# The Use of Cubic Splines and MATLAB for the Determination of BOD First-Order Model Coefficients

***Bashaar Y. Ammary*** [1)*] ***and La'aly A. Al-Samrraie*** [2)]

[1), 2)] Water and Environmental Engineering Department, Balqa Applied University, Al-Huson College, Al-Huson 21510, Jordan. E-Mails: bammary@yahoo.com, bammary@bau.edu.jo; laaly_samraie@yahoo.com
* Corresponding Author.

## ABSTRACT

The purpose of this paper is to introduce a new method for the determination of the values of the first-order Biochemical Oxygen Demand (BOD) model coefficients (k and L).Cubic spline interpolation and MATLAB were used for that purpose after the experimental and theoretical BOD values were made dimensionless. Experimental BOD values taken from the literature are made dimensionless by dividing them by experimental BOD1 through BOD5. Theoretical BOD values are made dimensionless by dividing them by the theoretical BOD1 through BOD5 for different theoretical k values. Experimental dimensionless BOD values are connected smoothly using cubic spline method. Using MATLAB, the closest theoretical dimensionless BOD curve to the cubic spline curve is selected and used for the determination of k and L. The closest curve is the one that has the same area under it as to that under the cubic spline curve with the areas that lie between the two curves being the smallest. The new method has proved to give more accurate values after it has been compared with a number of other methods used for that purpose.The new method produced the smallest error and the highest coefficient of determination than the other methods used. The newly introduced method is expected to replace all other methods, as these methods give different values for k and L depending on the method used.

**KEYWORDS:** Dimensionless BOD method, Cubic spline, MATLAB, BOD first-order model, BOD coefficients.

## INTRODUCTION

The Biochemical Oxygen Demand (BOD) test and its first-order model equation are still the most widely used to determine the organic content of wastewater (Ramalho, 1977; Steel and McGhee, 1991; Cutrera et al., 1999; Bassa and Chetty, 2002; Metcalf and Eddy, 2004; Singh, 2004; Siwiec et al., 2011). The BOD first-order model is shown in Equation (1).

$$\mathrm{BOD} = \mathrm{L}\left(1 - e^{-kt}\right); \qquad (1)$$

where:

BOD = amount of oxygen consumed (or BOD exerted) at time t (mg/l).

t = time elapsed since the start of the test (day).

L = ultimate BOD or $BOD_u$ (mg/l).

k = reaction constant (1/day).

As the ultimate BOD (L or $BOD_u$) and k in Equation (1) are unknown for a certain wastewater, a number of methods have been used to predict their values based on BOD measurements for the first five to six days.

A number of linearization techniques have been used for that purpose. These methods include the Reed-

Received on 12/10/2014.
Accepted for Publication on 25/2/2015.

Theriault least square method, the method of moments, Thomas method, the log-difference method, the slope method, the daily difference method, the rapid ratio method and the graphical method (Leduc et al., 1988). A number of non-linear methods (Marquardt, 1963; Constable and McBean, 1977a; Constable and McBean, 1977b; Barnwell, 1981) and stochastic models (Leuduc et al., 1986; Leduc et al., 1988) have also been used.

The following is a summary of the most widely used methods for that purpose; namely, Thomas method (Thomas, 1950), least square method and Fujimoto method (Marquardt, 1963; Cutrera et al., 1999; Bassa and Chetty, 2002; Metcalf and Eddy, 2004; Zainudin et al., 2010). In addition, a newly introduced method is also discussed. The method is based on the geometric series and logarithm series expansion of the BOD first-order model (Ammary and Al-Samrraie, 2014).

**Least Square Method:** In the least square method, the values of L and k for 5 days can be calculated as follows (Ammary and Al-Samrraie, 2014) using the data for the first five days and the data for the sixth day:

$$-k = \frac{5\Sigma Y'Y - \Sigma Y'\Sigma Y}{5\,\Sigma Y^2 - \Sigma Y\,\Sigma Y}; \quad (2)$$

$$L = \frac{5\Sigma Y'Y + k\,\Sigma Y^2}{k\,\Sigma Y}; \quad (3)$$

where:
$Y = BOD_t$.
$Y' = (Y_{n+1} - Y_{n-1}) / 2\Delta t$.
$\Delta t$ = Time between measurements, in this case 1.0 day.

Ammary and Al-Samrraie (2014) investigated the capability of the least square method (and the capability of other methods as mentioned below) in order to estimate the theoretical BOD first-order equation (not its ability to predict experimental data). They found that as the theoretical value of k increases, the least square method overestimates k by 1% to 6% for the theoretical values of k of 0.2 to 0.6/day, respectively. No error is introduced in estimating the ultimate BOD value.

**Thomas Method:** In this method, $(t\ /BOD)^{1/3}$ as ordinate and t as abscissa (t is the time in days) are plotted and fitted to a straight line (Thomas, 1950). The parameters are then estimated using the slope (b) and the intercept (a) of this line as follows:

$$k_{10} = 2.61\,\frac{b}{a}; \quad (4)$$

$$L = \frac{1}{2.3\ \ k_{10} a^3}; \quad (5)$$

where:
L = ultimate BOD.
$k_{10}$ = reaction constant to the base 10.

Ammary and Al-Samrraie (2014) have also shown that as the theoretical value of k increases, Thomas method underestimates the value of k by 1% to 9% for the theoretical values of k of 0.2 to 0.6/day, respectively. Thomas method also overestimates the value of L by 1% to 6% for the theoretical values of k of 0.2 to 0.6/day, respectively.

**Fujimoto Method:** In the Fujimoto method (Fujimoto, 1961), BOD values at time t+1 on the ordinate and BOD values at time t on the abscissa (t is time in days) are plotted and fitted to a straight line. The intersection of this straight line with a line of slope = 1 is an estimate of the ultimate BOD (L). The k value can then be determined using any measured BOD value and the corresponding time. Fujimoto method is essentially a graphical method, however, the ultimate BOD (L) can be found from the slope (b) and the Y intercept (a) of the arithmetic plot as follows (Ammary and Al-Samrraie, 2014):

$$L = \frac{a}{1-b}. \quad (6)$$

Ammary and Al-Samrraie (2014) have found that the estimates of k and L by Fujimoto method are exactly the same as the theoretical values. In reality, this method, however, introduces error in estimating the k value as it has to be found after applying the first-order model to any point where both BOD and t are known, which produces five different k values.

**Ammary and Al-Samrraie (2014) Method:** In this method (Ammary and Al-Samrraie, 2014), different equations are proposed for different k values, as shown in Table 1.

