EUDIOMETRIC THEORETIC-APPROACH TO MODELLING THE ASSIMILATIVE CAPACITY OF A RIVER: INCORPORATION OF BOOTSTRAPPING NEEDFUL FOR SENSITIVITY ANALYSIS

C. M. Chiejine¹*, A. C. Igboanugo² and L. I. N. Ezemonye³

¹DEPT OF ELECTRICAL/ELECTRONIC ENGINEERING, DELTA STATE POLYTECHNIC, OGWASHI-UKU, DELTA STATE, NIGERIA
²DEPARTMENT OF PRODUCTION ENGINEERING, UNIVERSITY OF BENIN, BENIN CITY, EDO STATE NIGERIA
³DEPARTMENT OF ANIMAL AND ENVIRONMENTAL BIOLOGY, UNIVERSITY OF BENIN, BENIN CITY, EDO STATE. NIGERIA

E-mail addresses: ¹pstchiejine@yahoo.com, ²dracigboanugo@yahoo.com, ³ezemlaw@yahoo.com

ABSTRACT

The mathematical physics underlying the adsorption and subsequent desorption of dissolved oxygen (DO) in a water body subject to effluent loading had been rarely investigated. The current state of play in this field although reflects use of different methods, the combine use of hat matrix and bootstrapping techniques to study the phenomenon of chemical adsorption and desorption of DO at molecular level in a polluted waterbody has not been thoroughly investigated. This study seeks to use a matrix projector, H-hat (Ĥ), to cast virtual spectral rays on pollutant loadings in a water body and in the process unravel the dynamics of chemical and biological gravitation of dissolved oxygen towards constituents of effluent pollutants in water body. This approach is anchored on the ordinary least squares methodology of multivariate linear regression. The method hypothesised is studded by a mathematical physics analysis of the phenomenon. Bootstrapping was used to establish means and variances of regression parameters, and subsequently, the confidence intervals of point estimates of parameters. Tricking technique adopted facilitated the development of extreme values of the dissolved oxygen and hence the supremum and infimum of assimilative capacity of the river which fluctuates with intensity of effluent loadings and season of the year (rainy, dry, and harmattan seasons). The result of bootstrapping revealed that assimilative capacity fluctuated widely from the values detected by point estimates of regression parameters thus suggesting that tricking of regression parameters, in turn, tunes up the regression model, and hence, fine tunes the value of assimilative capacity through necessary adjustments of model parameters. The results of this study obviates the need to deploy eudiometer for laborious direct measurement of dissolved oxygen in a body of polluted water. Thus an elegant technique for crossing the stream where it is shallowest has been developed in this study. The method is considered as a great improvement on previous approaches that seem to dawdle.

Keywords: eudiometry, bowl, assimilative capacity, bootstrapping, model tricking.

1. INTRODUCTION

The dynamics of interaction between pollutants from effluents and dissolved oxygen (DO) in a body of water has not been well studied. But it is known, however not in detail, that each pollutant in effluent discharge somehow react when mixed in water and in the process adsorb and subsequently desorb oxygen by virtue of relative valance between it and molecules of the DO in the body of water. And, as earlier stated, the mathematics underpinning the adsorption/desorption of oxygen molecule by each pollutant from the body of water is not well understood [1]. This study attempts to use projector matrix Ĥ, in the sense of ordinary least squares (OLS) of the multiple linear regression model, to unravel the phenomenon of dissolved oxygen adsorption and desorption in a body of water subject to effluent loading. The Ĥ matrix is conceived as virtual spectral rays incident upon a medium of water in which pollutant are injected as a plume. The virtual spectral signal (VSS) is assumed to energize the molecules of the pollutants such that, based on their relative valance, oxygen molecules will

*Corresponding author, Tel: +234-802-684-4167
gravitate towards each pollutant according to their respective potentials. Again, gobs and gobs of different pollutants in the effluent plume is treated as a column vector of independent variables \( x \), while the balance of DO in the medium which determines the assimilative capacity (AC), is treated as response or explanatory variable, \( y \). The problem in this modelling methodology is to apply virtual eudiometry as black box to estimate the amount of DO at any time based on sample realization from the polluted water body. Thus, the chief parameter of interest from the sample realization is the \( \beta \) which can be expressed as a linear combination of the column vector: \([x_1, x_2, x_3, ..., x_n]\) to give the estimate or ballpark of the population values of the parameter that estimates the DO (\( y \)). It is easy to visualize that estimate of the parameters are point values and we need a special technique of bootstrapping to determine mean and variance of each of the values of the point parameters so that the confidence intervals which facilitate tricking of the mother parameters can be determined.

