Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 17, 2023, page 175-185

Parameter Estimation in Mathematical Model of Wastewater Treatment Using Oxidation Pond Technique

Srivennila, Yeak Su Hoe*
Department of Mathematics, Faculty of Science, UTM, Skudai, Johor Bahru, Malaysia
*Corresponding author: srinila23@gmail.com

Abstract
This research aims to develop a mathematical model for wastewater treatment using the oxidation pond technique. The model is derived from published literature, and parameters are estimated. Simulated data values are obtained for seven variables. The research also focuses on reducing the cost of sustaining wastewater treatment technology in Malaysia by analyzing and forecasting system variables and parameters. The mathematical models, based on differential equations, explain the real system and mechanisms. Parameter estimation is successfully implemented, and the least value of the objective function is interpreted. Three variables (Temperature of River, Concentration of Chemical Oxygen Demand, Concentration of Dissolved Oxygen) show highly accurate forecasting, while two variables (Concentration of Microbes, Concentration of Nitrogen) have somewhat good forecasting based on Mean Absolute Percentage Error (MAPE). Additional factors, such as light presence and the impact of microalgae development, should be included in future studies.
Keywords: Wastewater treatment; Oxidation Pond; Mathematical model; Parameter estimation; Forecasting

Introduction

Water is a necessary chemical molecule composed of hydrogen and oxygen atoms. It exists in a variety of forms and is essential to the human body and the Earth's surface. However, unpredictable weather patterns and global warming have caused rivers all around the world to dry up. Water pollution and poor sanitation have also been related to disease transmission (World Health Organization, 2022). Access to safe drinking water is critical for good hygiene and general health.

The water we drink originates from lakes, rivers, and subterranean sources. This water becomes wastewater once it has been drunk and contaminated (Amit Sonune, 2004). Previously, cholera epidemics in England were linked to polluted well water, necessitating the creation of centralized sewage treatment plants (Ambulkar, 2022). These plants use physical, biological, and chemical methods to remove contaminants from water before returning it to the environment.

The goal of wastewater treatment is to remove suspended particles before releasing the treated water, known as effluent, back into the environment. Sewage treatment plants, industrial wastewater treatment plants, agricultural wastewater treatment plants, and leachate treatment plants are the four types of treatment facilities. Each is concerned with a distinct form of trash. Traditional wastewater treatment consists of three stages: primary, secondary, and tertiary treatment. Primary treatment includes removing solid waste from water, whereas secondary treatment purifies the water using biological and chemical methods. Tertiary treatment, often known as polishing, eliminates further contaminants from the water supply. This research focuses on secondary wastewater treatment, which uses oxidation mechanisms to cleanse water.

Materials and methods

Existing Mathematical Model

Based on (Amir Hamzah, 2015), a basic mathematical model that includes four variables, Microbes M(t), PSB P(t), DO D(t) and COD C(t) were developed

$$\frac{dM(t)}{dt} = M(t)[c_0 - c_1 M(t) - c_2 P(t)] + \frac{c_3 D(t) M(t)}{D(t) + c_4} + \frac{v_{in} M_0(t)}{v_r}$$

$$\frac{dP(t)}{dt} = P(t)[c_5 - c_6P(t) - c_7M(t)] + \frac{c_8D(t)P(t)}{D(t)+c_9} + \frac{pc_{10}v_{in}U(t)}{v_r}$$
$$\frac{dC(t)}{dt} = -c_{11}C(t) - c_{14}P(t)C(t) - \frac{c_{15}D(t)C(t)}{C(t)+c_{16}} + \frac{v_{in}C_0(t)}{v_r}$$
$$\frac{dD(t)}{dt} = c_{17}[D_s - D(t)] - c_{18}D(t)P(t) - c_{19}D(t)M(t) - c_{20}D(t)C(t) + \frac{v_{in}D_0(t)}{v_r}$$

where

$M(t)$	-	Concentration of Microbes in river
v_{in}	-	Average amount of sewage coming in
v_r	-	Volume of river
M_0	-	Concentration of microbes in untreated river water (initial point)
$P(t)$	-	Concentration of PSB in river
p	-	Concentration of PSB in 1L of U
$U(t)$	-	Amount of product (e.g., mPHO) containing PSB used in the river water
$C(t)$	-	Concentration of COD (chemical oxygen demand)
$D(t)$	-	Concentration of DO (dissolved oxygen)
D_s	-	Concentration of saturated oxygen
D_0	-	Concentration of COD in untreated river water (initial)
c_i	-	constants (calculated by parameter fitting-based experiment data in treated river water) where $i = 0, 1, 2, \dots, 20$

The Saturated oxygen, D_s, is estimated based on the method of Elmore and Hayes as cited in (Ranjith Shepur, 2019)

$$D_s = 14.652 - 0.41022T + 0.007991T^2 - 0.000077774T^3$$

where T is temperature.

