HIBISCUS PUBLISHER

JOURNAL OF ENVIRONMENTAL MICROBIOLOGY AND TOXICOLOGY

Website: http://journal.hibiscuspublisher.com/index.php/JEMAT/index

Testing the Normality of Residuals on Regression Model for the Growth of Sludge Microbes on PEG 600

Halmi M.I.E.[1], Shukor, M.S.[2], Masdor, N.A.[3], Shamaan, N.A.[4] and Shukor, M.Y.[2,5*]

[1]Department of Chemical Engineering and Process, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.
[2]Snoc International Sdn Bhd, Lot 343, Jalan 7/16 Kawasan Perindustrian Nilai 7, Inland Port, 71800, Negeri Sembilan, Malaysia.
[3]Biotechnology Research Centre, MARDI, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia
[4]Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, 13th Floor, Menara B, Persiaran MPAJ, Jalan Pandan Utama, Pandan Indah, 55100 Kuala Lumpur, Malaysia.
[5]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.

*Corresponding author:
Associate Prof. Dr. Mohd. Yunus Abd. Shukor
Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.
Email: yunus.upm@gmail.com
Tel: +603-89466722
Fax: +603-89430913

HISTORY

Received: 21st May 2015
Received in revised form: 22nd of June 2015
Accepted: 5th of July 2015

KEYWORDS

Polyethylene Glycol
modified Gompertz
sludge microbes
ordinary least squares method
normality test

ABSTRACT

Polyethylene glycols (PEGs) are employed in numerous sectors. PEGs are nephrotoxic and their biodegradation by microbes could be a potential tool for bioremediation. Numerous bacterial growth studies neglect primary modelling even though modelling exercises can reveal important parameters. Previously, we have utilized several growth models to model the growth of sludge microbes on PEG 600. We discovered that the modified Gompertz model via nonlinear regression utilizing the least square method was the best model to describe the growth curve. However, the use of statistical tests to choose the best model relies heavily on the residuals of the curve to be statistically robust. Normality tests for the residuals used in this work has indicated that the use of the modified Gompertz model in fitting of the growth curve of the sludge microbes on PEG 600 initially was not adequate due to the presence of an outlier. Upon removal of this outlier, the residuals conformed to normality test, visually and statistically.

INTRODUCTION

Synthetic polymer such as Polyethylene glycols (PEGs) are employed in a variety of industrial areas such as cosmetics, lubricants, pharmaceuticals, and antifreeze for automobile radiators plus in the creation of non-ionic surfactants. PEGs are nephrotoxic. Injured rabbit put through topically to polyethylene glycol-based antimicrobial cream model exhibited evidence nephrotoxicity with indications of failure. Several of the animals examined died within just 1 week of treatment [1]. Several millions of tons of PEGs are manufactured globally. Effluents contaminated with PEGS usually reach conventional sewage treatment systems making them a significant pollutant [2]. PEGs have the common structural formula of $HO(CH_2CH_2O)nCH_2CH_2OH$ and are water-soluble polymers but the difference is in their molecular weights. From the last three decades, concern has been expressed about the fate of these polymers in the environment and several studies have been performed on their biodegradability. Biodegradation of PEG was first documented in 1965 [3] and further isolations of PEG-degrading microorganisms have been reported [2].

Similar to numerous xenobiotics, the growth on this toxic substrate display a substantial lag phase because of the needs of the cell to endure and trigger detoxification and degradation of enzymes upon contact with the substrate before assimilation can occur. The growth profile displays a number of phases in which the specific growth rate begins at the value of zero accompanied by a stagnation of the rate linked to the lag time (λ). This is followed by acceleration to a maximal value (μ_m) for a given period of time. Finally the growth curves exhibit a final phase where the rate decreases and eventually reaches zero or an asymptote (A) [4]. A valuable parameter of the growth is the maximum growth rate (μ_m) [5]. This value is important for the development of secondary models such as growth kinetics [6]. Previously, we have utilized several growth models to model the growth of sludge microbes on PEG 600. The data was obtained from the literature. We discovered that the modified

Gompertz model via nonlinear regression utilizing the least square method was the best model to describe the growth curve. However, the use of statistical tests to choose the best model relies heavily on the residuals of the curve to be distributed normally. We perform statistical diagnosis tests for normality such as the Kolmogorov-Smirnov, Wilks-Shapiro and D'Agostino-Pearson on the residuals from the regression model utilized in modelling the growth data.

