Investigating the Growth of Bacteria using Double Sigmoid Model with Reparameterization

Masithoh Yessi Rochayani 1,*, Dahlia Gladiola Rurina Menufandu 2 and Rahmila Dapa 1

1 Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, 50275 Jawa Tengah, Indonesia; rahmiladapa@gmail.com (R.D.)
2 Department of Statistics, Faculty of Natural Science and Mathematics, Universitas Brawijaya, 65145 Jawa Timur, Indonesia; menufandu.dahlia@gmail.com (D.G.R.M.)
* Correspondence: yessirochayani@lecturer.undip.ac.id (M.Y.R.)


Received: 18 September 2023 Accepted: 5 December 2023 Published: 31 December 2023

Abstract: The growth of an organism can be modeled using a growth curve. However, bacteria's growth pattern differs from other organisms. Bacterial growth is divided into four phases: lag, logarithmic, stationary, and death. The experts re-parameterized the growth curve to match the growth phase of the bacteria. Bacterial growth patterns generally do not show a single sigmoid pattern but form two curves. Therefore, the double sigmoid model is more suitable. This study modeled the growth of the *Pseudomonas putida* bacteria by observing the optical density of the medium. Model parameters are estimated using the Non-Linear Least Square (NLS) method with the Gauss-Newton algorithm. The modeling results show that the double sigmoid model fits the growth curve of *Pseudomonas putida* better than the single sigmoid model. The Double Logistic model outperforms all models with the highest adjusted R² and the smallest RMSE, AIC, and BIC values.

Keywords: Growth curve, Bacterial growth, *Pseudomonas putida*, Double Sigmoid Model.

Copyright: © 2022-2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Growth is a process of increasing size, volume, mass, height, or other sizes, which can be expressed quantitatively and is irreversible (cannot change back to a previous condition or state). A growth curve is a graph that describes how a phenomenon grows over time. Growth curves are often used to describe poultry weight or plant height. In addition, the growth curve is also used to describe population growth or the development of disease outbreaks over time. In the 19th to 20th centuries, many scientists developed growth models including the Gompertz model developed by Gompertz in 1825, Logistics developed by Verhulst in 1838, Morgan-Mercer-Flodin (MMF) in 1910-1913, Brody in 1945, Weibull in 1951, Von Bertalanffy 1957, Schnute 1981, Stannard 1985, and MCDill-Amateis 1992 (Panik, 2014). In various studies, more researchers use these growth models to observe the growth of animals than plants (Tjørve & Tjørve, 2017). The research used the Gompertz and Logistics models to observe duck growth patterns (Prayogo et al., 2017). Another study described the growth curve of the Kacang goat's body weight using Gompertz and logistic models.
Another research used Brody, Von Bertalanffy, Logistics, and Gompertz growth curves to describe Brazilian tropical goats' body weight (de Sousa et al., 2021).

Besides observing animal growth patterns, growth models are also used to observe microorganism growth patterns. In prokaryotic organisms such as bacteria, growth is an increase in cell volume and the number of cells. The growth of bacterial cells usually follows a certain growth pattern in the form of a sigmoid growth curve. Zwietering et al. (1990) modeled the growth of the bacterium Lactobacillus plantarum using a modified Logistic, Gompertz, Richard, Stannard, and Schnute models. The results of his research show that the modification of the Gompertz model gives the best results. Pla et al. (2015) observed the growth of Bacillus cereus, Listeria monocytogenes, and Escherichia coli using a three-phase linear model, Gompertz, Logistic, Richard, and Baranyi. The results showed that all models showed good results when observing bacterial growth using optical density (Pla et al., 2015).

The growth pattern of bacteria is different from that of animals or plants. Bacterial growth is divided into four phases (Madigan et al., 2021). The first phase is the lag phase, where bacteria adapt to the media to slow bacterial growth. Then, the bacteria go through a logarithmic phase. In the logarithmic phase, the bacteria divide continuously until they reach their maximum amount. The stationary phase occurs after the logarithmic phase, which is the right phase for harvesting. In this phase, the number of bacteria that die equals the number of bacteria that divide. So, there is no increase in the number of bacteria. The final stage is the death phase, in which the bacteria die but no longer any bacteria are still dividing. So, in this phase, the number of bacteria decreases. Experts reparameterized growth models to fit the phenomenon of the bacterial growth phase. Zwietering et al. (1990) reparameterized the Logistics model, Gompertz, Richard, Stanard, and Schnute. Vázquez et al. (2012) reparameterized the Weibull and von Bertalanffy model. The modifications of these models have been used by Longhi et al. (2017), Njalam'mano & Chirwa (2019), and Pla et al. (2015). The original Logistic and Gompertz models are shown in Equations (1) and Equation (2).

