0.5 to 10 mg/L, also with reports of concentrations up to 50 mg/L (WHO, 2003). However due to the influence of dumpsite adjacent location of ground water (L1), there was a significant (t stat =9.9, p <.000) difference between the two sampling locations.

Excepting E.coli (t-stat=3.9, p=.002), there were no significant differences between THB and TCB counts in the two locations respectively. This may be attributed to the fact that these later parameters consist of enumerations of broad based microbial organisms (LeChevallier et al, 1996; Allen et al., 2004), in which changes in the individual constituents was not taken into account. Excepting Fe, all parameters passed normality tests as there were no significant P-values from Anderson-Darling test. Fe correlated strongly and significantly with TCB (0.7) and E.coli (0.7), as with TCB and E.coli (0.6). This is expected as E.coli is a subset of TCB (LeChevallier et al, 1996).

Table 1: Field data used for modelling with normality tests results

<table>
<thead>
<tr>
<th>Sampling Location</th>
<th>Fe (mg/l)</th>
<th>THB</th>
<th>TCB</th>
<th>E.coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>0.80</td>
<td>20</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>L1</td>
<td>1.04</td>
<td>21</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>L1</td>
<td>1.09</td>
<td>26</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td>L1</td>
<td>1.27</td>
<td>32</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>L1</td>
<td>1.09</td>
<td>29</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>L1</td>
<td>1.21</td>
<td>21</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>L1</td>
<td>1.11</td>
<td>19</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>L2</td>
<td>0.09</td>
<td>9</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>L2</td>
<td>0.17</td>
<td>18</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>L2</td>
<td>0.22</td>
<td>28</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>L2</td>
<td>0.43</td>
<td>33</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>L2</td>
<td>0.49</td>
<td>31</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>L2</td>
<td>0.37</td>
<td>22</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>L2</td>
<td>0.35</td>
<td>20</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Min</td>
<td>0.09</td>
<td>9</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Max</td>
<td>1.27</td>
<td>33</td>
<td>40</td>
<td>12</td>
</tr>
<tr>
<td>Mean</td>
<td>0.7</td>
<td>23.5</td>
<td>23.64</td>
<td>6.5</td>
</tr>
<tr>
<td>SD</td>
<td>0.43</td>
<td>6.65</td>
<td>8.63</td>
<td>2.68</td>
</tr>
<tr>
<td>Anderson-Darling, P-Values</td>
<td>0.04</td>
<td>0.30</td>
<td>0.96</td>
<td>0.27</td>
</tr>
<tr>
<td>Spearman Rank Correlations</td>
<td>Fe</td>
<td>1</td>
<td>0.36</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>THB</td>
<td>1</td>
<td>0.78</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>TCB</td>
<td>1</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.coli</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Distribution Fitting for Input Parameters
Due to the small size of field data, distribution fittings for the input variables (THB, TCB; E.coli) were automatically determined by DiscoverSim®. This revealed that Box-Cox distribution was most suitable for TCB and TCB, while Skew-Normal distribution was most suitable for E.coli. Details are as presented on Table 2.

Table 2: Distribution fitting for input variables

<table>
<thead>
<tr>
<th>Distribution</th>
<th>THB</th>
<th>TCB</th>
<th>E.coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution fit</td>
<td>Box-Cox</td>
<td>Box-Cox</td>
<td>Skew-Normal</td>
</tr>
<tr>
<td>Location value</td>
<td>28.562</td>
<td>16.045</td>
<td>3.099</td>
</tr>
<tr>
<td>Scale SD</td>
<td>9.067</td>
<td>5.435</td>
<td>4.272</td>
</tr>
<tr>
<td>Lambda/Skew value</td>
<td>1.101</td>
<td>0.852</td>
<td>10.489</td>
</tr>
</tbody>
</table>
3.3 MLR Analysis and Model Function
MLR equation 3 expresses the result of the relationship between the three input variables (THB, TCB; E.coli) and the output variable (Fe). Detail of the MLR analysis is presented on Table 3.

\[
Fe = (0.096) + ( 0.033) \times THB + (0.0471) \times TCB + (0.039) \times E.coli
\]  

Generally, there is a strong degree of association between Fe and microbial quality indicators as the regression coefficient, \( R^2 \) Adjusted = 62.9%, indicated that about 60% of Fe variations in the model can be explained by the predictor parameters (THB, TCB; E.coli). Though the result also shows that the constant (intercept, 0.10) is not significant (\( P =.77 \)) in the model, but VIF (variance inflation factor) and Tolerance, measures of co-linearity, <5 and <0.2 respectively, showed diverse co-linearity between the parameters; THB (3.61, 0.28), TCB (6.27, 0.16) and E.coli (2.72, 0.37). Analysis of variance for the model revealed a significant \( p < .00 \), indicating that the model is statistically significant. Durbin-Watson Tests shows significant p-values for auto-correlations suggesting that there was a trend in the residuals; a time series correlations between input variables as expected. On the whole, these provide authentication of the model for use in explaining the relationships between Fe and the predictor parameters (THB, TCB; E.coli), thus, supporting the fact that Fe is essential to these microbial quality indicators (Church et al., 2000, Kirchman et al., 2000, Andrews et al., 2003). From the MLR equation 3, the minimum concentration of Fe in ground water from this region would be 0.51mg/l per unit count of these microbial water quality indicators. This value is about 27.1% lower than the sample average of 0.70 mg/l.