Monod Equation & Modified Mathematical Model

Monod equation in general is a mathematical model for growth of microorganism where the empirical equation is

$$\mu = \mu_{max}\frac{[S]}{K_s + [S]}$$

where

μ	-	growth rate of a considered microorganism
μ_{max}	-	maximum growth rate of this microorganism
K_s	-	half velocity constant
[S]	-	concentration of the limiting substrate S for growth

Referring to (Fauzi, 2021), a modified mathematical model from the existing model with the combination of Monod equation in $\frac{dC(t)}{dt}$ and $\frac{dD(t)}{dt}$equations, a modified mathematical model is obtained.

$$\frac{dM(t)}{dt} = M(t)[c_0 - c_1M(t) - c_2P(t)] + \frac{c_3D(t)M(t)}{D(t)+c_4} + \frac{v_{in}M_0(t)}{v_r}$$
$$\frac{dP(t)}{dt} = P(t)[c_5 - c_6P(t) - c_7M(t)] + \frac{c_8D(t)P(t)}{D(t)+c_9} + \frac{pc_{10}v_{in}U(t)}{v_r}$$
$$\frac{dC(t)}{dt} = -c_{11}C(t)\frac{c_{12}}{c_{13}+C(t)} - c_{14}P(t)C(t) - \frac{c_{15}D(t)C(t)}{C(t)+c_{16}}\frac{c_{12}}{c_{13}+C(t)} + \frac{v_{in}C_0(t)}{v_r}$$
$$\frac{dD(t)}{dt} = c_{17}[D_s - D(t)] - c_{18}D(t)P(t) - c_{19}D(t)M(t) - c_{20}D(t)C(t)\frac{c_{12}}{c_{13}+C(t)} + \frac{v_{in}D_0(t)}{v_r}$$

Ammonium & Nitrate Equations

Based on (Mona Radwan, 2001), two equations are added in order to take into account of the presence of ammonium and nitrate in the river wate with a modification of introducing a new unknown parameter, c_{21}, in the equations in order to obtain a more optimal equation based on the data obtained.

$$\frac{dA(t)}{dt} = c_{21}[-k_n A(t) x_n^{T(t)-20} + y_g k_g B(t) x_g^{T(t)-20}]$$
$$\frac{dN(t)}{dt} = c_{21}[k_n A(t) x_n^{T(t)-20} - k_n N(t) x_d^{T(t)-20}]$$

where

$A(t)$	-	Concentration of ammonium
k_n	-	Nitrification rate 20℃ per day = 0.54
x_n	-	Arrhenius temperature coefficient of nitrification process = 1.13
y_g	-	Nitrogen content in organic matter = 0.29
k_d	-	Denitrification rate per day = 1
$B(t)$	-	Concentration of biological oxygen demand
x_g	-	Arrhenius temperature coefficient of degradation process = 1.024
$N(t)$	-	Concentration of nitrate
k_g	-	Degradation rate of organic matter at 20℃ per day = 0.5
x_d	-	Arrhenius temperature coefficient of denitrification process = 1.16

Temperature Equation

Additionally, based on (Ranjith Shepur, 2019) a new variable, temperature, equation is added and modified

$$\frac{dT(t)}{dt} = -k_T(T(t) - T_{air})$$

since first-order rate coefficient for temperature, k_T, is inversely proportional to water depth, $k_T = c^2 H^{-c*^2}$, where c and c* are the unknown constants that will be determined using parameter estimation.

Refined Mathematical Model

The refined mathematical model that is presented here is the result of the modifications that were done to the model in the sense of the incorporation of three new variables, $T(t)$, $A(t)$ and $N(t)$, the temperature of river, ammonium concentration and nitrate concentration in the river water, and performing a slight modification to the equations such as introducing new parameters. Moreover, the modifications too took place by adding in several equations from the different published literatures according to its suitability.

$$\frac{dT(t)}{dt} = -c_{21}^2 H^{-(c_{22})^2}[T(t) - T_{air}] \quad (1)$$

$$\frac{dM(t)}{dt} = M(t)[c_0 - c_1 M(t) - c_2 P(t)] + \frac{c_3 D(t) M(t)}{D(t) + c_4} + \frac{v_{in} M_0(t)}{v_r} \quad (2)$$

$$\frac{dP(t)}{dt} = P(t)[c_5 - c_6 P(t) - c_7 M(t)] + \frac{c_8 D(t) P(t)}{D(t) + c_9} + \frac{c_{10} v_{in} p U(t)}{v_r} \quad (3)$$

$$\frac{dC(t)}{dt} = -c_{11} C(t)\left(\frac{c_{12}}{c_{13} + C(t)}\right) - c_{14} P(t) C(t) + \frac{c_{15} C(t) D(t)}{C(t) + c_{16}}\left(\frac{c_{12}}{c_{13} + C(t)}\right) + \frac{v_{in} C_0(t)}{v_r} \quad (4)$$

$$\frac{dD(t)}{dt} = c_{17}[D_s - D(t)] - c_{19} D(t) M(t) - c_{18} D(t) P(t) + \frac{v_{in} D_0(t)}{v_r} + c_{20} D(t) C(t)\left(\frac{c_{12}}{c_{13} + C(t)}\right) \quad (5)$$