\[
y_t = \frac{a}{1 + \exp(-b(t-h))} \\
y_t = a \exp\left[\exp\left(-b(t-h)\right)\right]
\]

where \(y_t\) is the optical density at time \(t\), \(t\) is incubation time (hours), \(a\) is the maximum optical density, \(h\) is the time at which the absolute growth rate is maximum (hours), and \(b\) is the relative growth rate that is determined at the time (per hour). Equations (3) and Equation (4) state the Logistic and Gompertz model modifications.

\[
y_t = \frac{A}{1 + \exp\left(\frac{4\mu}{A}(\lambda - t) + 2\right)} \\
y_t = A \exp\left[-\exp\left(\frac{\mu e}{A}(\lambda - t) + 1\right)\right]
\]

where \(y_t\) is the optical density at time \(t\), \(t\) is the incubation time (hours), \(A\) is the maximum optical density, \(\mu\) is the maximum growth rate (per hour); \(\lambda\) is lag time (hours) (Njalam'mano & Chirwa, 2019).

The modified parameters of the Logistic growth model are explained as follows. The maximum optical density \((A)\) is the same as \(a\) in Equation (1) and (2), the maximum growth rate \((\mu)\) is derived by Equation (5), and lag time \((\lambda)\) is derived by Equation (6).

\[
\mu = \frac{bc}{4} \\
\lambda = h - \frac{2}{b}
\]

Where \(h\) is the time at which the absolute growth rate is maximum, \(b\) is the relative growth rate, \(c\) is the asymptotic amount of growth, and \(e = 2.718\) (Njalam'mano & Chirwa, 2019). Meanwhile, the parameter of the Gompertz model is modified as follows. The \(A\) is the same as \(a\), while the \(\mu\) and \(\lambda\) are derived by Equation (7) and Equation (8) respectively.

\[
\mu = \frac{bc}{e} \\
\lambda = h - \frac{1}{b}
\]
Vázquez et al. (2012) stated that in some cases, animal growth patterns do not show a single sigmoid pattern but tend to form two sigmoid curves (double sigmoid). This double sigmoid pattern can be expressed as the sum of the two sigmoid curves. The Double Logistics and Double Gompertz models can be seen in Equation (9) and Equation (10), respectively.

\[
y_t = \frac{A_1}{1 + \exp\left(\frac{4\mu_1}{A_1}(\lambda_1 - t) + 2\right)} + \frac{(A_2 - A_1)}{1 + \exp\left(\frac{4\mu_2}{(A_2 - A_1)}(\lambda_2 - t) + 2\right)}
\]

(9)

\[
y_t = A_1 \exp\left(-\exp\left[\frac{\mu_1 e}{A_1}(\lambda_1 - t) + 1\right]\right) + (A_2 - A_1) \exp\left(-\exp\left[\frac{\mu_2 e}{(A_2 - A_1)}(\lambda_2 - t) + 1\right]\right)
\]

(10)

where \( y_t \) is the optical density at time \( t \), \( t \) is the incubation time (hours), \( A_1 \) is the maximum optical density in the first sigmoid, \( \lambda_1 \) is the lag phase in the first sigmoid (hours), \( \mu_1 \) is the maximum growth rate in the first sigmoid (per hour), \( A_2 \) is the maximum optical density, \( \lambda_2 \) is the lag phase in the second sigmoid (hours), \( \mu_2 \) is the maximum growth rate in the second sigmoid (per hour). This study aims to model the growth pattern of \( Pseudomonas putida \) and compare the single sigmoid curve (Logistic and Gompertz) and the double sigmoid curves (Double Logistic and Double Gompertz). The selection of the best model is done based on four evaluation metrics, including \( R^2_{adj} \), RMSE, AIC, and BIC.