The values of k and L are evaluated using this method by first drawing1/BOD *versus* 1/t in days. Using any equation in Table 1, the value of L is calculated from the Y intercept of the resulting line. The value of k is then calculated from the slope of the resulting line using the same equation and the value of L just calculated. Depending on the new k value calculated, a new appropriate equation in Table 1 is then used. From the new equation in Table 1, a different L value (from the same Y intercept) and then k value (from the same slope) are calculated. After a few trials, the values of k and L are determined. For other values of k, the Y intercept and slope can be calculated using interpolation. More details on the derivation and use of this method can be found elsewhere (Ammary and Al-Samrraie, 2014).

**Table 1. Equations used to calculate ultimate BOD (L) for different k values for the method proposed by Ammary and Al-Samrraie (2014)**

| k value (1/day) | Equation for the corresponding k value |
|---|---|
| 0.05 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.9201\mathrm{L}}+\frac{1}{1.0009\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.10 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.8467\ \mathrm{L}}+\frac{1}{1.0036\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.15 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.7787\ \mathrm{L}}+\frac{1}{1.0083\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.20 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.7170\ \mathrm{L}}+\frac{1}{1.0146\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.25 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.6603\ \mathrm{L}}+\frac{1}{1.0228\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.30 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.6085\ \mathrm{L}}+\frac{1}{1.0328\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.35 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.5611\ \mathrm{L}}+\frac{1}{1.0445\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.40 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.5177\ \mathrm{L}}+\frac{1}{1.0581\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.45 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.4778\ \mathrm{L}}+\frac{1}{1.0735\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.50 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.4415\ \mathrm{L}}+\frac{1}{1.0903\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.55 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.4083\ \mathrm{L}}+\frac{1}{1.1092\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |
| 0.6 | $\frac{1}{\mathrm{BOD}}=\frac{1}{1.3782\ \mathrm{L}}+\frac{1}{1.1295\ \mathrm{Lk}}\frac{1}{\mathrm{t}}$ |

**The Dimensionless Approach**

The dimensionless approach is based on dividing the experimental $BOD_1$ through $BOD_5$ by one of the experimental values of $BOD_1$, $BOD_2$, $BOD_3$, $BOD_4$ or $BOD_5$, thus making them dimensionless. These experimental dimensionless BOD values (along with the value (0.0, 0.0) for day 0.0) are then connected smoothly using cubic spline method (Shikin and Pils, 1995; Knott, 1999). The area below the cubic spline, consisting of five cubic polynomial functions connected smoothly at the point of intersection, is calculated. This area is then compared to the theoretical dimensionless curves of the BOD first-order model equation (Equation 1) for different k values. The theoretical dimensionless curves are obtained by dividing the theoretical BOD values up to $BOD_5$ by the corresponding theoretical value of $BOD_1$through $BOD_5$ for different k values of the BOD first-order model equation (Equation 1). The resulting theoretical curves drawn between $BOD/BOD_{1\text{-}5}$ as ordinate *versus* time as abscissa for the different k values thus start from 0.0 at day 0.0 and end at 1.0 at day 1 if the values are divided by $BOD_1$ (at 1.0 at day 2 if the values are divided by $BOD_2$, at 1.0 at day 3 if the theoretical values are divided by $BOD_3$,… and so on). The areas below these curves are compared to the area below the dimensionless experimental BOD values. The theoretical dimensionless curve that has similar area below it as to the area below the experimental dimensionless curve obtained using cubic spline method is used to calculate the value of k and then L (see Figure 5). An example of how the theoretical dimensionless curves are calculated is shown below for the case when k = 0.2 / day when the theoretical values are divided by $BOD_5$:

$$BOD_1 = L\left(1-e^{-k\,t}\right) = L\left(1-e^{-0.2(1)}\right) = 0.1813\ L. \quad (7)$$

$$BOD_3 = L\left(1-e^{-k\,t}\right) = L\left(1-e^{-0.2(3)}\right) = 0.4512\ L. \quad (8)$$

$$BOD_5 = L\left(1-e^{-k\,t}\right) = L\left(1-e^{-0.2(5)}\right) = 0.6321\ L \quad (9)$$

The value of $BOD_1/BOD_5$ would thus be equal to 0.2868, while the value of $BOD_3/BOD_5$ would be equal to 0.7138. When the theoretical dimensionless curves are drawn, the curves are drawn for continuous values of time and the corresponding dimensionless values of $BOD/BOD_5$, not just at days 1, 2, 3, 4 and 5. This produces the true smooth first-order curve that starts from (0.0, 0.0) at day 0.0 and ends at (5.0, 1.0) at day 5.0.

Once the value of k is determined from the theoretical curve that has the same area below it as that of the experimental dimensionless points connected smoothly using cubic spline method, the ultimate BOD can then be determined. For example, the values of $BOD_u/BOD_5$ or $L/BOD_5$ can be calculated using Equation (9) for the same k value mentioned above (0.2/day) as:

$$\frac{L}{BOD_5} = \frac{1}{(1-e^{-k\,t})} = \frac{1}{(1-e^{-0.2(5)})} = 1.5820. \quad (10)$$

The value of L is then calculated by multiplying the value of experimental $BOD_5$ by the theoretical value of $L/BOD_5$, as shown in Equation (10) for the case of k = 0.2 / day.

In Equation 10, once the value of k was determined, the L value was calculated by multiplying the experimental $BOD_5$ by the theoretical $L/BOD_5$. If the theoretical values were made dimensionless by dividing them by $BOD_1$ or $BOD_3$, for example, then the value of L is obtained by multiplying the experimental value of $BOD_1$ or $BOD_3$ by the theoretical values of $L/BOD_1$ or $L/BOD_3$, respectively.

For example, if the dimensionless approach is based on $BOD_1$ and assuming that the k value is equal to 0.2/day as was assumed above in Equations 7-10, then the value of $BOD_3/BOD_1$ would be equal to 2.4887 and the value of $BOD_5/BOD_1$ would be equal to 3.4865. The value of $BOD_u/BOD_1$ or $L/BOD_1$ can also be calculated using Equation (11) for the same conditions to be equal to:

$$\frac{L}{BOD_1} = \frac{1}{(1-e^{-k\,t})} = \frac{1}{(1-e^{-0.2(1)})} = 5.517. \quad (11)$$

In the present paper, the theoretical values of $BOD/BOD_1$ through $BOD/BOD_5$ for continuous k values (from 0 up to 0.7) were calculated using MATLAB (Hahn and Valentine, 2007; McMahon, 2007). Only a limited number of k values are shown in this paper for reasons of clarity and for demonstration purposes.