Whereas oodles of studies have addressed the pollution in Ikpoba River using proximate analysis and related approaches, the balance of literature is still grossly deficient in the determination of assimilative capacity and mathematical physics analysis of pollutant dispersion in both far-field and near-field. See for example [1-4].

Again, few authors have applied linear regression in investigating effluent dispersion (see, for example also [5-8]). Apart from [9] which applied bootstrapping technique in a study to analyze the 'Sobol' sensitivity of a complex environmental model, very little have been researched using bootstrapping techniques to undertake sensitivity analysis on fluctuation of the DO in a body of polluted water. The current study therefore is a kind of effort to breach the perceived frontier of knowledge. The knowledge of the dynamics of relative adsorption of DO by the constituent pollutants, as this study affords, is important in the sense that it provides insight and enlightenment into better understanding of how the constituent pollutants contribute most to water body pollution and hence serve as a useful guide to modelling and corrective actions. Again, it will also help to test the robustness of the optimal solution.

The solution to the nagging problem of surface water pollution resulting from industrial effluents and municipal surface run-off require reliable and scientifically proven information. Such information can only be made available through regular monitoring programmes. Regular monitoring usually result in large volume of complex data matrix which requires accurate interpretation and significant conclusions.

Relatedly, [10] stated that application of multivariate statistical techniques help in the interpretation of complex data matrix and better understanding of water quality and its management. Hence, the study applied Cluster Analysis (CA), Discriminant Analysis (DA) and Principle component Analysis (PCA) to evaluate the variations in the water quality of Mumbai coast; and was able to give meaningful and interesting interpretations to the information obtained.


Moreover, the stalking of the level of dissolved oxygen in a body of water based on the observed variability in the level of pollutant in effluent loading appears to fully describe the mechanics of fluctuation of assimilative capacity of a river subject to effluent loading. Some of the foregoing statistical techniques categorically corroborate our approach of deployment of regression model and the accompanying resampling techniques in the nature of bootstrapping to show how adjustment of regression parameters can help to explain why assimilative capacity of a body of water fluctuates. The procedure therefore tries to bring out the level of significance of bootstrapping as applied therein. The procedure also brings out the important role which \( R^2 \), i.e. the multivariate coefficient of determination, plays in explaining to what extent the multivariate linear regression is able to account for the perceived variability or fluctuation of assimilative capacity given some changes in the level of pollutants in effluent loading. It is evident from the foregoing analysis of the previous works that the balance of literature is deficient on the use of regression analysis and bootstrapping to undertake sensitivity analysis to
do tricking in the prediction of dissolved oxygen from sample realizations.

The aim of this study is to use multivariate linear regression to show that geometric projection is a lucid means of identifying unit changes in pollutant variables by which we can further use bootstrapping technique to establish confidence intervals that facilitate tricking of the model parameters with a view to bringing out the necessary adjustments required for the optimization of DO prediction.

The current study offers the use of multivariate linear regression and bootstrapping statistical analysis to investigate the degradation of assimilative capacity of water body subject to effluent loading. The theoretical formulation underpinning the statistical application is provided and it makes the application to be better understood by readers.

2. METHODOLOGY

2.1 Data Collection

The accompanying data depicted in Table 1 were readings obtained from Ikpoba River from three point sources. At each point upstream and downstream, samples of polluted water were obtained 20m upstream and downstream respectively. Again, samples were collected 200m downstream of the third point source. These were taken over three seasons—namely wet, dry, and harmatan season. These samples were analysed in a chemical laboratory according to Table 1.

The observations of the sample pollutants were organized according to the format of Table 2 which is compatible with hat matrix application. The data were fitted into the accompanying Multivariate Linear Regression model.

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_7 y_7
\]

Where \([x_1, x_2, \ldots, x_n]\) are the observations shown in Tables 1 and 2. The problem was solved by both manual and MATLAB software.