2. Materials and Methods

The growth model is one of the intrinsically non-linear regression models, a non-linear model that cannot be transformed into a linear form. The parameters of this model are obtained by the non-linear least square (NLS) method. The idea of this method is the same as in ordinary least squares (OLS), which is to minimize the sum of the residual squares. However, because the function used is non-linear, an iterative method is needed to estimate the parameters. Assume that given the model \( M \) with parameter \( \theta = (\theta_1, ..., \theta_m) \) where \( m \) stated the number of parameters. The estimated value of the model at time-\( i \) or \( t_i \) is expressed by \( M_i(\theta) \). The results of measurements at time \( t_i \) are expressed by Equation (11).

\[
y_i = M_i(\theta) + \eta_i(\theta)
\]

(11)

where \( i = 1, 2, ..., n \) and \( r_i(\theta) \) is the residual at time \( t_i \). The function of the residual can be expressed by Equation (12).

\[
r_i(\theta) = y_i - M_i(\theta)
\]

(12)

The Residual Sum of Square (RSS) is expressed by Equation (13).

\[
f(\theta) = \frac{1}{2} \sum_{i=1}^{n} (r_i(\theta))^2 = \frac{1}{2} \|r(\theta)\|^2 = \frac{1}{2} r(\theta)^T r(\theta)
\]

(13)

where and \( r(\theta) = (r_1(\theta), ..., r_n(\theta))^T \).

Minimizing the RSS means the model parameters close to the observed data are obtained. The gradient (a vector of first-order partial derivatives) of RSS is stated by Equation (14).

\[
G(\theta) = \nabla r(\theta) r(\theta) = J(\theta)^T r(\theta)
\]

(14)

where \( J(\theta) \) is the Jacobian matrix of the residual function \( r(\theta) \) and is stated by Equation (15).

\[
J(\theta) = \nabla r(\theta)^T = \\
\begin{bmatrix}
\frac{\partial r_1(\theta)}{\partial \theta_1} & ... & \frac{\partial r_1(\theta)}{\partial \theta_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial r_m(\theta)}{\partial \theta_1} & ... & \frac{\partial r_m(\theta)}{\partial \theta_n}
\end{bmatrix}
\]

(15)

Therefore, the Gradient \( G(\theta) \) in Equation (14) can be expressed as Equation (16). Meanwhile, the Hessian \( H(\theta) \) matrix (a matrix of second-order partial derivatives of a function) is stated by Equation (17) (Siregar et al., 2018).

\[
G(\theta) = \begin{bmatrix}
\frac{\partial r_1(\theta)}{\partial \theta_1} & ... & \frac{\partial r_m(\theta)}{\partial \theta_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial r_1(\theta)}{\partial \theta_n} & ... & \frac{\partial r_m(\theta)}{\partial \theta_n}
\end{bmatrix}
\begin{bmatrix}
r_1(\theta) \\
\vdots \\
r_m(\theta)
\end{bmatrix} = \\
\begin{bmatrix}
\sum_{i=1}^{n} \frac{\partial r_i(\theta)}{\partial \theta_1} r_i(\theta) \\
\vdots \\
\sum_{i=1}^{n} \frac{\partial r_i(\theta)}{\partial \theta_n} r_i(\theta)
\end{bmatrix}
\]

(16)
\[ H(\theta) = \sum_{i=1}^{n} \nabla r_i(\theta)\nabla r_i(\theta)^T + \sum_{i=1}^{n} r_i(\theta)\nabla^2 r_i(\theta) \]
\[ = \nabla(\theta)^T J(\theta) + \sum_{i=1}^{n} r_i(\theta)\nabla^2 r_i(\theta) \]  

(17)

By expressing the objective function \( f(\theta) \) as the second-order Taylor series expansion, the quadratic model is obtained as Equation (18).

\[ q(\theta) = \frac{1}{2} r(\theta)^T r(\theta) + J(\theta)^T r(\theta) (\theta - \theta_k) + \frac{1}{2} (S(\theta_k) + J(\theta_k)^T J(\theta_k))(\theta - \theta_k)^2 \]  

(18)

where \( S(\theta_k) = r(\theta)^T r(\theta) \) (Lai et al., 2017).

Deriving Equation (18) with respect to \( \theta \) and equating to zero, then Equation (19) is obtained.

\[ J(\theta_k)^T r(\theta_k) + (S(\theta_k) + J(\theta_k)^T J(\theta_k))(\theta - \theta_k) = 0 \]  

(19)

Rearrange (19), we obtain Equation (20).

\[ (S(\theta_k) + J(\theta_k)^T J(\theta_k))(\theta - \theta_k) = -J(\theta_k)^T r(\theta_k) \]  

(20)

After some algebra manipulations, by letting \( \theta = \theta_{k+1} \), then Equation (20) can be rewritten by Equation (21).

\[ \theta_{k+1} = \theta_k - (S(\theta_k) + J(\theta_k)^T J(\theta_k))^{-1} J(\theta_k)^T r(\theta_k) \]  

(21)

Since the term \( S(x) \) is very small, then it can be ignored (Chong & Zak, 2013). Therefore, the Gauss-Newton recursion equation is obtained in Equation (22).

\[ \theta_{k+1} = \theta_k + [J(\theta_k)^T J(\theta_k)]^{-1} [-J(\theta_k)^T r(\theta_k)] \]  

(22)

From the description above, the Gauss-Newton algorithm can be briefly described as follows.