**Application of the Dimensionless Curves Based on $BOD_1$-$BOD_5$ on Experimental BOD Data**

As the experimental values of the $BOD_1$ through $BOD_5$ are measured, the dimensionless values of these experimental BOD data are calculated. Five different dimensionless sets are calculated. The first one is obtained by dividing each experimental value by the experimental $BOD_1$ value. The second dimensionless set is obtained by dividing the same experimental data by $BOD_2$. Similarly, the third through fifth dimensionless sets are obtained by dividing the same experimental data by $BOD_3$ through $BOD_5$ values, respectively. For each of these five dimensionless sets, a cubic spline curve is drawn. Each of these five curves is then drawn on the corresponding five theoretical dimensionless curves drawn using MATLAB for different k values. The k value of the closest theoretical dimensionless curve to the experimental cubic spline curve is then used to calculate k and L of the experimental data. The closest curve is the curve that has the same area below it as the cubic spline curve. In addition, the closest curve is that one that has minimum area between the cubic spline curve and the theoretical curve, as will be shown below. The cubic spline curve and the area under it, in addition to the areas under the theoretical dimensionless BOD curves and the areas between these two curves were found using a MATLAB code written for that purpose.

In the cubic spline method, a cubic polynomial function connects every two measured adjacent dimensionless BOD values. Every two adjacent points are connected by a different cubic polynomial function. The different cubic polynomial functions should pass through each measured dimensionless BOD value and be continuous in first and second derivatives at the interior dimensionless BOD values (De Boor, 1978). For the case when $BOD_5$ is used as a basis to convert the experimental data to dimensionless data (in other words when all experimental values are divided by $BOD_5$), a cubic polynomial function connects points $BOD_1/BOD_5$ and $BOD_2/BOD_5$. Another cubic polynomial function connects points $BOD_2/BOD_5$, $BOD_3/BOD_5$… and so on. Equation 12 shows the form of a cubic polynomial function.

$$S_i(x)=a_i(x-x_i)^3+b_i(x-x_i)^2+c_i(x-x_i)+d_i \quad \text{for} \quad i=1,2,\ldots,n-1; \qquad (12)$$

where a, b, c and d are the coefficients of the third order polynomial functions. To evaluate these coefficients, equations for the interior points or nodes and for the two boundary conditions should be written. For the boundary conditions, it was assumed here that the second derivative at the endpoints is equal to that at the points immediately adjacent to them. This method is called Parabolic Runout Spline (Knott, 1999; Gerald and Wheatley, 2004). This type of boundary condition results in a parabolic curve at the endpoint. It is assumed that this type is more appropriate than the many other types of boundary conditions usually used in cubic spline interpolation. A MATLAB code has been developed to solve the matrix that results for the interior nodes and the two boundary conditions. MATLAB is a very powerful tool for many engineering applications that include matrices, as the case here is.

The cubic spline method draws a smooth curve that passes through each of the dimensionless experimental data. For the case when $BOD_5$ is used as a basis to convert the experimental data to dimensionless data (in other words when all experimental values are divided by $BOD_5$), the cubic spline curve starts from day zero on the time axis, and thus the start coordinates are (0.0, 0.0), and the end coordinates are (5.0, 1.0). For the case when $BOD_3$ is used as a basis to convert the experimental data to dimensionless data (in other words when all experimental values are divided by $BOD_3$), the cubic spline curve starts from day zero on the time axis, and thus the start coordinates are (0.0, 0.0). The cubic spline curve passes through the coordinates (3.0, 1.0) and ends at the coordinates (5.0, $BOD_5/BOD_3$). This smooth cubic spline curve is actually a number of cubic polynomial curves that connect smoothly at the point of intersection.

The MATLAB code developed here performs the following tasks:

1- The code starts by calculating the first set of the dimensionless experimental BOD values by dividing the experimental $BOD_1$ through $BOD_6$ values by the experimental $BOD_1$ value. The second set of dimensionless experimental BOD values are also calculated by dividing the same experimental values by the experimental $BOD_2$ value. The third, fourth and fifth sets of dimensionless experimental BOD values are calculated by dividing the experimental BOD values by $BOD_3$, $BOD_4$ and $BOD_5$, respectively. The sixth day BOD (i.e., $BOD_6/BOD_{1-5}$) is used to ensure that when cubic spline curves are drawn, the value of cubic spline curve before day 5 does not go above the maximum theoretical value (for the case based on $BOD_5$, does not go above 1.0 just before day 5). In addition, the use of $BOD_6$ value minimizes the effect of the boundary conditions on the cubic spline curves. For example, dimensionless BOD values based on $BOD_5$ (i.e., when experimental BOD values are divided by $BOD_5$), cubic spline curves for high k values could theoretically go above 1.0 just before the fifth day if cubic splines are drawn for only the 5 days' values of BOD. Therefore, the value of $BOD_6$ was used in drawing all the cubic spline curves between 0.0 and 5.0 days. It should be noted that only the area below these curves (between 0.0 and 5.0 days) was calculated and drawn (see below). Table 2 shows BOD values taken from the literature for the first six days and Table 3 shows an example on how to apply the dimensionless method based on $BOD_1$ through $BOD_5$ on Data No. 1 of Table 2.

**Table 2. BOD values *versus* incubation time taken from literature**

| Data No. | Reference | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 |
|---|---|---|---|---|---|---|---|
| 1 | Tchobanoglous and Schroeder (1985) | 15.8 | 26.7 | 37.4 | 45.9 | 50.1 | 56.1 |
| 2 | Metcalf and Eddy (2004) | 65 | 109 | 138 | 158 | 178 | 190 |

**Table 3. Example on applying the dimensionless approach to BOD data no. 1 in Table 2; BOD values for days 0 to 6 are divided by $BOD_1$ through $BOD_5$ values**

| Dimensionless BOD data set | Day | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| | BOD (mg/l) | 0.0 | 15.8 | 26.7 | 37.4 | 45.9 | 50.1 | 56.1 |
| 1 | $BOD/BOD_1$ | 0.0 | 1.000 | 1.6899 | 2.3671 | 2.9051 | 3.1709 | 3.5506 |
| 2 | $BOD/BOD_2$ | 0.0 | 0.5918 | 1.0000 | 1.4007 | 1.7191 | 1.8764 | 2.1011 |
| 3 | $BOD/BOD_3$ | 0.0 | 0.4225 | 0.7139 | 1.0000 | 1.2273 | 1.3396 | 1.5000 |
| 4 | $BOD/BOD_4$ | 0.0 | 0.3442 | 0.5817 | 0.8148 | 1.0000 | 1.0915 | 1.2222 |
| 5 | $BOD/BOD_5$ | 0.0 | 0.3154 | 0.5329 | 0.7465 | 0.9162 | 1.0000 | 1.1198 |

2- For each of the five sets of dimensionless experimental data (as shown in Table 3, for example), the code then determines the cubic polynomial functions that connect each of the dimensionless experimental data. For the five sets of dimensionless experimental data, five cubic polynomial curves are determined. Each of these curves consists of five different cubic polynomial functions connected smoothly at the point of intersection. Figure 1 shows an example of the five cubic polynomial functions connecting the dimensionless BOD values based on $BOD_1$ (the experimental BOD values ($BOD_1$-$BOD_5$) divided by experimental $BOD_1$ value), as shown in dimensionless BOD data set 1 of Table 3.