2.2 Data Analysis

The Multivariate linear regression model was applied using sample realizations in order to develop a model that can predict the DO, which is the prime determinant of the assimilative capacity of the water body.

The various pollution parameters were treated as predictor variable \(x_i\) while the resultant DO was treated as the response variable, \(y\). Also, \(\beta_j\) \(j = 1, 2, \ldots, 7\), represents the propensity of each sensitivity index to, in the first instance, adsorb oxygen and subsequently to desorb same in order to meet both chemical and biological oxygen demand and thereby causing diminution of DO in the body of water. With this estimate, the level of DO in the river can be predicted howbeit as point estimates.

On the other hand, by bootstrapping technique of [20], there were 100 times resampling with replacement from the bowl With this, for each resampling, the tendency of the pollutants to deplete the DO (\(\beta\)) was calculated; and the overall values of \(\beta_i\) determined by means of scree plots see Figure 3 (a-h). The upper and lower control limits (the confidence intervals) of the resultant output of \(\gamma\) were determined by tricking, involving the substitution of the parameters in order to get the extreme values of \(\gamma\). By substituting the extreme values of each of the model parameters, the sensitivity analysis on the value of \(\gamma\) was seen to be evaluated. This defines the “bandwidth” of the pollutants depletive effects on DO in the polluted water body.

Previous to this stage, the multivariate coefficient of determination \(R^2\) was evaluated in order to ascertain the model adequacy. In other words, \(R^2\) which is given by, \(R^2 = \frac{1 - \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}}\), ascertains to what extent the model is able to account for the perceived variabilities in the value of response variable \(y\), which is the dissolved oxygen (DO).

<table>
<thead>
<tr>
<th>Table 1: Data Matrix of Ikpoba River Quality Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO_4</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>P1 (ups)</td>
</tr>
<tr>
<td>P1 (dns)</td>
</tr>
<tr>
<td>P2 (ups)</td>
</tr>
<tr>
<td>P2 (dns)</td>
</tr>
<tr>
<td>P3 (ups)</td>
</tr>
<tr>
<td>P3 (dns)</td>
</tr>
<tr>
<td>200M dns</td>
</tr>
</tbody>
</table>
3. THEORETICAL FORMULATION

Linear regression, in a geometrical sense, represents projection of vector of response variable, \( y \), onto the space defined by \( n \times m \) vector of predictor variable \( x \). Geometrically, we can do the following representation.

\[
\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon
\]

\( \beta \) is a ballpark of \( \beta \). In point of fact, \( \beta \) is the potential or valence or adsorptive capacity which each pollutant has for oxygen adsorption and when the resultant oxygen adsorption and desorption from DO is done, the balance of DO in the water body is \( y \), and the estimate (predicted value) is \( \hat{y} \).

\[
\hat{y} = b_0 + bx + \varepsilon
\]

In the natural process, we can represents these as column vectors.

\[
y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}
\]

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\]

Notice that \( x \) is a matrix that is considered tall and skinny; \( n \gg m \). The sole purpose of this research is to build a machine that can replicate the perceived natural process (model building). The realization is a sample obtained from the polluted river.

In the natural setting, the model is

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \ldots + \beta_m x_m + \varepsilon_i \]

\[
\varepsilon \approx N(O_2 \sigma^2)
\]

**Table 2: Realizations, \( X = f(S_1, y_1, \ldots, y_n) \)**

<table>
<thead>
<tr>
<th>S/N</th>
<th>Observations of Sample Pollutants</th>
<th>Resultant Depleted DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S_11 S_12 S_13 S_14 S_15 S_16 S_17</td>
<td>y_1</td>
</tr>
<tr>
<td>2</td>
<td>S_21 S_22 S_23 S_24 S_25 S_26 S_27</td>
<td>y_2</td>
</tr>
<tr>
<td>3</td>
<td>S_31 S_32 S_33 S_34 S_35 S_36 S_37</td>
<td>y_3</td>
</tr>
<tr>
<td>n</td>
<td>S_n1 S_n2 S_n3 S_n4 S_n5 S_n6 S_n7</td>
<td>y_n</td>
</tr>
</tbody>
</table>

In the natural setting, the model is

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \ldots + \beta_m x_m + \varepsilon_i \]

\[
\varepsilon \approx N(O_2 \sigma^2)
\]
This model is assumed, to a good extent, to mimic the natural process of oxygen diminution in the body of water as a result of adsorption and desorption of DO.