METARIALS AND METHODS

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [7] which helps to digitize scanned plots into table of data with good enough precision [4]. Data were acquired from the works of Huang et al. [8], from Figure 1 which show the effect of different concentrations of the substrate PEG 600 on the growth of sludge microbes measured over several days, replotted, and then assessed using several growth models where the modified Gompertz model was found to be the best (**Fig. 1,** with permission) (Halmi, M,I.E., Shukor, M.S., Shamaan, N.A. and Shukor, M.Y. 2015. Evaluation of several mathematical models for fitting the growth of sludge microbes on PEG 600. Manuscript in preparation).

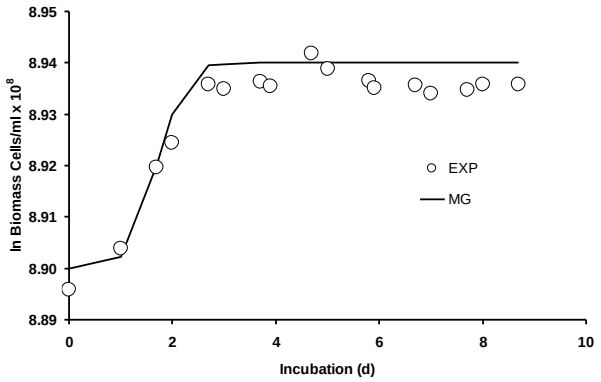

Fig. 1. Growth curves of sludge microbes on PEG 600 fitted by the modified Gompertz growth model.

Normality test

Residuals from the modified Gompertz model were subjected to three normality tests- Kolmogorov-Smirnov [9,10], Wilks-Shapiro [11] and the D'Agostino-Pearson omnibus K2 test [12]. Two ways to check for normality are through graphical and numerical means. Graphical methods such as the normal quantile–quantile (Q-Q) plots, histograms or box plots are the simplest and easiest way to assess normality of data. The detail mathematical basis of these normality test statistics is extensive and is available in the literature [13]. The normality tests were carried out using the GraphPad Prism® 6 (Version 6.0, GraphPad Software, Inc., USA).

Residuals are very important in assessing the health of a curve from a particular used model. Mathematically, residual for the i^{th} observation in a given data set can be defined as follows (Eqn. **1**);

$$e_i = y_i - f\left(x_i;\hat{\beta}\right) \qquad (1)$$

where y_i denotes the i^{th} response from a given data set while x_i is the vector of explanatory variables to each set at the i^{th} observation corresponding values in the data set.

RESULTS AND DISCUSSION

The fit of a statistical model is usually determined precisely using tests which use residuals. Residuals are the contrast between a predicted and observed quantity utilizing a particular mathematical model. The general rule is that the larger the differenced between the predicted and observed values, the poorer the model. Plot of residuals (observed-predicted) were checked and the analysis showed that the data were not randomly distributed for all tests (**Table** 1). This could indicate the presence of an outlier in the residual. The Grubbs' test was applied (results published elsewhere) in order to identify the outlier(s). The Grubbs' test statistic identifies the largest absolute deviation from the sample mean in units of the sample standard deviation [14]. The Grubbs' test identify an outlier for the residual data 0.05521 at a significance level of 5% (α=5%). The presence of this outlier was graphically indicated using the residual plot (**Fig**. 2).

Table 1. Numerical normality test for the residual from the Buchanan- three phase model.