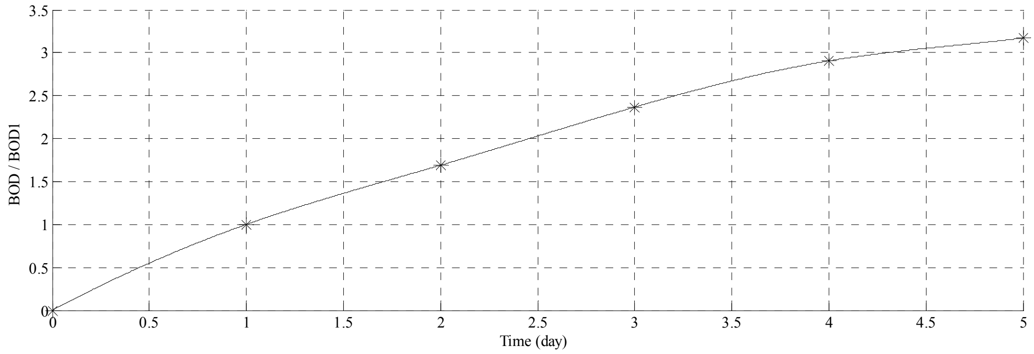



**Figure (1): Cubic spline curve connecting the dimensionless BOD data based on $BOD_1$ (set no. 1 in Table 3). Five cubic polynomial functions connected smoothly at the points of intersection**

3- The code then determines the area below the cubic spline curve between day zero and day five for each set of the dimensionless BOD values (set 1 to set 5 in Table 3).

4- Then, for each set of experimental dimensionless data, a theoretical dimensionless curve with the same area below it as that with the cubic spline curves is determined. The code selects this curve by trying different values of k and calculating the area below each of these curves. Figure 2 shows, as an example, the theoretical dimensionless curves for five different k values (namely 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5/ day) drawn along with the cubic spline curve shown in Figure 1.

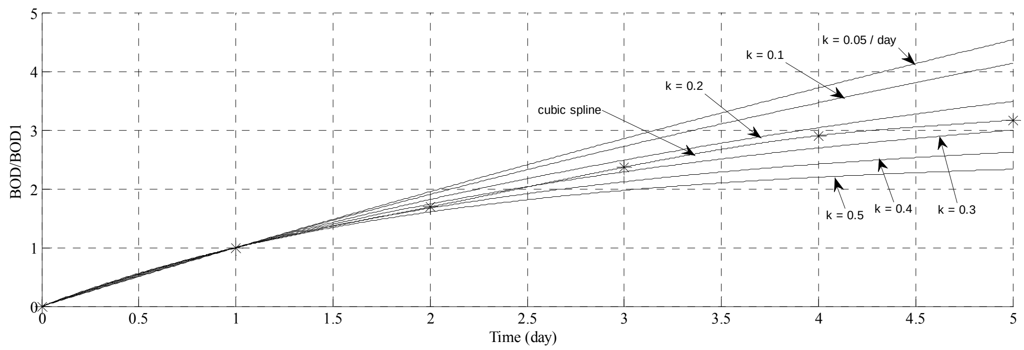



**Figure (2): Cubic spline curve connecting the dimensionless BOD data based on $BOD_1$(set no. 1 in Table 3) and the corresponding theoretical curves with k values of 0.05/day to 0.5/day**

The code then calculates the area below each of these curves and compares it with the area under the cubic spline curve. In reality, the code determines the area under hundreds or thousands of these theoretical dimensionless curves, and the curve with the area under it closest to the area under the cubic spline curve is determined.

5- The k value of this curve that has the area closest to the area below the cubic spline curve is used as a possible value of k of the first-order model equation. As the areas below these two types of curves are equal, this implies that the area that lies below the cubic spline curve and above the theoretical dimensionless curve (area A in Figure 3) is equal to the area that lies above the cubic spline curve and below the theoretical dimensionless curve (area B in Figure 3).

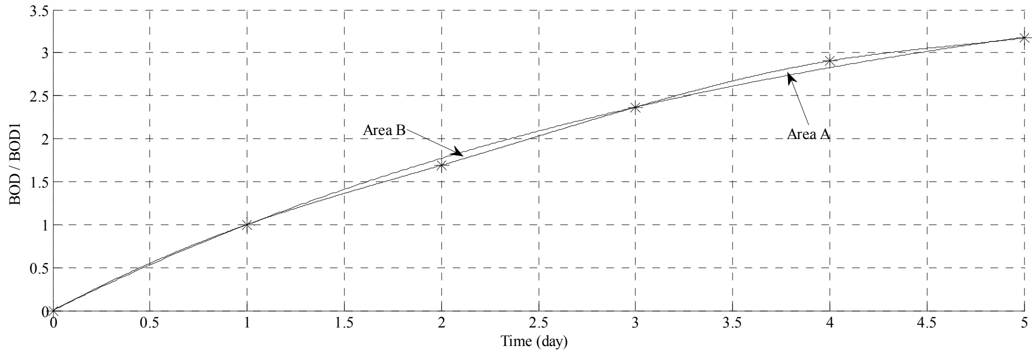



**Figure (3)**: **Cubic spline curve connecting the dimensionless BOD data based on $BOD_1$(set no. 1 in Table 3) and the corresponding theoretical curve that has the same area under it, with a k value of 0.259 / day. Area A and area B should be equal in order for the areas below the cubic spline curve and the theoretical dimensionless curve to be equal. Minimum area A (or area B) indicates that the two curves are closest**

6- Now, the code has 5 theoretical dimensionless curves each with a different value of k for the 5 sets of experimental dimensionless curves. Figures 4 and 5 show two of these theoretical dimensionless curves based on $BOD_3$ and $BOD_5$, respectively, as well as the corresponding cubic spline curves of the experimental BOD data. Each of these 5 theoretical curves has the same area below it as the corresponding cubic spline curve based on the same BOD denominator. Table (4) is a summary of these 5 theoretical dimensionless curves that have the same area below them as the area below the cubic spline curves for the five dimensionless BOD data sets shown in Table 3.