And with sample realizations we want to usually have \( \frac{n}{m} > 100 \) and in any case where \( \frac{n}{m} \leq 5 \), we are bothered that we do not have sufficient observations for estimating the number of parameters in \( \beta \).

Thus the multiple linear regression equation, in short hand, compact or matrix form, is \( y = x\beta + \epsilon \) And in expanded linear algebraic form we have:

\[
\text{cov}(\epsilon_i, \epsilon_j) = \sigma^2 \delta_{ij}; \sigma_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}
\]

Recall that \( b \approx \beta \), implying that \( b \) is a ballpark of \( \beta \). Also \( \beta \) is taken as an estimate.

We shall estimate \( b \) using sample realization such that \( \hat{y} \approx y_i \). Of course, the model fitting is seen from the point of view of ordinary least squares (OLS).

Let's choose a criterion, \( Q \)

\[
Q = \sum_{i=1}^{n} (y_i - \hat{y})^2
\]

where \( Q \) is deemed to be the sum of squared deviations or error sum of squares. This quantity, \( Q \), can be seen as a distance or length which reflects the degree or length of departure of estimate \( \hat{y} \) from the original observation.

\[
\hat{y} = b_0 + b_1 x_1 + \cdots + b_m x_m + \epsilon_i
\]

\[
y - \hat{y} = U
\]

\( U \) needs to be small so that \( \hat{y} \) will be closest to \( y_i \) as possible.

Observe that \( \hat{y} \) is the projection of \( y \) on column vector space spanned by \( \{x_1, x_2, \ldots, x_m\} \)

We wish to estimate the values of \( b \) that will minimize the value of \( U \)

\[
\sum U^2 = Q
\]

Hence,

\[
Q = \sum_{i=1}^{n} (y_i - \hat{y})^2
\]

\[
= \sum_{i=1}^{n} (y_i - x \beta)^2
\]

Note that equation (7) is in matrix form. Thus \( y = x \beta \) is a matrix difference. Consider:

\[
\sum (y - x b) (y - x b) = Q
\]

And by matrix operation, this is not conformable. Thus we can use the transpose of one of them to multiply the other since \( Q \) is a squared length of vector difference.

Hence we have,

\[
Q = (y - x b)^T (y - x b)
\]

Equation (11) is now conformable \((1 \times n) \times (n \times 1) = 1 \times 1 = \text{a number}\)

In calculus, we know that by equating the differential coefficient to zero, we can obtain the value of \( b \) that minimizes \( Q \). Thus we obtain the following sequence:

\[
\frac{\partial Q}{\partial b_0}, \frac{\partial Q}{\partial b_1}, \frac{\partial Q}{\partial b_2}, \ldots, \frac{\partial Q}{\partial b_m}
\]

When each of these expressions are equated to zero, we can after some cumbersome work, obtain the needed values of \( b \) that minimize \( Q \). To circumvent this imputed messy calculations, we can work backwards as follows;

Define the hat matrix \( \hat{H} \)

\[
\hat{H} = X(X^TX)^{-1}X^T
\]

\( X^T \) is the transpose of \( X \); it is the set of observations of pollutants represented as a vector matrix.

Some properties of \( \hat{H} \):

1. The inverse \((X^TX)^{-1}\) exists; in other words, \( X \) has full rank
2. \( \hat{H} \) is idempotent; hence \( \hat{H}^2 = \hat{H} \)
3. Diagonal element of \( \hat{H} = h_{ii} \); it has \( i \)th leverage value.

Set \( y = \hat{H} y + (1-\hat{H})y \)

We can now develop \( Q \) based on the following lemmas:

Lemma 1: \( \hat{H}^T = \hat{H} \)

Lemma 2: \((xb)^T = b^T x^T \)

Hence; we state as follows:

\[
Q = (y - x b)^T (y - x b)
\]
Notice that the second and third summands vanish as a result of the following manipulation/substitution:

\[\hat{y} - X(X^TX)^{-1}X^Ty\]

\[(1 - \hat{H})X = X - \hat{H}X = X - X(X^TX)^{-1}X^TX\]

\[= X - X(X^TX)^{-1}(X^TX)^{-1}X^TX = X - X = 0\]  \(14\)

Hence we have:

\[Q = (\hat{y} - x\beta)^T(\hat{y} - x\beta) + [(1 - \hat{H})y]^T[(1 - \hat{H})y]\]  \(15\)

If \(Q\) should be minimized with respect to \(\beta\), it should be achieved through the first summand (first term of equation 15).