$$\frac{dA(t)}{dt} = c_{23}\left[-k_n A(t) z_n^{T(t)-20} + y_g k_g B(t) z_g^{T(t)-20}\right] \quad (6)$$

$$\frac{dN(t)}{dt} = c_{23}\left[k_n A(t) z_n^{T(t)-20} - k_n N(t) z_d^{T(t)-20}\right] \quad (7)$$

where

$T(t)$	-	Temperature of River
$M(t)$	-	Concentration of Microbes in River
$P(t)$	-	Concentration of PSB in River
$C(t)$	-	Concentration of COD
$D(t)$	-	Concentration of DO
$A(t)$	-	Concentration of Ammonium

$N(t)$	-	Concentration of Nitrate
$U(t)$	-	Amount of product, mPHO, containing PSB used in the river water
$B(t)$	-	Concentration of BOD
H	-	Depth of Water = 30
T_{air}	-	Temperature of Air = 20
v_{in}	-	Average amount of sewage coming in river = 1
v_r	-	Volume of River = 1
$M_0(t)$	-	Concentration of Microbe in Untreated River Water (Initial of $M(t)$)
p	-	Concentration of PSB in 1L of mPHO, U(t) = 1
$C_0(t)$	-	Concentration of COD in Untreated River Water (Initial of $C(t)$)
D_s	-	Concentration of Saturated Oxygen = $14.652 - 0.41022T + 0.007991T^2 - 0.000077774T^3$
$D_0(t)$	-	Concentration of DO in Untreated River Water (Initial of $D(t)$)
k_n	-	Nitrification Rate 20℃ per day = 0.54
k_g	-	Degradation rate of organic matter at 20℃ per day = 0.5
k_d	-	Denitrification rate per day = 1
z_n	-	Arrhenius temperature coefficient of nitrification process = 1.13
z_g	-	Arrhenius temperature coefficient of degradation process = 1.024
z_d	-	Arrhenius temperature coefficient of denitrification process = 1.16
y_g	-	Nitrogen content in organic matter = 0.29
t	-	Time, $0 \leq t \leq 70$
c_i	-	Constants (calculated by parameter fitting-based experiment data in treated river water) where $i = 0,1,\dots,23$

Conversion of ODE System to Matrix Form

In order to simplify the ODE system, first w_i are introduced where $i = 0,1,2,3,4,5,6$ to replace all the seven variables, $T(t), M(t), C(t), D(t), A(t)$ and $N(t)$.

Therefore, letting $[w] = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} T(t) \\ M(t) \\ P(t) \\ C(t) \\ D(t) \\ A(t) \\ N(t) \end{bmatrix}$

From equation 1, 2, 3, 4, 5, 6 and 7, isolating $\frac{d}{dt}[w_i]$ from RHS to the LHS,

$$\frac{dT(t)}{dt} = -c_{21}^2 H^{-(c_{22})^2}[T(t) - T_{air}] \quad (1a)$$

$$\frac{dM(t)}{dt} = M(t)[c_0 - c_1 M(t) - c_2 P(t)] + \frac{c_3 D(t) M(t)}{D(t) + c_4} + \frac{v_{in} M_0(t)}{v_r} \quad (2a)$$

$$\frac{dP(t)}{dt} = P(t)[c_5 - c_6 P(t) - c_7 M(t)] + \frac{c_8 D(t) P(t)}{D(t) + c_9} + \frac{c_{10} v_{in} p U(t)}{v_r} \quad (3a)$$

$$\frac{dC(t)}{dt} = -c_{11} C(t)\left(\frac{c_{12}}{c_{13} + C(t)}\right) - c_{14} P(t) C(t) + \frac{c_{15} C(t) D(t)}{C(t) + c_{16}}\left(\frac{c_{12}}{c_{13} + C(t)}\right) + \frac{v_{in} C_0(t)}{v_r} \quad (4a)$$

$$\frac{dD(t)}{dt} + \frac{dA(t)}{dt} = c_{17}[D_s - D(t)] - c_{19} D(t) M(t) - c_{18} D(t) P(t) + \frac{v_{in} D_0(t)}{v_r} + c_{20} D(t) C(t)\left(\frac{c_{12}}{c_{13} + C(t)}\right) \quad (5a)$$

$$\frac{dA(t)}{dt} = c_{23}\left[-k_n A(t) z_n^{T(t)-20} + y_g k_g B(t) z_g^{T(t)-20}\right] \quad (6a)$$

$$\frac{dN(t)}{dt} = c_{23}\left[k_n A(t) z_n^{T(t)-20} - k_n N(t) z_d^{T(t)-20}\right] \quad (7a)$$