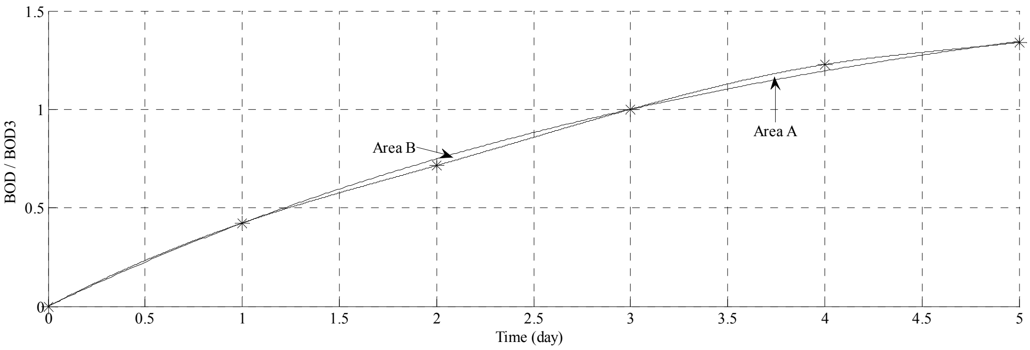



**Figure (4)**: **Cubic spline curve connecting the dimensionless BOD data based on $BOD_3$ (set no. 3 in Table 3) and the corresponding theoretical curve that has the same area under it. Area A and area B should be equal**

7- The code then selects the curve that has the minimum values of area A (or area B). The theoretical dimensionless curve that produces minimum area A means that the cubic spline curve and the theoretical dimensionless curve are the closest compared to any other theoretical dimensionless curve. Table 4 shows that the theoretical dimensionless curve that produces minimum area A is the one based on $BOD_5$. The corresponding k value is equal to 0.264 / day and the $BOD_5$ factor ($BOD_u/BOD_5$) is equal to 1.3645. Therefore, the ultimate BOD is equal to:

$BOD_u$ = ($BOD_5$) ($BOD_u/BOD_5$) = ($BOD_5$) ($BOD_5$ Factor) = (50.1 mg/l) (1.3645) = 68.4 mg/l.

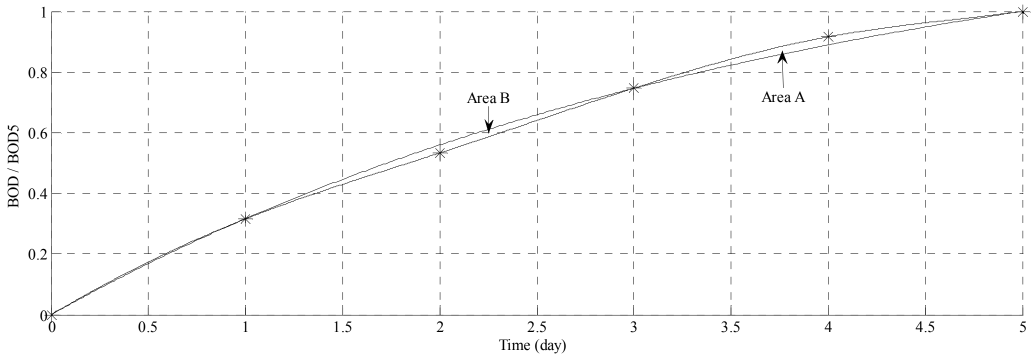



**Figure (5): Cubic spline curve connecting the dimensionless BOD data based on $BOD_5$ (set no. 5 in Table 3) and the corresponding theoretical curve that has the same area under it. Area A and area B should be equal**

**Table 4. Summary of the 5 theoretical dimensionless curves having the same area under them as the area under the corresponding cubic spline curve set shown in Table 3**

<table>
<tr><td>BOD data (mg/l)</td><td>0</td><td>15.8</td><td>26.7</td><td>37.4</td><td>45.9</td><td>50.1</td><td>56.1</td></tr>
<tr><td>Dimensionless Basis</td><td colspan="2">BOD/BOD1</td><td>BOD/BOD2</td><td colspan="2">BOD/BOD3</td><td>BOD/BOD4</td><td>BOD/BOD5</td></tr>
<tr><td>k (1/day)</td><td colspan="2">0.259031</td><td>0.158081</td><td colspan="2">0.257739</td><td>0.175063</td><td>0.264002</td></tr>
<tr><td>BOD Factor</td><td colspan="2">4.3821</td><td>3.6892</td><td colspan="2">1.8571</td><td>1.9859</td><td>1.3645</td></tr>
<tr><td>BODu (mg/l)</td><td colspan="2">69.2</td><td>98.5</td><td colspan="2">69.5</td><td>91.1</td><td>68.4</td></tr>
<tr><td>Area A</td><td colspan="2">0.0988</td><td>0.0755</td><td colspan="2">0.0415</td><td>0.0374</td><td>0.0336</td></tr>
<tr><td>Area B</td><td colspan="2">0.1000</td><td>0.0759</td><td colspan="2">0.0415</td><td>0.0374</td><td>0.0329</td></tr>
</table>

In order to illustrate this method further, Table (5) shows the dimensionless BOD sets for data no. 2 of Table 2 above. Again, this data is taken from the literature.

Table 6 is a summary of the results obtained by applying the cubic spline method and MATLAB on the dimensionless BOD data sets shown in Table 5.

For the sets in Table 5, a cubic spline curve connects each of the dimensionless BOD data sets obtained by dividing the experimental data by $BOD_1$ through $BOD_5$, and the area under each of these curves is calculated using the MATLAB code. The MATLAB code then finds, for each of the cubic spline curves, the corresponding theoretical dimensionless curve that has the same area under it as that under the cubic spline curve. The areas under these curves and the corresponding k values and BOD factors are all shown in Table 6.

From Table 6, the curve that has the minimum values of area A (or area B) is the one based on $BOD_2$.

The corresponding k value is equal to 0.388327/ day and the $BOD_2$ factor ($BOD_u/BOD_2$) is equal to 1.8517. Therefore, the ultimate BOD is equal to: $BOD_u$ = ($BOD_2$) ($BOD_u/BOD_2$) = ($BOD_2$) ($BOD_2$ Factor) = (109mg/l) (1.8517) = 201.8mg/l.