It is noticed that if we set

\[b = (X^TX)^{-1}X^Ty\]  \(16\)

Then,

\[(H\hat{y} - x\beta)^T(H\hat{y} - x\beta)\]

\[= [H\hat{y} - X(X^TX)^{-1}X^Ty][H\hat{y} - X(X^TX)^{-1}X^Ty]\]

\[= [X(X^TX)^{-1}X^Ty - X(X^TX)^{-1}X^Ty][X(X^TX)^{-1}X^Ty - X(X^TX)^{-1}X^Ty]\]  \(17\)

Hence \(b = (X^TX)^{-1}X^Ty\) is the least squared estimate of \(\beta\). One cannot agree more that this result is most significant in the sense that it is a direct computation of regression parameters, \(\beta\), using the format of matrix of realizations \(x = f(S_i|y_i)\) depicted in Table 2. It is easily computed with MATLAB software.

### 3.1 Interpretation of the Model Estimation

\(\beta\) is the estimation of the rate, potential or valence with which the pollutants adsorb oxygen from the body of water. This rate has been estimated from the sample realization to be \(\hat{y}\). With this estimate of model parameters, we can predict the level of dissolved oxygen as

\[\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n + \epsilon_i\]  \(18\)

Finally, \(y\) is the set of observations of the balance of dissolved oxygen level remaining after some quantities had been adsorbed and desorbed by pollutants.

At this juncture, we note that \(b\) and \(\beta\) are mere point estimates of parameters. We need to use bootstrapping methodology to estimate the confidence interval, the purpose of which is needful for model tricking and adjustments in order to optimize \(y\), the prediction of dissolved oxygen, DO.

### 4. RESULTS AND DISCUSSION

The results of the study are presented in the following order:

(i) Regression Analysis
(ii) Bootstrapping
(iii) Sensitivity Analysis
(iv) Check for model Adequacy

We take them seriatim.

#### 4.1 Regression Analysis

In line with theory of multiple linear regression (MLR) using hat matrix \(H\), as matrix projector which elucidated the ordinary least squares (OLS) methodology applied, the regression parameters were computed using both manual and MATLAB software application. The results obtained with MATLAB are as shown in Table 3.

The two results are essentially the same mutatis mutandis, the respective differences being attributed to rounding off errors. Thus confirming that the theory straddling the two approaches are in concordance. Further, conflating both sets of results, it is evident that the manually computed results support the results obtained through MATLAB, confirming the authenticity of the values of the regression parameters. Again, each of these parameters represents the rate at which each variable increases with unit change in its value. The \(\beta_o\) in particular represents an autonomous value that may or may not have direct practical meaning depending on the situation under consideration. It is a value that provides necessary adjustments to the value of DO and that is why it has been referred to as an autonomous value. It could be seen also as a boundary condition.

#### 4.2 Bootstrapping

Further, we reiterate that the results of the regression parameters are mere point estimates. Thus, we need interval estimates of these parameters. Statistically, this appears difficult because we need a special methodology for generating the means and standard deviations in order to obtain the required confidence interval. In consequence, we had a recourse to the bootstrapping methodology developed by [20]. This method had also been applied in [10] in a study to...
analyze the ‘Sobol’ sensitivity of a complex environmental model. In this study, however, the sensitivity indices were used to rank the contribution and influence of each pollutant to the overall degradation of the water body. Never the less, this application corroborates the approach adopted by the current study. The results of bootstrapping are shown in Figure 4.