Representing the equations above in terms of w_i and matrix form,

$$\frac{d}{dt}\begin{bmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{bmatrix}\begin{bmatrix}w_0\\w_1\\w_2\\w_3\\w_4\\w_5\\w_6\end{bmatrix}=\begin{bmatrix}-(c_{21})^2H^{-(c_{22})^2}[w_0-T_{air}]\\w_1[c_0-c_1w_1-c_2w_2]+\frac{c_3w_4w_1}{w_4+c_4}+\frac{v_{in}M_0(t)}{v_r}\\w_2[c_5-c_6w_2-c_7w_1]+\frac{c_8w_4w_2}{w_4+c_9}+\frac{pc_{10}v_{in}U(t)}{v_r}\\-c_{11}w_3\frac{c_{12}}{c_{13}+w_3}-c_{14}w_2w_3-\frac{c_{15}w_4w_3}{w_3+c_{16}}\frac{c_{12}}{c_{13}+w_3}+\frac{v_{in}C_0(t)}{v_r}\\c_{17}[D_s-w_4]-c_{18}w_4w_2-c_{19}w_4w_1-c_{20}w_4w_3\frac{c_{12}}{c_{13}+w_3}+\frac{v_{in}D_0(t)}{v_r}\\c_{23}\left[-k_nw_5z_n^{w_0-20}+y_gk_gB(t)z_g^{w_0-20}\right]\\c_{23}\left[k_nw_5z_n^{w_0-20}-k_nw_6z_d^{w_0-20}\right]\end{bmatrix}$$

Letting $L=\begin{bmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{bmatrix}$ and all the equation on the RHS as f_i where $i=0,1,\dots,6$

$$\frac{d}{dt}L\begin{bmatrix}w_0\\w_1\\w_2\\w_3\\w_4\\w_5\\w_6\end{bmatrix}=\begin{bmatrix}f_0\\f_1\\f_2\\f_3\\f_4\\f_5\\f_6\end{bmatrix}$$

Multiplying L^{-1} both sides since $L^{-1}L=I$,

$$\frac{d}{dt}\begin{bmatrix}w_0\\w_1\\w_2\\w_3\\w_4\\w_5\\w_6\end{bmatrix}=L^{-1}\begin{bmatrix}f_0\\f_1\\f_2\\f_3\\f_4\\f_5\\f_6\end{bmatrix}\qquad(8)$$

Obtaining the inverse of *L*, L^{-1}, using Elementary Row Operation method,
starting with $L^{-1}=(I|L)$ reducing until $L^{-1}=(L|I)$ in which the *L* will become L^{-1}

$$L^{-1}=\left(\begin{array}{ccccccc|ccccccc}1&0&0&0&0&0&0&1&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&1&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&1&0&0&0&0\\0&0&0&1&0&0&0&0&0&0&1&0&0&0\\0&0&0&0&1&0&0&0&0&0&0&1&1&0\\0&0&0&0&0&1&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&1&0&0&0&0&0&0&1\end{array}\right)$$

$$\xrightarrow[-R_6+R_5]{}\left(\begin{array}{ccccccc|ccccccc}1&0&0&0&0&0&0&1&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&1&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&1&0&0&0&0\\0&0&0&1&0&0&0&0&0&0&1&0&0&0\\0&0&0&0&1&-1&0&0&0&0&0&1&1&0\\0&0&0&0&0&1&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&1&0&0&0&0&0&0&1\end{array}\right)$$

Therefore, $L^{-1}=\begin{pmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&-1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{pmatrix}$

Substituting L^{-1} into 8,

$$\frac{d}{dt}\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{bmatrix}$$

Solving it,

$$\frac{d}{dt}\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 - f_5 \\ f_5 \\ f_6 \end{bmatrix} \qquad (9)$$

Here, finally equation 9 is the ODE system that were converted to matrix system which will then be used to in parameter estimation, minimization.

Digitalization of Published Data

To ensure that the estimated parameters are in sync with real life situation, it is crucial to incorporate real life data instead of pseudo data. Thus, parameter fitting was done to the graphs found in several scientific articles in order to obtain the experiment data. The steps below explain the whole process of parameter fitting.

step 1: Save a screenshot of the graph found from any of the article online in jpeg format

step 2: Open the picture using paint application. Obtain the pixel point of the starting and end point of the x and y-axis of the graph. Section the graphs into suitable section if too much noise exists.

For example, in the graph below the pixel coordinates for 18mg/l is 113, 38px, 0 hours and 0 mg/l are 113, 344px and finally for 20 000 hours is 693, 344px. Moreover, since the graph has quite a number of noises, it was sectioned into 4 sections which are indicated according to the green, blue and pink line where the first section is the points before the green line, second section is the points that lies in between the green and blue lines, third section is the points that lies in between the blue and pink lines and the fourth section is the points that lies on the RHS of the pink line.