**Table 5. Example on applying the dimensionless approach to BOD data no. 2in Table 2; BOD values for days 0 to 6 are divided by $BOD_1$ through $BOD_5$ values**

| **Dimensionless BOD data set** | **Day** | **0** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|---|
| | BOD (mg/l) | 0.0 | 65 | 109 | 138 | 158 | 178 | 190 |
| 1 | $BOD/BOD_1$ | 0.0 | 1.000 | 1.6769 | 2.1231 | 2.4308 | 2.7385 | 2.9231 |
| 2 | $BOD/BOD_2$ | 0.0 | 0.5963 | 1.0000 | 1.2661 | 1.4495 | 1.6330 | 1.7431 |
| 3 | $BOD/BOD_3$ | 0.0 | 0.4710 | 0.7899 | 1.0000 | 1.1449 | 1.2899 | 1.3768 |
| 4 | $BOD/BOD_4$ | 0.0 | 0.4114 | 0.6899 | 0.8734 | 1.0000 | 1.1266 | 1.2025 |
| 5 | $BOD/BOD_5$ | 0.0 | 0.3652 | 0.6124 | 0.7753 | 0.8876 | 1.0000 | 1.0674 |

**Table 6. Summary of the 5 theoretical dimensionless curves having the same area under them as the area under the corresponding cubic spline curve set shown in Table 5**

| BOD data (mg/l) | 0 | 65 | 109 | 138 | 158 | 178 | 190 |
|---|---|---|---|---|---|---|---|
| Dimensionless Basis | $BOD/BOD_1$ | $BOD/BOD_2$ | $BOD/BOD_3$ | $BOD/BOD_4$ | $BOD/BOD_5$ | | |
| K | 0.389105 | 0.388327 | 0.164535 | 0.412381 | 0.335646 | | |
| BOD Factor | 3.1024 | 1.8517 | 2.5669 | 1.2378 | 1.2296 | | |
| $BOD_u$ | 201.7 | 201.8 | 354 | 195.6 | 218.9 | | |
| Area A | 0.0310 | 0.0187 | 0.1699 | 0.0217 | 0.0269 | | |
| Area B | 0.0336 | 0.0194 | 0.1700 | 0.0220 | 0.0274 | | |

**Table 7. Ultimate BOD values using different methods. The values of the new method are shown in Tables 5 and 6 for data 1 and 2 in Table 2, respectively**

| | **L (Ultimate BOD value), mg/l** | | | | |
|---|---|---|---|---|---|
| **Data No.** | **Cubic Spline Method** | **Least Square Method** | **Fujimoto Method** | **Thomas Method** | **Ammary and Al-Samrraie (2014) Method** |
| 1 | 68.4 | 73.8 | 74 | 73.5 | 68 |
| 2 | 201.8 | 219.3 | 217 | 214 | 202 |

**Table 8. k values using different methods. The values of the new method are shown in Tables 5 and 6 for data 1 and 2 in Table 2, respectively**

| | **k value (1/day)** | | | | |
|---|---|---|---|---|---|
| **Curve** | **Cubic Spline Method** | **Least Square Method** | **Fujimoto Method** | **Thomas Method** | **Ammary and Al-Samrraie (2014) Method** |
| Curve 1 | 0.264 | 0.23 | 0.24 | 0.236 | 0.26 |
| Curve 2 | 0.388 | 0.337 | 0.333 | 0.356 | 0.38 |

**Comparison with Other Methods**

The results obtained by the cubic spline method above were compared to those of a number of different methods used for the determination of k and L values of BOD data. These methods were: Least square method, Thomas method, Fujimoto method and Ammary and Al-Samrraie (2014) method. Tables 7 and 8 show a comparison between the results obtained by the new method, for data no. 1 and data no. 2 in Table 2, and the results obtained by the least square, Thomas and Fujimoto methods, in addition to the method proposed by Ammary and Al-Samrraie (2014). It should be noted that the value of k obtained from the dimensionless method was used as k value for the first trial in the method proposed byAmmary and Al-Samrraie (2014).

## STATISTICAL ANALYSIS

In order to study the goodness of fitness of the different methods, a number of BOD data points that represent a truely first-order model were used. An ultimate BOD value (L) of 1000 mg/l and three k values of 0.2, 0.4 and 0.6/day were used. These BOD values for days 1-6 were calculated and are present under the first-order (FO) columns of Table 9 for the three k values.

**Table 9. $BOD_1$-$BOD_6$ obtained by the first-order (FO) model for k values of 0.2, 0.4 and 0.6/day and L = 1000 mg/l used to calculate k and L using least square (LS), Thomas (Th) and Ammary and Al-Samrraie (A-S) methods. The $BOD_1$-$BOD_6$ listed for these methods were back calculated using the obtained k and L values for the different methods**

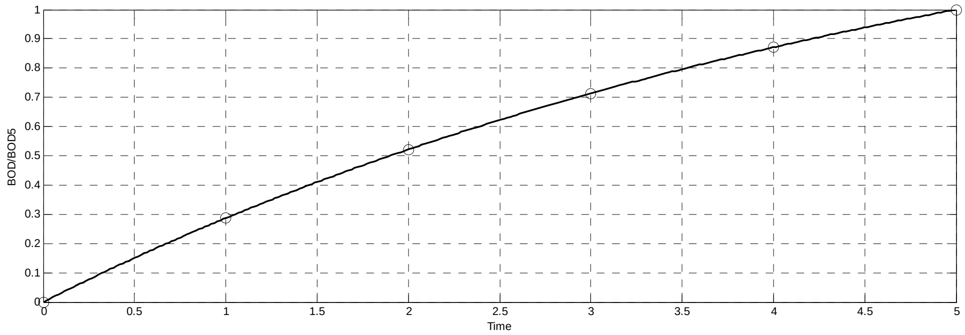
| | k = 0.2 / day | | | | k = 0.4 / day | | | | k = 0.6 / day | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Day** | **FO** | **LS*** | **Th$** | **A-S**** | **FO** | **LS*** | **Th$** | **A-S**** | **FO** | **LS*** | **Th$** | **A-S**** |
| 1 | 181.3 | 182.4 | 181.1 | 181.7 | 329.7 | 336.8 | 328.3 | 331.9 | 451.2 | 470.9 | 446.4 | 473.4 |
| 2 | 329.7 | 331.5 | 329.7 | 329.3 | 550.7 | 560.2 | 552.1 | 548.2 | 698.8 | 720.1 | 705.2 | 730.6 |
| 3 | 451.2 | 453.4 | 451.6 | 449.4 | 698.8 | 708.3 | 704.7 | 689.2 | 834.7 | 851.9 | 855.2 | 870.3 |
| 4 | 550.7 | 553.1 | 551.6 | 546.9 | 798.1 | 806.6 | 808.8 | 781.1 | 909.3 | 921.7 | 942.1 | 946.2 |
| 5 | 632.1 | 634.6 | 633.7 | 626.2 | 864.7 | 871.7 | 879.8 | 841.0 | 950.2 | 958.6 | 992.5 | 987.5 |
| 6 | 698.8 | 701.2 | 701.1 | 690.7 | 909.3 | 914.9 | 928.2 | 880.0 | 972.7 | 978.1 | 1022 | 1010 |

* For least square method, the calculated values for k were 0.201334, 0.4107 and 0.63666/day, respectively and with calculated L as 1000 mg/l for all k values.