Normality test	Analysis
D'Agostino & Pearson omnibus normality test	
K2	21.06
P value	< 0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****
Shapiro-Wilk normality test	
W	0.5707
P value	< 0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****
KS normality test	
KS distance	0.4351
P value	< 0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****

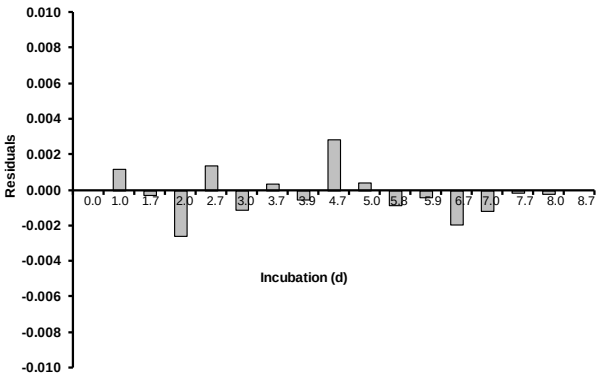

Fig. 2. Residual plot for the modified Gompertz model model.

The outlier identified by the Grubb's test was removed and the same tests were again applied in order to assess the normality. The results are presented in **Table** 2. The removal of this outlier as indicated using the residual plot shows uniform random distribution (**Fig**. 3).

Table 2. Numerical normality test for the residual from the modified Gompertz model after removal of an outlier.

Normality test	Analysis
D'Agostino & Pearson omnibus normality test	
K2	0.4997
P value	0.7789
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
Shapiro-Wilk normality test	
W	0.9731
P value	0.9011
Passed normality test (alpha=0.05)?	Yes
P value summary	ns
KS normality test	
KS distance	0.1291
P value	> 0.1000
Passed normality test (alpha=0.05)?	Yes
P value summary	ns

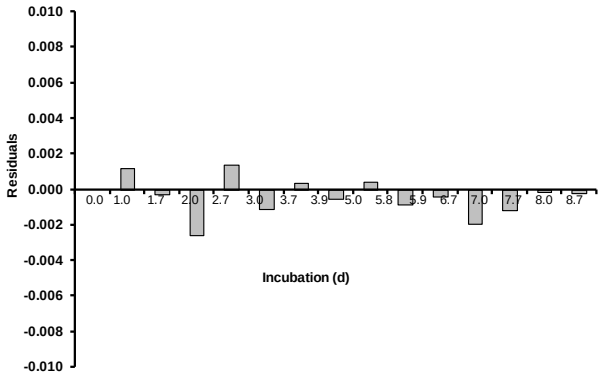

Fig. 3. Residual plot for the modified Gompertz model after removal of an outlier.

The normal probability Q-Q plot of residuals for the modified Gompertz model was almost in a straight line and appears to show no underlying pattern (**Fig**. 4). The resulting histogram overlaid with the resulting normal distribution curve (**Fig**. 5) indicates the residuals were truly random and the model used was appropriately fitted.

Graphical diagnostic of residuals normality

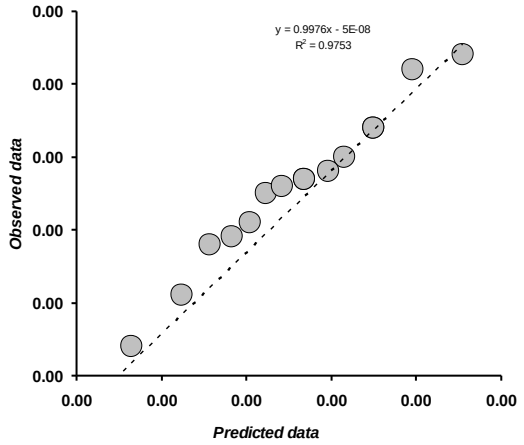

Fig 4. Normal Q-Q plot for the observed sample against theoretical quantiles.

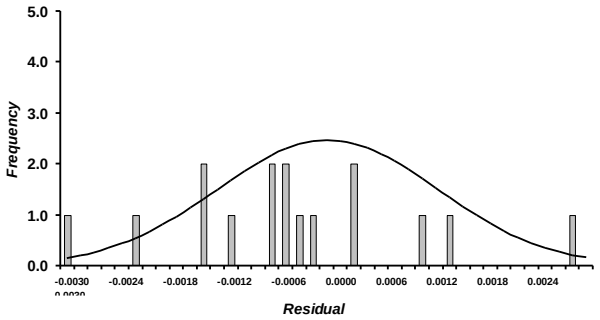

Fig. 5. Histogram of residual for the modified Gompertz model overlaid with a normal distribution (mean 0.000227 and standard deviation 0.003344).