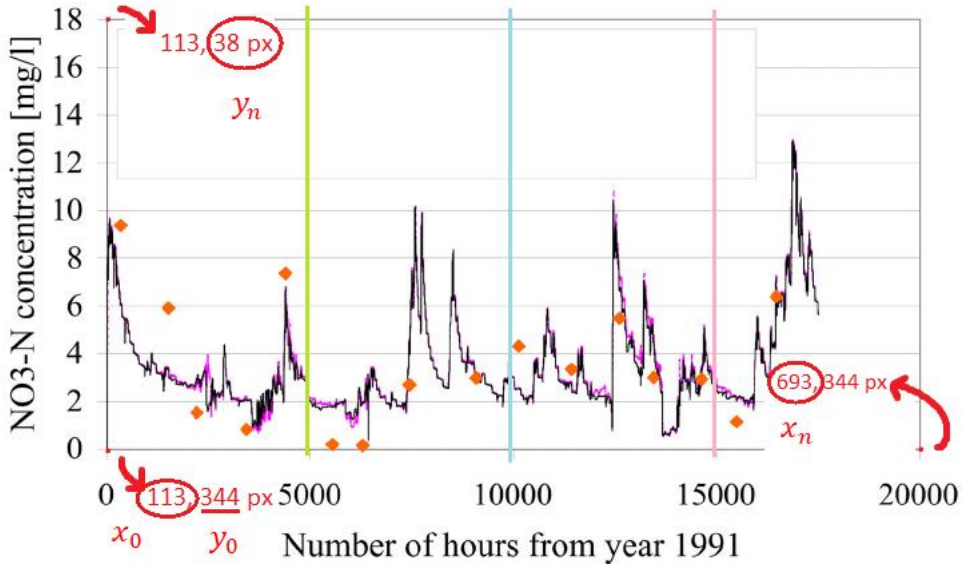


Figure 1 Concentration of Nitrogen, N(t), found from Published Literature

step 3: Perform two matrix multiplication for x-axis and y-axis to obtain the equation that converts the pixel coordinates to cartesian coordinates where,

$x_0 = 113$ pixel, $y_0 = 344$ pixel, $x_n = 693$ pixel, $y_n = 38$ pixel,

For y-coordinate:

$$\begin{pmatrix} 1 & 38 \\ 1 & 344 \end{pmatrix}\begin{pmatrix} y_n \\ y_0 \end{pmatrix} = \begin{pmatrix} 18 \\ 0 \end{pmatrix}$$

For x-coordinate:

$$\begin{pmatrix} 113 & 1 \\ 693 & 1 \end{pmatrix}\begin{pmatrix} x_0 \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 20000 \end{pmatrix}$$

Solving for $\begin{pmatrix} y_n \\ y_0 \end{pmatrix}$ and $\begin{pmatrix} x_0 \\ x_n \end{pmatrix}$ would result in obtaining two equations

equation 1: $y_{coordinate} = y_n - y_0\left(y_{pixel\ coordinate}\right)$

equation 2: $x_{coordinate} = x_0\left(x_{pixel\ coordinate}\right) - x_n$

step 4: Obtain needed pixel coordinates to recreate a similar line graph with the article obtained online and finally convert the pixel coordinates to cartesian coordinates using the 2 equations obtained in the end of step 3.

Using Microsoft Excel, all of the data are collected from substituting the $x_{pixel\ coordinate}$ and $y_{pixel\ coordinate}$ into equation 1 and 2. Then the data is plotted into scatter plot graph where the line of best fit which is known as the trendline in Microsoft Excel is added to obtain an equation that could show the pattern of the data. Here, for the graphs that were sectioned, several equations could be obtained which would finally will build a piecewise function.

Positive Values Interpolation Technique

While performing the parameter fitting, some of the data plotted were observed to produce trendline equation that are partly negative. One of the examples of a similar situation is shown in the image below where it could be observed that part of the trendline provides a negative y value.