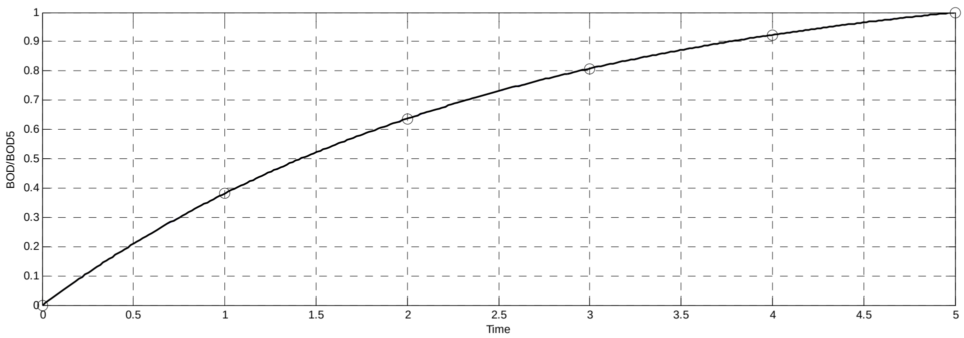
$ For Thomas method, the calculated values were k = 0.1978/day and L = 1009mg/l, k = 0.38283/day and L = 1032mg/l and k = 0.54533/day and L = 1062 mg/l, respectively.

** For Ammary and Al-Samrraie (2014) method, the calculated values were k = 0.2072/day and L = 971mg/l, k = 0.428/day and L = 953mg/l, k = 0.6101/day and L = 1037 mg/l, respectively.

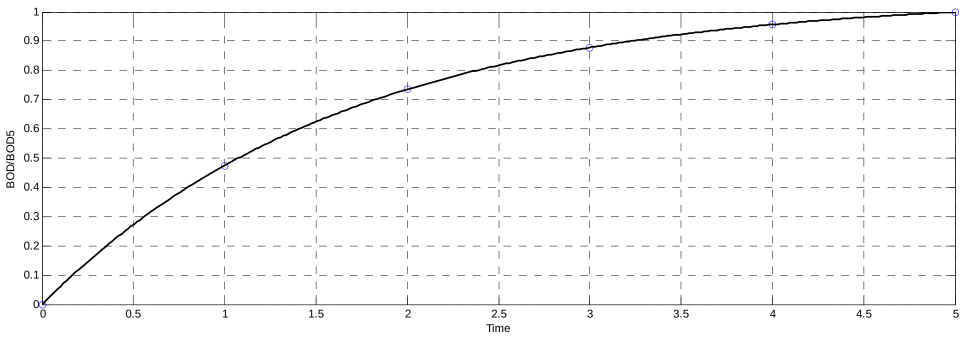
From these first-order values shown in Table 9, the values of k and L were back calculated using these different methods. Using Equations (2) and (3) above, the k and L values for the least square method were calculated and found to be (k = 0.201334/day and L = 1000 mg/l; k =0.4107/day and L = 1000 mg/l; k= 0.63666/day and L = 1000 mg/l), respectively. For Thomas method, $(t/BOD)^{1/3}$ values as ordinate were drawn against t as abscissa and a straight line is obtained. Using equations (4) and (5), the values of k and L were calculated and found to be (k = 0.1978/day and L = 1009 mg/l; k = 0.38283/day and L = 1032 mg/l; k = 0.54533/day and L = 1062 mg/l), respectively. Fujimoto method was also applied to the same first-order (FO) data and k = 0.2, 0.4 and 0.6/day and L = 1000 mg/l for all k values were obtained. For Ammary and Al-Samrraie (2014) method, the calculated values were: (k = 0.2072/day and L = 971 mg/l; k = 0.428/day and L = 953mg/l; k = 0.6101/day and L = 1037 mg/l), respectively. Using the cubic spline method (dimensionless method), the values of k obtained were 0.2, 0.4 and 0.6/day and L = 1000 mg/l for all k values as shown in Figure 6 (a-c). Figure 6 is based on $BOD_5$, but the same results were obtained based on $BOD_{1-4}$. In Figure 6, the experimental dimensionless BOD values (which are theoretical in this case), the theoretical BOD curve and the cubic spline curves produced a complete match, so that only one curve can be observed.




**(a): k = 0.2/day, L =1000 mg/l based on $BOD_5$**




**(b): k = 0.4/day, L =1000 mg/l based on $BOD_5$**




**(c): k = 0.6/day, L = 1000 mg/l based on $BOD_5$**

**Figure (6): Cubic spline curves connecting the dimensionless theoretical BOD first-order (FO) data shown in Table 9 based on $BOD_5$ and the corresponding theoretical curves. Both curves are identical and therefore only one curve can be seen in Figure 6 (a-c)**

The BOD values for days 1-6 were then calculated using the results obtained for k and L for the different methods. The difference between the BOD values for days 1-6 that were used to calculate the k and L values in the first place are then compared to the BOD values for days 1-6 for the new k and L values. As Table 9 shows, there are differences between the predicted values of $BOD_{1-6}$ by the different methods. Table 9 does not show Fujimoto results and the cubic spline (dimensionless) method results as both gave exact values as the first-order (FO) values. To assess the goodness of fit of each of these methods, a number of statistical parameters were calculated and compared (Cutrera et al., 1999). These include the Sum of the Squares of the Errors ($err^2$), the Coefficient of Determination (CD) and the Model Selection Criteria (MSC). Equations 13-15 were used to calculate and compare these values for the different methods (Cutrera et al., 1999).

$$err^2 = \sum_{i=1}^{n}(BOD_{obsi} - BOD_{calci})^2; \quad (13)$$

$$CD = \frac{\sum_{i=1}^{n}\left(BOD_{obsi} - \overline{BOD_{obs}}\right)^2 - \sum_{i=1}^{n}(BOD_{obsi} - BOD_{calci})^2}{\sum_{i=1}^{n}\left(BOD_{obsi} - \overline{BOD_{obs}}\right)^2}; \quad (14)$$

$$MSC = \ln\left[\frac{\sum_{i=1}^{n}\left(BOD_{obsi} - \overline{BOD_{obs}}\right)^2}{\sum_{i=1}^{n}(BOD_{obsi} - BOD_{calci})^2}\right] - \frac{2p}{n}; \quad (15)$$

where:
$BOD_{obsi}$ = observed (experimental) values.
$\overline{BOD_{obs}}$ = average of observed (experimental) values.
$BOD_{calci}$= calculated values of each fitting procedure.
$p$ = number of parameters (2 in this case).
$n$= number of data points.