<table>
<thead>
<tr>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
<th>( \beta_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-279.537</td>
<td>80.2411</td>
<td>-131.07</td>
<td>4.0714</td>
<td>3.7552</td>
<td>27.0139</td>
<td>-3.1380</td>
<td>-1.3322</td>
</tr>
</tbody>
</table>

Table 3: MATLAB computed results of Regression analysis

Table 3: Depiction of Confidence Limits across Some Significance Levels

<table>
<thead>
<tr>
<th></th>
<th>UCL</th>
<th>UCL</th>
<th>UCL</th>
<th>UCL</th>
<th>LCL</th>
<th>LCL</th>
<th>LCL</th>
<th>LCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>193.6897</td>
<td>193.0304</td>
<td>192.3905</td>
<td>191.8902</td>
<td>195.5626</td>
<td>184.3446</td>
<td>192.967</td>
<td></td>
</tr>
<tr>
<td></td>
<td>229.2187</td>
<td>228.5594</td>
<td>227.9195</td>
<td>227.4192</td>
<td>231.0916</td>
<td>219.8736</td>
<td>228.4959</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>UCL</th>
<th>UCL</th>
<th>UCL</th>
<th>UCL</th>
<th>LCL</th>
<th>LCL</th>
<th>LCL</th>
<th>LCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>299.1487</td>
<td>298.4894</td>
<td>297.8495</td>
<td>297.3493</td>
<td>301.0216</td>
<td>289.8036</td>
<td>298.426</td>
<td></td>
</tr>
</tbody>
</table>

**Gloss:** UCL: Upper Control Limit, LCL: Lower Control Limit, PNT. EST.: Point Estimate

The results of the scree plots are as follows:

(a) PO\(_4\)

(b) NO\(_3\)

(c) Temp

(d) Feecal Coliform
The behaviours of $\beta_1$, $\beta_4$ and $\beta_7$ show that their associated variables (PO$_4$, Fecal Coliform and COD) are of stable characteristics and therefore, do not cause wild fluctuations in the depletion of DO.

However, we need to watch out for pollutants with the sensitivity indices $\beta_2$, $\beta_3$, $\beta_5$ and $\beta_6$ (i.e. NO$_3$, Temperature, pH and BOD in that sequence). Their values fluctuate significantly and could cause instability in the value of dissolved oxygen, DO. They engender wild fluctuations which lead to high variability in the value of DO and hence they are very sensitive parameters. These parameters account for very severe variations in the value of DO. In other words, they can be said to have higher propensity to adsorb the DO from the body of water by virtue of their erratic characteristics as evidenced in figures 3 (b,c,e and f). $\beta_0$ is an autonomous parameter that depends on the boundary condition of pollutant loading. It may or may not have practical interpretations.

### 4.3 The Confidence Intervals for Sensitivity Analysis

By substituting at once, the values of the upper control limit (UCL) and the lower control limit (LCL) of the sensitivity indices, $\beta_i$ each set at a time, into the regression model, the extreme values of DO can be obtained and these extreme values define the limit beyond which the DO cannot exceed. Thus, when the DO values obtained with point estimate (PNT. EST.), which is as an average value, we will then claim that such average values will not fluctuate beyond the extreme values determined by the UCL and LCL. This is the real essence of the sensitivity analysis because it answers the “what if” question.

The foregoing confidence limits appear to be the pith of the results in the sense that they serve as a tool for model tricking in order to obtain the pattern of fluctuation of the values of DO given the fact that each of the parameters have a range of values within which they fluctuate. Thus by the method of model tricking, it will be possible to eyeball the manner or pattern in which DO deficiency happens as effluents are ejected into the river. This process of model tricking is indeed an optimization technique that provides a means for sensitivity analysis. Thus as the values of the regression parameters fluctuate, so do the value of DO varies to affect the assimilative capacity of the river.

### (iv) Check for Model Adequacy

The computed result of the multivariate coefficient of determination, $R^2$, show that the value is 0.71 (71%) indicating that a good fit was achieved by the regression model developed.

### 5. CONCLUSION

This study has ably demonstrated a combined use of multivariate linear regression modelling and bootstrapping as an effective statistical analytics for conducting sensitivity analysis in order to evaluate the robustness of assimilative capacity modelling in a body of water. The philosophy behind sensitivity analysis is a close examination of what amount of changes in the critical values of regression parameters
corresponding to sensitive pollutants, namely NO₃. Temperature, pH and BOD can cause wild fluctuations in the value of DO without affecting the model robustness. The method proposed crosses the stream where it is shallowest. This study has struggled, in a savoir-faire manner, as it were out of a puzzling issue into unhoped-for unfoldment of a vista of knowledge beyond the specific ways of looking at the world. Earlier methods appear to dawdle.

6. Reference