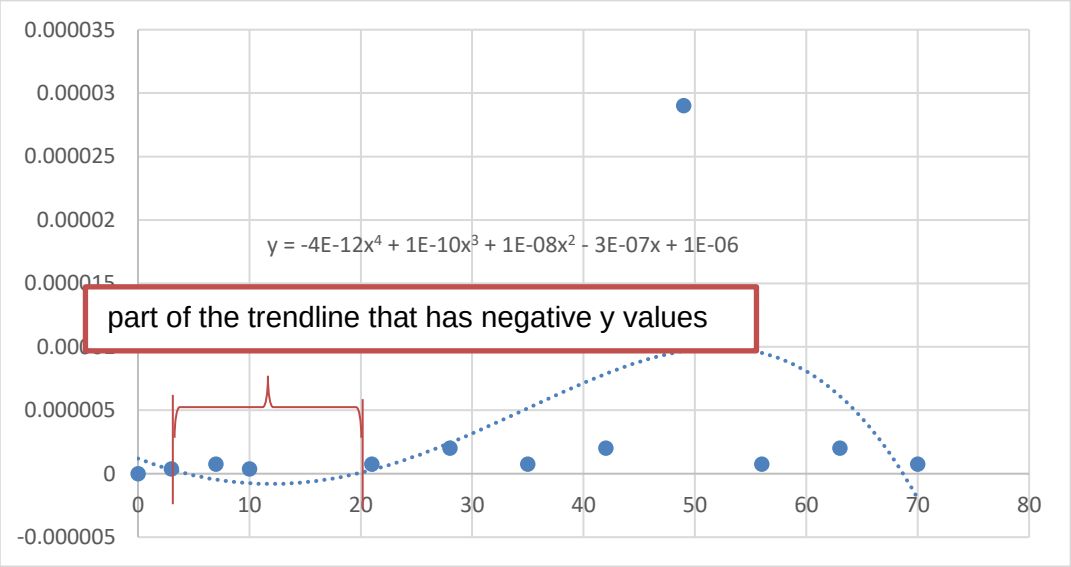


Figure 2 Graph of Concentration of PSB, P(t), the Actual Data

In order to prevent from obtaining a negative value, mathematical optimization was performed using Python's code and SciPy library in which the mathematical theory is basically the fitting of experiment values found through digitalization of published data.

Minimize

$$obj(a_0, a_1, a_2, a_3) = \sum_{i=1}^{j} \left|(a_0 + a_1 x_g(i) + a_2 x_g(i)^2 + a_3 x_g(i)^3) - y_g(i)\right|^2$$

where

$a_0, a_1, a_2\ and\ a_3$ – unknown constants that is fitted

$x_g(i)$ – experiment values of x

$y_g(i)$ – experiment values of y

The code specifically fits a polynomial curve to a set of data points found in (Graeme Wake, et al., 2014) using least square method subject to a constraint on the coefficients of the polynomial. The algorithm is as stated as below

step 1: Import NumPy and SciPy libraries.
step 2: Define two arrays, x_g and y_g that consists of ten x and y data points respectively.
step 3: Define a fourth-degree polynomial function 'poly', $a_0 + a_1x + a_2x^2 + a_3x^3$.
step 4: Define the objective function, 'obj', that returns the squared error between data points and polynomial curve, 'poly', and input coefficients, 'coeff'.
step 5: Define a function, 'ineqf', that returns coefficient of polynomial.
step 6: Inequality constraint, 'ineq-const', is defined on 'ineqf' function to impose a constraint on coefficient of polynomial whose $ineqf(a) \geq 0$.
step 7: Optimization is performed using minimize function. 'Obj' minimized with subject to 'ineq-const' using SLSQP method.

Finally, a function that displays the graph as shown below is obtained.

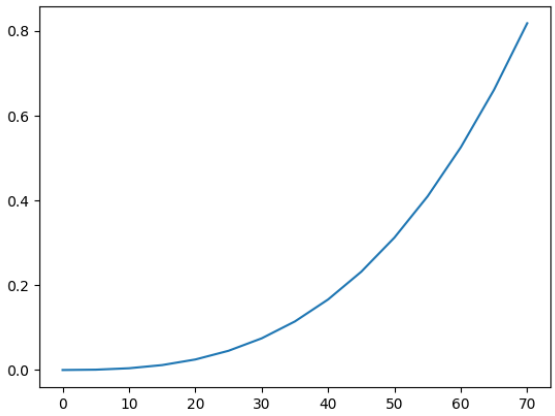


Figure 3 Positive Value Graph of Concentration of PSB, P(t)

Parameter Estimation

The unknown constants presented in the mathematical model, c_i, known as the parameters of the model in the finalized modified mathematical model are obtained using parameter estimation via Python coding. The optimal values of the parameters are calculated using a parameter fitting approach, derivative-free optimization algorithm, after acquiring the tested data values from published literatures, (John R. Ockendon, 2014) and (Mona Radwan, 2001). Following this, an objective function that calculates the error between the actual and simulated data values is introduced and minimized with subject to the mathematical model in 3.6 using python's integrated Sequential Least Square Programming (SLSQP) method of solving the minimization introduced providing the least error value along with the optimal values of the parameters. The objective function is as shown below in which $T(t), M(t), P(t), D(t)\ and\ C(t)$ is the simulated values whereas $T^*(t), M^*(t), P^*(t), D^*(t)\ and\ C^*(t)$ are the actual values obtained from the experimented data values in the published literatures.

$$f(c_0, c_1, \dots, c_{22}) = \sum_{i=1}^{j} |T(t_i) - T^*(t_i)|^2 + \sum_{i=1}^{j} |M(t_i) - M^*(t_i)|^2 + \sum_{i=1}^{j} |P(t_i) - P^*(t_i)|^2 + \sum_{i=1}^{j} |D(t_i) - D^*(t_i)|^2 + \sum_{i=1}^{j} |C(t_i) - C^*(t_i)|^2$$

where
$T^*(t_i)$ – temperature of river water measured at point of effluent
$M^*(t_i)$ – concentration of microbes at point of effluent
$P^*(t_i)$ – concentration of PSB at point of effluent
$D^*(t_i)$ – concentration of DO at point of effluent
$C^*(t_i)$ – concentration of COD at point of effluent