1. Input the initial value of parameter \( \theta_0 \) , set the tolerance \( \varepsilon > 0 \) and let the iteration index \( k = 0 \).
2. Calculate the gradient \( G(\theta_k) \) using Equation (16), then compute the norm of the gradient \( \|G(\theta_k)\| \). If \( \|G(\theta_k)\| < \varepsilon \), stop.
3. Calculate \( [J(\theta_k)^T J(\theta_k)]^{-1} [-J(\theta_k)^T r(\theta_k)] \) and update the parameters using Equation (22) and repeat Step 2 until convergence.

### 3. Results and Discussion

#### 3.1. Modeling Pseudomonas putida Growth

The data used in this study were obtained from research by Puteri (2015). The growth of *Pseudomonas putida* bacteria was observed by measuring the optical density of the growth medium at a wavelength of 450 nm. The optical density of a medium is the logarithmic ratio of the intensity of incident light \( I_0 \) to the intensity of the light transmitted through the medium \( I_0 \) (Zhang & Hoshino, 2019). In the first 24 hours, the optical density was measured every hour, while in the second 24 hours, the optical density was measured every 2 hours, resulting in 36 observations. In the stationary phase, the number of dividing bacteria is almost equal to the number of dead bacteria. The start of the stationary phase is reached at 34 hours of incubation time. Therefore, the harvesting of bacteria can be done after an incubation time of 34 hours (Puteri, 2015).

<table>
<thead>
<tr>
<th>Table 1. Parameter estimates of the Logistic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Logistic</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Gompertz</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The results of the parameter estimate of the Logistic and Gompertz single model are summarized in Table 1. At the same time, the parameter estimates of Double Logistic and Double Gompertz are presented in Table 2. All the parameters are generally significant, with a p-value of zero.
Table 2. Parameter estimates of the Double Logistic model

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Logistic</td>
<td>$A_1$</td>
<td>0.3271</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>0.0141</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\lambda_1$</td>
<td>5.7417</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$A_2$</td>
<td>0.6326</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>0.1832</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\lambda_2$</td>
<td>19.3262</td>
<td>0.000</td>
</tr>
<tr>
<td>Double Gompertz</td>
<td>$A_1$</td>
<td>0.2936</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>0.0107</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\lambda_1$</td>
<td>2.7300</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$A_2$</td>
<td>0.6447</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>0.1497</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\lambda_2$</td>
<td>18.8938</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Figure 1 presents the plot of the data. The growth of Pseudomonas putida is characterized by an increase in the density of the medium as the incubation time increases. The growth curve describes the phases in the life cycle of Pseudomonas putida, including the lag, logarithmic, and stationary phases.

Figure 1 shows that the lag phase occurs at an incubation time of 0 to 3 hours. A significant increase in population characterizes the logarithmic phase. The logarithmic phase occurs at an incubation time of 3 to 34 hours. Figure 2 shows the single sigmoid model of Logistics and Gompertz. On the basis of Figure 2, neither model can perfectly fit Pseudomonas putida's growth. Look again at the growth pattern of Pseudomonas putida in Figure 1. Figure 1 shows that the growth of the Pseudomonas putida bacteria forms a double sigmoid curve. For this reason, a more appropriate approach for this growth pattern is a double sigmoid curve. The observational data form the first sigmoid curve at 0 to 14 hours of incubation and the second sigmoid curve at 14 to 48 hours. Therefore, the growth pattern of Pseudomonas putida can be approximated by Double Logistic and Double Gompertz growth curves. Figure 3 shows the Double Logistics and Double Gompertz models. For the double sigmoid model, the parameters interpreted are: $\lambda_3$ states the lag phase length, $A_2$ states the maximum optical density and $\mu_2$ states the maximum growth rate. The lag phase is estimated to occur at 0 to 5.7 hours by the Double Logistic model and 0 to 2.7 hours by the Double Gompertz model. At the beginning of incubation (lag phase), the density has not increased significantly because Pseudomonas
*Pseudomonas putida* has just adapted to the new medium so that the cells have not reproduced (Madigan *et al.*, 2021). The Double Logistics and Double Gompertz models estimate the maximum optical density of *Pseudomonas putida* to be 0.6326 and 0.6447, respectively. The maximum growth rates estimated by the Logistics and Gompertz models are 0.1832 and 0.1497, respectively.