After the removal of the outlier, all of the normality tests used showed that the residuals were normally distributed (**Table** 2). Number of bins and samples analyzed decided the shape of the distribution. In the Wilks-Shapiro test, a W^2 statistic is determined depending on the anticipated values of the order statistics in between identically-distributed random variables as well as their independent covariance plus the standard normal distribution, correspondingly. In the event the test statistics value-W^2 is high, then the agreement is rejected [11]. The Kolmogorov-Smirnov statistic is a non-parametric numerical test that compares the cumulative frequency of residuals. It calculates the agreement between the model and observed values. It could also be used as a measure between two series of observation. The *p* value is calculated for the difference between two cumulative distributions and sample size [9,10].

The skewness and kurtosis of the distribution is computed as a method to quantify the difference between the sample distributions to a normal distribution In the D'Agostino-Pearson normality test method. A *p*-value from the sum of these discrepancies is then computed. The most often form of the D'Agostino-Pearson normality tests is the omnibus K2 test as D'Agostino developed several normality tests [12].

In conclusion, normality tests for the residuals used in this work has indicated that the use of the modified Gompertz model in fitting of the growth curve of the sludge microbes on PEG 600 initially was not adequate due to the presence of an outlier. Upon removal of this outlier, the residuals conform to normality test, visually and statistically. It is well known that many publications did not elaborate further on the use of statistical diagnosis of the residuals from the model used. This could results in data violating the Gaussian or normal distribution. This assumption is an important requirement for many of the parametric statistical evaluation methods used in non linear regression. Methods such as the Pearson's correlation coefficient either normal or adjusted, root mean square analaysis, F-test and t-test rely on the residuals to be normally distributed. These assumptions could avoid errors of the Type I and II errors. Furthermore, in the event that the dignostic tests shows that the residuals violated some of the assumptions various nonparametric treatments could be used or changing to a different model can in practice remedy the situation.

ACKNOWLEDGEMENT

This project was supported by a grant from Snoc International Sdn Bhd.

REFERENCES

[1] Herold DA, Rodeheaver GT, Bellamy WT, Fitton LA, Bruns DE, Edlich RF. Toxicity of topical polyethylene glycol. Toxicol Appl Pharmacol. 1982;65(2):329–335.

[2] Watanabe M, Kawai F. Study on biodegradation process of polyethylene glycol with exponential growth of microbial population. 2010. 145-157.

[3] Payne WJ, Williams JP, Mayberry WR. Primary alcohol sulfatase in a Pseudomonas species. Appl Microbiol. 1965;13(5):698–701.

[4] Halmi MIE, Shukor MS, Johari WLW, Shukor MY. Evaluation of several mathematical models for fitting the growth of the algae *Dunaliella tertiolecta*. Asian J Plant Biol. 2014;2(1):1–6.

[5] Zwietering MH, Jongenburger I, Rombouts FM, Van't Riet K. Modeling of the bacterial growth curve. Appl Environ Microbiol. 1990;56(6):1875–1881.

[6] Ahmad SA, Ahamad KNEK, Johari WLW, Halmi MIE, Shukor MY, Yusof MT. Kinetics of diesel degradation by an acrylamide-degrading bacterium. Rendiconti Lincei. 2014;25(4):505–512.

[7] Rohatgi, A. WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer/app/ Accessed June 2 2014.

[8] Huang Y-L, Li Q-B, Deng X, Lu Y-H, Liao X-K, Hong M-Y, et al. Aerobic and anaerobic biodegradation of polyethylene glycols using sludge microbes. Process Biochem. 2005;40(1):207–211.

[9] Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G Dell' Ist Ital Degli Attuari. 1933;4:83–91.

[10] Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat. 1948;19:279–281.

[11] Royston P. Wilks-Shapiro algorithm. Appl Stat. 1995;44(4):R94.

[12] D'Agostino RB. Tests for Normal Distribution. In: D'Agostino RB, Stephens MA, editors. Goodness-Of-Fit Techniques. Marcel Dekker; 1986.

[13] Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J Off Publ Fed Am Soc Exp Biol. 1987;1(5):365–374.

[14] Grubbs F. Procedures for detecting outlying observations in samples. Technometrics. 1969;11(1):1–21.