Simulation of Data Values

Based on the optimal values of the parameters that are estimated based on 3.10., simulated data values for all the seven variables are computed using the python coding by solving the mathematical model which is a system of differential equations using LSODA from FOTRAN library odepack. LSODA

is one of the solving methods which automatically switched from the nonstiff method, Adams Method, to either remaining to the nonstiff method or to stiff method, Backward Difference Formula (BDF). (Millman, n.d.) Stiffness here refers to the difficulty of solving an equation numerically. Since, the model or the system of differential equations presented in 3.6 could be concluded to be stiff as it seems too complex to solve, the method of solving that is used in order to obtain the simulated data values is the Backward Difference Formula (BDF). BDF basically is a method that is often used to solve the ODE problems that are stiff in which the solution changes rapidly. BDF method approximates the derivative of a function by comparing its current value with a previous value and the formula is as shown below. (Numerical Differentiation)

$$\frac{df}{dt} = \frac{f(t) - f(t - \Delta t)}{\Delta t}$$

MAPE Values Calculation and Interpretations

Finally, after estimating the parameters and obtaining the simulated data values, an error analysis was performed. In this research, the calculation of Mean Absolute Percentage Error, MAPE, are used to evaluate the accuracy of the simulated data values for each variable, *T(t), P(t), M(t), C(t), D(t), A(t), N(t)*. The formula for the MAPE value calculation of one of the variables for an instance A(t) is as stated below.

$$MAPE = \frac{100\%}{n} \sum_{1}^{n} \left| \frac{A^*(t) - A(t)}{A^*(t)} \right|$$

where *n* is total number of data, $A^*(t)$ is Actual data and $A(t)$ is Simulated data

Based on the MAPE values calculated, a conclusion on the accuracy of the simulated data for each variable were concluded based on (Montaño, 2013) which is presented in table 1.

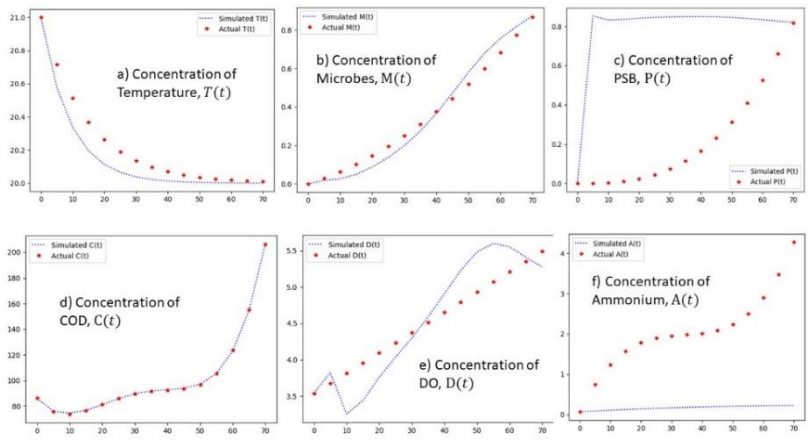
Table 1: MAPE Values' Conclusion

MAPE Value	Conclusion
Less than 10%	Very Accurate Forecast
10 – 20%	Good Forecast
20 – 50%	Fair Forecast
More than 50%	Inaccurate Forecast

Results and discussion

Using the experiment data values, the simulated data values is obtained through the substitution of the parameters estimated numerically. An error analysis is then performed with the experiment values and the simulated values by computing the percent error and finally obtaining the Mean Absolute Percentage Error (MAPE). In figure 4 the plotted graph of the comparison between the actual value with the simulated value and the percent error for all the seven variables in table 2 are presented and finally the estimated parameters are presented in table 3.

Based on interpretation of typical MAPE values presented in (Montaño, 2013) and table 2, the simulated data for the variable *T(t)*, *C(t)* and *D(t)* are highly accurate forecasting as the MAPE value is less than 10%, 0.37%, 0.34% and 6.38% respectively. The simulated values that are obtained for *N(t)* and *M(t)* is considered as a good forecasting as the MAPE value is 10.64% and 20.03% respectively which is in the interval of 10 to 20%. The simulated values for *P(t)* and *A(t)* are however seemed to have inaccurate forecasting as the MAPE values obtained is more than 50% which is 9663.67.83% and 85.40% respectively. Based on this error analysis it could be clearly be observed that *T(t)* has the best forecasting and *P(t)* has the worst forecasting compared to the simulated values obtained.



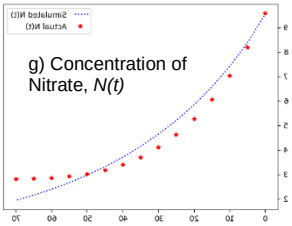


Figure 4 Comparison graphs of actual data and simulated data versus time for (a) Temperature of River, *T(t)*, (b) Concentration of Microbes, *M(t)*, (c) Concentration of PSB, *P(t)*, (d) Concentration of COD, *C(t)*, (e) Concentration of DO, *D(t)*, (f) Concentration of Ammonium, *A(t)* and (g) Concentration of Nitrate, *N(t)*

Table 2: MAPE Values for 7 variables.