3.2. Model Evaluation

The models were evaluated using four evaluation metrics, including the adjusted coefficient of determination ($R_{adj}^2$), root mean square error (RMSE), Akaike's information criterion (AIC), and Bayesian information criterion (BIC). $R_{adj}^2$ is a correction for the coefficient of determination ($R^2$) where the calculation of $R_{adj}^2$ involves the number of parameters. $R_{adj}^2$ is calculated using Equation (23).
\[ R_{adj}^2 = 1 - (1 - R^2) \left( \frac{n - 1}{n - m - 1} \right) \]  
(23)

where

\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \]  
(24)

\( n \) is the number of observations, \( m \) is the number of the parameters, \( \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \) is the residual sum of squared (RSS), \( \sum_{i=1}^{n} (y_i - \bar{y})^2 \) is the total sum of squared (TSS), \( y_i \) is the observed values at the \( i \)-th time, \( \hat{y}_i \) is the predicted values at the \( i \)-th time, and \( \bar{y} \) is the mean of the predicted values at the \( i \)-th time.

The range of values for the \( R_{adj}^2 \) is between 0 and 1, and the closer \( R_{adj}^2 \) is to 1, the more accurate the model is (Anderson, 2014). RMSE is calculated using Equation (25).

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \]  
(25)

where \( \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \) is RSS. The best model is the one with the lowest RMSE. The formula of AIC is presented by Equation (26).

\[ AIC = n \log \left( \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \right) + 2m \]  
(26)

where \( m \) is the number of parameters in the model (Burnham & Anderson, 2004). The AIC ranges of \((-\infty, +\infty)\). A smaller AIC value indicates a better model. The BIC is calculated by Equation (27),

\[ BIC = n \log \left( \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \right) + m \log n \]  
(27)

where \( n \) is the number of observations, and \( m \) is the number of parameters in the model (Priestley, 1981). A model with a lower BIC provides a better fit.

Table 3 presents the model evaluation. Based on Table 3, the double sigmoid models provide a higher \( R_{adj}^2 \) compared to the single sigmoid models. It means that the double sigmoid models better fit Pseudomonas putida’s growth curve. The Double Logistic model gives a \( R_{adj}^2 \) of 0.998, which is the highest \( R_{adj}^2 \) among all models. On the basis of RMSE, AIC, and BIC, the double sigmoid models give smaller values than the single sigmoid models, meaning the double sigmoid models fit better. The Double Logistic outperforms all the models with the smallest value of RMSE, AIC, and BIC.

<table>
<thead>
<tr>
<th>Model</th>
<th>( R_{adj}^2 )</th>
<th>RMSE</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>0.978</td>
<td>0.037</td>
<td>-230.56</td>
<td>-225.80</td>
</tr>
<tr>
<td>Gompertz</td>
<td>0.967</td>
<td>0.046</td>
<td>-215.48</td>
<td>-210.73</td>
</tr>
<tr>
<td>Double Logistic</td>
<td>0.998</td>
<td>0.010</td>
<td>-316.86</td>
<td>-307.36</td>
</tr>
<tr>
<td>Double Gompertz</td>
<td>0.997</td>
<td>0.014</td>
<td>-297.80</td>
<td>-288.29</td>
</tr>
</tbody>
</table>

4. Conclusions

In conclusion, this study has successfully proposed the double sigmoid models with reparameterization to model the growth pattern of bacteria. The results show that the double sigmoid models fit the growth curve of Pseudomonas putida better than the single sigmoid model based on the \( R_{adj}^2 \), RMSE, AIC, and BIC. The Double Logistic model outperforms all the models with the highest value of \( R_{adj}^2 \) and the smallest value of RMSE, AIC, and BIC. For further research, these double sigmoid models can be applied to model the growth of other bacteria and microorganisms. Other growth models, such as Brody, Richard, and Von Bartaleny, can be used instead of Logistics and Gompertz models.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Petra Christian University Indonesia, for supporting this research and publication. The authors would also like to thank the reviewers for all their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References


