

Short Communication

Outlier Analysis of the Modified Gompertz Model used in fitting 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*}

¹Department of Chemical Engineering and Process, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

²Snoc International Sdn Bhd, Lot 343, Jalan 7/16 Kawasan Perindustrian Nilai 7, Inland Port, 71800, Negeri Sembilan, Malaysia.

³Biotechnology Research Centre, MARDI, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia

⁴Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, 13th Floor, Menara B, Persiaran MPAJ, Jalan Pandan Utama, Pandan Indah, 55100 Kuala Lumpur, Malaysia.

⁵Department 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: 20th of May 2015
Received in revised form: 20th of June 2015
Accepted: 1st of July 2015

KEYWORDS

Polyethylene Glycol
modified Gompertz
sludge microbes
ordinary least squares method
Grubbs 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. More often than not, the residuals must be tested for the presence of outliers (at 95 or 99% of confidence). In this work, the Grubb's test to detect the presence of outlier in the growth model was carried out. The test detected an outlier. This datum point will be removed in all future statistical tests such as normality, runs test, tests for homoscedasticity and presence of autocorrelation. In addition, remodeling of the data using the modified Gompertz model will be carried out.

INTRODUCTION

Polyethylene glycols (PEGs) are utilized in numerous industrial sectors for example cosmetics, lubricants, pharmaceuticals, and antifreeze for automobile radiators as well as in the production of non-ionic surfactants. PEGs are nephrotoxic. Injured rabbit subjected topically to polyethylene glycol-based antimicrobial cream model demonstrated proof of nephrotoxicity with signs and symptoms of failure. Some 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 $\text{HO}(\text{CH}_2\text{CH}_2\text{O})_n\text{CH}_2\text{CH}_2\text{OH}$ 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 catechol 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, of equal variance (homoscedastic) and random. More often than not, the residuals must be tested for the presence of outliers [at 95 or 99% of confidence). This is normally done using the Grubb's test.

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).

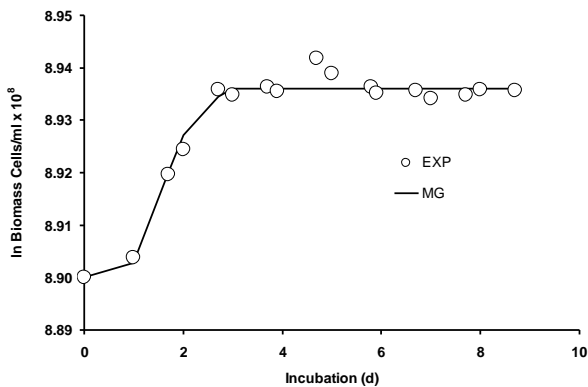


Fig. 1. Growth curves of sludge microbes on PEG 600 fitted by the modified Gompertz growth model.

Grubbs' Statistic

Data distortions by a single data point either the mean or a single data point from a triplicate can lead to gross error in the fitting of a nonlinear curve. Checking for outlier is thus an important part of curve fitting. Grubbs test is used to detect outlier in univariate environment and the data is assumed to be normally distributed [9]. The test can be applied to the maximal or minimal observed data from a Student's t distribution (Eqn. 1) and to test for both data simultaneously (Eqn. 2).

$$G_{\min} = \frac{\bar{X} - \min(X)}{s} \tag{1}$$

$$G_{\max} = \frac{\max(X) - \bar{X}}{s} \tag{1}$$

$$p_G = 2n \cdot p_t \left(G \frac{\sqrt{n(n-2)}}{n-1}, n-2, 1 \right)$$

$$G_{\text{all}} = \frac{\max(\bar{X} - \min(X), \max(X) - \bar{X})}{s}$$

$$p_G = n \cdot p_t \left(G \frac{\sqrt{n(n-2)}}{n-1}, n-2, 2 \right) \tag{2}$$

RESULTS AND DISCUSSION

Statistics of nonlinear regression relies heavily on the use of residuals data. Residuals are the difference between predicted and observed values of a mathematical model. Statistical tests should be carried out to test for the adequacy of the residuals in randomness, does not contain outlier, obeying normality and do not show autocorrelation. As a rule of thumb, the larger the difference between the predicted and observed values, the poorer the model [10].

Data distortions by a single data point either the mean or a single data point from a triplicate can lead to gross error in the fitting of a nonlinear curve. The Grubbs' test shows the absence of outlier for the residual data. Checking for outlier is thus an important part of curve fitting [9]. Grubbs' test identifies one outlier at any given time. This outlier is expunged from the dataset and also the analyze is iterated till no outliers are discovered. Nevertheless, numerous iterations alter the likelihood of detection, and also the test must not be employed for sample sizes of six or less as it regularly tags the majority of the points as outliers.

Grubbs' test was applied 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. In the event the ensuing test statistic g is larger compared to the critical value, the related value is usually considered to generally be an outlier. The Grubbs' test indicated that the data for day 5 incubation period as an outlier. This datum point will be removed in all future statistical tests such as normality, runs test, tests for homoscedasticity and presence of autocorrelation. In addition, remodeling of the data using the modified Gompertz model will be carried out.

A possible outlier is really an extreme data point that the investigator tags as improbable in view of a few detailed requirements. Much more specifically, an outlier in a sample is really an extreme value that is definitely too severe. For example, the maximum is considered an outlier when it is statistically too big for that distribution on the maximum in the population model (Barnett & Lewis, 1996, Saporta, 2011) [11]. A simple method to label potential outliers in measurements is to use boxplot, while a bit more advanced method is also used such as the Chauvenet's criterion in engineering and the 3-sigma criterion together with the Z-score in chemometrics. Although these methods are simple and quick, a more correct method is to use a statistical test for outlier detection. Specific tests include the Dixon's Q-test or Grubbs' ESD-test for one outlier. The main restriction of the Grubbs test would be that the

thought quantity of outliers, k , should be described precisely. If k is not described properly, this could distort the findings of the tests. In the event outliers are multiple or the exact number of outliers are not known, the Rosner's generalized Extreme Studentized Deviate or ESD-test is recommended [12].

In conclusion, the Grubbs' test indicated that the data for day 5 incubation period as an outlier. This datum point will be removed in all future statistical tests such as normality, runs test, tests for homoscedasticity and presence of autocorrelation. In addition, remodeling of the data using the modified Gompertz model will be carried out. The test is important because data distortions by a single data can lead to gross error in the fitting of a nonlinear curve.

ACKNOWLEDGEMENT

This project was supported by a fund 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] Grubbs F. Procedures for detecting outlying observations in samples. *Technometrics*. 1969;11(1):1–21.
- [10] López S, Prieto M, Dijkstra J, Dhanoa MS, France J. Statistical evaluation of mathematical models for microbial growth. *Int J Food Microbiol*. 2004;96(3):289–300.
- [11] Barnett V, Lewis T. *Outliers in Statistical Data*. 3rd ed. Chichester; New York: Wiley; 1994. 604.
- [12] Rosner B. *Fundamentals of biostatistics*. 7th ed. Boston: Brooks/Cole; 2011.