Variables	MAPE Values (%)
T(t)	0.37
M(t)	20.03
P(t)	9663.67
C(t)	0.34
D(t)	6.38
A(t)	85.40
N(t)	10.64

Table 3: Estimated values of the parameters

Parameter	Value	Parameter	Value	Parameter	Value
c[0]	4.77778933	c[8]	1.4816844	c[16]	0.86096705
c[1]	0.44830202	c[9]	5.49947516	c[17]	-0.47533362
c[2]	2.95335137	c[10]	2.16016567	c[18]	2.81843168
c[3]	-1.86977327	c[11]	11.4275905	c[19]	0.2281656
c[4]	-1.9214677	c[12]	0.29627809	c[20]	-5.74989628
c[5]	-2.82771011	c[13]	-11.41186572	c[21]	0.81615635
c[6]	0.37457956	c[14]	1.12213811	c[22]	0.72953154
c[7]	0.25855492	c[15]	9.16724767	c[23]	0.04354689

Conclusion

The primary goals of this research are to modify mathematical models derived from various published literatures, estimate model parameters, and obtain simulated data values for all variables considered, which in this case would be seven variables. Following that, by the end of the research, all three objectives had been met. In terms of obtaining the optimum value of the model's parameter, parameter estimation was successful, and the least value of the objective function could also be interpreted. In terms of obtaining the optimum value of the model's parameter, parameter estimation was successfully performed and the least value of objective function also could be interpreted the error of the model were approximately 79 which were obtained after numerous modifications and trials since the starting value of the objective function were as large as 1×10^{14}. In this research, it could be concluded that three variables, *T(t)*, *C(t)* and *D(t)*, were successfully get a highly accurate forecasting and two variables with somewhat good forecasting, *M(t)* and *N(t)*. Despite many successes, two variables, *A(t)* and *P(t)*, particularly *P(t)*, have failed to achieve an accurate predicting.

References

Ambulkar, J. A. (2022, September 5). *wastewater treatment*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/technology/wastewater-treatment

Amir Hamzah, A. B. (2015, 3). A mathematical model for wastewater treatment. *Jurnal Teknologi, 78*, 65-70. doi:10.11113/jt.v78.7815

Amit Sonune, R. G. (2004). Developments in wastewater treatment methods. *Desalination, 167*, 55-63.

Bloganica. (2017, June 27). *PRIMARY, SECONDARY, AND TERTIARY WASTEWATER TREATMENT: HOW DO THEY WORK?* Retrieved from ORGANICA: https://www.organicawater.com/primary-secondary-tertiary-wastewater-treatment-work/

Fauzi, S. H. (2021). Wastewater Treatment Process: A Modified Mathematical Model for Oxidation Ponds. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 86*(1), 76-86.

Graeme Wake, John R. Ockendon, Kok Lay Teo, Ryan Loxtan, Adérito Araújo, Basuki Widodo, . . . J-BIOTECH ENVIRONMENT SDN. BHD. (2014). Mathematical Modeling and Optimizationfor Biological-based Treatment of Taman Timor Oxidation Pond, Johor. *Malaysian 2nd Mathematics in Industry Study Group*, (pp. 1-37).

John R. Ockendon, G. W. (2014). Mathematical Modeling and Optimization for Biological-based Treatment of Taman Timor Oxidation Pond, Johor. *Malaysian 2nd Mathematics in Industry Study Group (MISG 2014)*, 1-37.

Millman, J. (n.d.). *scipy*. Retrieved from github: https://github.com/scipy/scipy/blob/v0.19.0/scipy/integrate/odepack/readme

Mona Radwan, A. E.-S. (2001, 12). Modeling of Nitrogen in River Water Using a Detailed and a Simplified Model. *The Scientific World Journal, 1 Suppl 2*, 200-6. doi:10.1100/tsw.2001.351

Montaño, J. a. (2013, 11). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema, 25*, 500-506. doi:10.7334/psicothema2013.23

Municipal Solid Waste Landfills. (2022, April 25). Retrieved from EPA United States Environmental Protection Agency: https://www.epa.gov/landfills/municipal-solid-waste-landfills

Natalie Croker, R. R. (2022, August 20). *CNN World.* Retrieved from The world's rivers are drying up from extreme weather. See how 6 look from space: https://edition.cnn.com/2022/08/20/world/rivers-lakes-drying-up-drought-climate-cmd-intl/index.html

Numerical Differentiation. (n.d.). Retrieved July 12, 2023, from Lehigh University: https://www.lehigh.edu/~ineng2/clipper/notes/NumDif.htm

Ranjith Shepur, A. S. (2019, 01). Water Quality Model for Streams: A Review. *Journal of Environmental Protection, 10*, 1612-1648. doi:10.4236/jep.2019.1012097

World Health Organization. (2022, March 21). Retrieved from Drinking-water: https://www.who.int/news-room/fact-sheets/detail/drinking-water