The results are shown in Table 10. These results show that the dimensionless method using cubic splines and Fujimoto method produced the best fit to the data. Comparing the other three methods, Table 10 shows that at k = 0.2/day, Thomas method produced better fitness than the other methods, as it had the smallest squared error, the highest CD and the highest MSC. For higher k values (0.4 and 0.6/day), the least square method resulted in better fitness to the data than the other two methods (lower squared error and higher CD and MSC).

**Table 10. The sum of squares of the error, coefficient of determination and model selection criteria for different methods for first-order BOD values with k = 0.2, 0.4 and 0.6/day and L = 1000 mg/l**

| | **Sum of the squares of the errors** | | | | |
|---|---|---|---|---|---|
| **k value (1/day)** | **Least Square** | **Thomas** | **Fujimoto** | **Ammary and Al-Samrraie (2014)** | **Cubic Spline Method** |
| K = 0.2 | 26.63 | 8.65 | 0.0 | 118.5 | 0.0 |
| K = 0.4 | 385.3 | 742.8 | 0.0 | 1807 | 0.0 |
| K = 0.6 | 1391.6 | 5753 | 0.0 | 6912 | 0.0 |
| | **Coefficient of Determination (CD)** | | | | |
| **k value (1/day)** | **Least Square** | **Thomas** | **Fujimoto** | **Ammary and Al-Samrraie (2014)** | **Cubic Spline Method** |
| k = 0.2 | 0.99986 | 0.99995 | 1.0 | 0.99937 | 1.0 |
| k = 0.4 | 0.99839 | 0.99690 | 1.0 | 0.99245 | 1.0 |
| k = 0.6 | 0.99295 | 0.97085 | 1.0 | 0.96498 | 1.0 |
| | **Model Selection Criteria (MSC)** | | | | |
| **k value (1/day)** | **Least Square** | **Thomas** | **Fujimoto** | **Ammary and Al-Samrraie (2014)** | **Cubic Spline Method** |
| k = 0.2 | 8.1979 | 9.322 | Infinity | 6.705 | Infinity |
| k = 0.4 | 5.7659 | 5.1096 | Infinity | 4.220 | Infinity |
| k = 0.6 | 4.288 | 2.8687 | Infinity | 2.685 | Infinity |

## CONCLUSIONS

A new dimensionless method is introduced for the determination of the corresponding coefficients of the first-order BOD equation. A MATLAB code determines the closeness of the cubic spline interpolation of the experimental dimensionless BOD data to the theoretical dimensionless BOD first-order equation. The MATLAB code is capable of determining the k value that best describes the experimental data and calculates the ultimate BOD directly. The method produced the least error and the highest coefficient of determination compared to the other methods studied. The advantages of such method include better visualization, understanding and comparison of experimental data to the BOD first order model.

## REFERENCES

Ammary, B.Y., and Al-Samrraie, L.A. (2014). "Evaluation and comparison of methods used for the determination of BOD first-order model coefficients". Int. J. Environment and Waste Management, 13 (4), 362-375.

Barnwell, T. O. (1981)."Least square estimates of BOD parameters". J. Environ. Eng. Div. (ASCE), 107, 1197-1202.

Bassa, F., and Chetty, S. (2002). "Determination of biodegradable COD fraction of industrial effluents by measurements of BOD and COD". Proceedings of the Biennial Conference of the Water Institute of Southern Africa (WISA), 135-143.

Constable, T.W., and McBean, E. A. (1977a). "Bayesian model discrimination for BOD analysis". Can. J. Civil Eng., 4, 371-379.

Constable, T. W., and McBean, E. A. (1977b). "Parameter estimation for the first-order BOD equation using non-linear techniques". Can. J. Civil Eng., 4, 462-470.

Cutrera, G., Manfredi, L., Del Valle, C. E., and González, J.F. (1999). "On the determination of the kinetic parameters for the BOD test". Water SA, 25 (3), 377-379.

De Boor, C. (1978). "A practical guide to splines". Springer Verlag, New York.

Fujimoto,Y. (1961)."Graphical use of first-stage BOD equation". JWPCF, 36 (1), 69-71.

Gerald, C.F., and Wheatley, P.O. (2004). "Applied numerical analysis". Addison-Wesley, New York.

Hahn, B.D., and Valentine, D.T. (2007). "Essential MATLAB for engineers and scientists". Elsevier.

Knott, G.D. (1999). "Interpolating cubic splines". Birkhauser, Boston, USA.

Leduc, R., Unny, T.E., and McBean, E.A. (1986). "Stochastic model for first-order BOD kinetics". Water Research, 20, 625-632.

Leduc, R., Unny, T.E., and McBean, E.A. (1988)."Stochastic models for first-order kinetics of biochemical oxygen demand with random initial conditions, inputs and coefficients". Applied Mathematical Modelling, 12, December, 565-572.

Marquardt, D. (1963). "An algorithm for least-squares estimation of non-linear parameters". SIAM Journal on Applied Math, 11, 431-438.

McMahon, D. (2007). "MATLAB demystified, a self-teaching guide". McGraw Hill Company, New York, USA.

Metcalf and Eddy, Inc. (2004). "Wastewater engineering treatment disposal and reuse". 3rd ed., McGraw Hill, New York.

Ramalho, R.S. (1977). "Introduction to wastewater treatment". Academic Press, New York.

Shikin, E.V., and Pils, A.I. (1995). "Handbook on splines for the user". CRC Press.

Singh, B. (2004). "Determination of BOD kinetic parameters and evaluation of alternate methods". A Thesis Submitted to Thapar Institute of Engineering and Technology, Patiala.

Siwiec, T., Kiedrynska, L., Abramowicz, K., Rewicka, A., and Nowak, P. (2011). "BOD measuring and modeling methods-review". Ann. Warsaw Univ. of Life Sci. – SGGW, Land Reclam.,43 (2), 143-153.

Steel, E. W., and McGhee, T. J. (1991). “Water supply and sewerage”. McGraw-Hill Company, New York.

Tchobanoglous, G., and Schroeder, E.D. (1985). “Water quality: characteristics, modeling, modification”. Addison-Wesley, USA.

Thomas, H. A. (1950). "Graphical determination of BOD curve constants". Water Sewage Works, 97, 123-129.

Zainudin, Z., Mohamed, M., and Ramli, M. R. (2010)."Effects of induced salinity on BOD5 reaction kinetics of river water samples". Malaysian Journal of Analytical Sciences, 14 (1), 24-31.