For the fact that Fe is the second most abundant metal in the earth crust, the frequency distribution of simulation suggests the presence of Fe in every sampled ground water.

3.4 Monte Carlo Simulation
The model function (equation 3), used to perform the simulation produced distribution frequency curve shown in Figure 1 and detail statistical output on Table 4. The distribution curve indicated that within the study area, up to 65% of ground water has Fe concentration in the typical range of 0.5 to 10 mg/L (WHO, 2003). Hence about 34.5% of ground water within this region is expected to be out of this specification limits.

<table>
<thead>
<tr>
<th>Table 3: Result of regression analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regression coefficients</strong></td>
</tr>
<tr>
<td>R-Square</td>
</tr>
<tr>
<td>R-Square Adjusted</td>
</tr>
<tr>
<td>S (Root Mean Square Error)</td>
</tr>
<tr>
<td><strong>Parameter Estimates:</strong></td>
</tr>
<tr>
<td>Predictor Term</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>THB</td>
</tr>
<tr>
<td>TCB</td>
</tr>
<tr>
<td>E.coli</td>
</tr>
<tr>
<td><strong>Analysis of Variance for Model:</strong></td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>Total (Model + Error)</td>
</tr>
<tr>
<td><strong>Durbin-Watson Test for Autocorrelation in Residuals:</strong></td>
</tr>
<tr>
<td>DW Statistic</td>
</tr>
<tr>
<td>P-Value Positive Autocorrelation</td>
</tr>
<tr>
<td>P-Value Negative Autocorrelation</td>
</tr>
</tbody>
</table>
Fig 1: Frequency distribution for Fe concentration (mg/l)

Table 4: Result of Monte Carlo simulation

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>10000</td>
</tr>
<tr>
<td>Mean</td>
<td>9.129</td>
</tr>
<tr>
<td>SD</td>
<td>2.213</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.302</td>
</tr>
<tr>
<td>25th Percentile (Q1)</td>
<td>7.661</td>
</tr>
<tr>
<td>50th Percentile (Median)</td>
<td>9.164</td>
</tr>
<tr>
<td>75th Percentile (Q3)</td>
<td>10.575</td>
</tr>
<tr>
<td>Maximum</td>
<td>17.663</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Normality Tests</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson-Darling Normality Test</td>
<td>1.087</td>
</tr>
<tr>
<td>p-value (A-D Test)</td>
<td>0.008</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.028</td>
</tr>
<tr>
<td>p-value (Skewness)</td>
<td>0.248</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.046</td>
</tr>
<tr>
<td>p-value (Kurtosis)</td>
<td>0.347</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual Specification Limits (Empirical)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Total (out of specification)</td>
<td>34.500</td>
</tr>
<tr>
<td>% Total (within specification)</td>
<td>65.500</td>
</tr>
</tbody>
</table>

However, the model revealed a general average of 9.13±2.21 mg/l concentration of Fe in ground water samples based on the field data for this study area.

The model also revealed that Fe concentration distribution is not normal (Anderson-Darling, 1.087, p =0.008), due to the diverse influence of anthropogenic activities.

The model is suggestive of the fact that so long as these microbial water quality indicators are found in ground water samples in this study area, the minimum average of Fe concentration is expected to be in the neighborhood of 1.3mg/l. This supports the fact that about 5% of the earth's crust contains Fe, making it the second most abundant metal (Elinder, 1986), consequently the cause of its presence in most environmental matrices.

4. CONCLUSION

This study achieved the utilization of a MLR equation as a model function for the development of a simulation model for Fe concentration in ground water, using field data of microbiological water quality indicators as predictor parameters.

The model supports existing finding that microbial water quality indicators are strongly associated with the presence of Fe in ground water.

Hence this model revealed that up to 35% of ground water sourced for drinking water purposes is expected to contain more than 10mg/l of Fe. This condition is a cause for concern as the microbial quality of such water sources is expected to increase as well.

However, further studies are required to validate the present findings with this model.

REFERENCES


[27] SigmaXL (2013). Discovery through simulation, DiscoverSim® version 1.1 workbork, SigmaXL Inc. 803, Toronto, Ontario, M4W-3C7, Canada. www.sigmaxl.com


