HIBISCUS PUBLISHER

JOURNAL OF ENVIRONMENTAL MICROBIOLOGY AND TOXICOLOGY

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

Short Communication

Test of Randomness of Residuals for the Modified Gompertz model used in the 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*]

[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: 26th of June 2015

KEYWORDS

Polyethylene Glycol
modified Gompertz
sludge microbes
ordinary least squares method
runs 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. Nonlinear regression using the least square method normally uses the assumption that data points do not depend on each other or the value of a data point is not dependent on the value of preceding or proceeding data points or do not exhibit autocorrelation. In this work, the Durbin–Watson statistic for the presence of autocorrelation in the growth model was carried out.

INTRODUCTION

Manmade polymer including 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 $HO(CH_2CH_2O)nCH_2CH_2OH$ 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]. The growth profile of microorganism on this substrate 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. The method of mathematically fitting nonlinear curve using the ordinary least squares method relies heavily on the residuals for the curve to be normally distributed of equal variance (homoscedastic), and does not show autocorrelation [7–9]. Aside from this, an important consideration that has not been highlighted enough is that the residuals must be random. In order for randomness to be met we perform the Wald–Wolfowitz runs test [10] statistical diagnosis tests.

MATERIALS AND METHODS

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [11] 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. [12], 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).

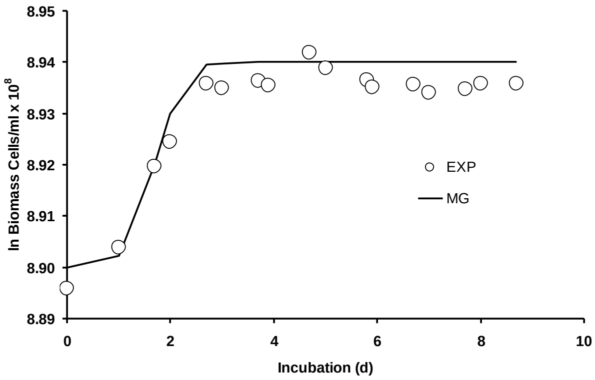

Fig. 1. Growth curves of sludge microbes on PEG 600 fitted by the modified Gompertz growth model.

Runs test

The runs test [13] was carried out to the residuals of the regression in order to detect nonrandomness. This could detect a systematic deviation of over or under estimation sections of the curve when using a specific model [10]. The runs test look at the sequence of the residuals that are usually positive and negative. A good runs is usually signifies by alternating or a balance number of positive and negative residual values. The number of runs of sign is usually expressed in the form of a percentage of the maximum number possible. The runs test calculates the probability for the presence of too many or too few runs of sign. The presence of too many of a run sign could indicate the presence of negative serial correlation whilst the presence of too few runs could indicate a clustering of residuals with the same sign or the presence of systematic bias.

The test statistic is

H_0= the sequence was produced randomly
H_a= the sequence was not produced randomly

$$Z=\frac{R-\bar{R}}{sR} \qquad (1)$$

Where Z is the test statistic, $\bar{R}$ is the expected number of runs, R is the observed number of runs and sR is the standard deviation of the runs. The computation of the values of $\bar{R}$ and sR (n_1 is positive while n_2 is negative signs) is as follows;

$$\bar{R}=\frac{2n_1.n_2}{n_1+n_2}+1 \qquad (2)$$

$$s^2R=\frac{2n_1.n_2\left(2n_1.n_2-n_1-n_2\right)}{\left(n_1+n_2\right)^2\left(n_1+n_2-1\right)} \qquad (3)$$

As an example

Test statistic: Z = 3.0
Significance level: α = 0.05
Critical value (upper tail): $Z_{1-\alpha/2}$ = 1.96
Critical region: Reject H_0 if $|Z|$ > 1.96
Since the test statistic value (Z) is larger than the critical value then the null hypothesis is rejected at the 0.05 significance level or the sequence was produced in a nonrandom manner.

RESULTS AND DISCUSSION

The fit of a statistical model can be diagnosed accurately using tests that use residuals. Residuals are the difference between a predicted and observed quantity using a particular mathematical model. The rule of thumb is that the larger the difference between the predicted and observed values, the poorer the model.

Runs test

From **Table** 2, the number of runs was 13, the expected number of runs under the assumption of randomness was 7.461538, indicating the series of residuals had adequate runs. The z-value indicates how many standard errors the observed number of runs is below the expected number of runs, the corresponding p-value indicate how extreme this z-value is. The interpretation is the same like other o-values statistics. If the p-value is less than 0.05 then the null hypothesis that the residuals are indeed random can be rejected. Since the p-value was greater than 0.05, therefore the null hypothesis is not rejected indicating no convincing evidence of non-randomness of the residuals and they do represent noise.

Table 2. Runs test for randomness.

Runs test	Residual data set
observations	12
below mean	8
above mean	4
no of runs	5
E(R)	6.333
var(R)	2.101
stdev(R)	1.449
Z-value	-0.920
p-value	0.179

The runs test calculates the probability for the presence of too many or too few runs of sign. The presence of too many of a run sign could indicate the presence of negative serial correlation whilst the presence of too few runs could indicate a clustering of residuals with the same sign or the presence of systematic bias. The runs test is an important tool in nonlinear regression to detect nonrandomness of the residuals [13]. The runs test could detect systematic deviation of the curve such as over or under estimation of the sections when using a specific model. The runs test look at the sequence of the residuals that are usually positive and negative. A good runs is usually signifies by alternating or a balance number of positive and negative residual values. The number of runs of sign is usually expressed in the form of a percentage of the maximum number possible [10].

The runs test has also been utilized as a technique to test for autocorrelation in time-series regression models. However, simulation studies using Monte Carlo have shown that the runs test produces distinctly asymmetrical error rates in the two tails [14]. The investigation is carried out to analyse the empirical properties of the runs test utilizing (a) sample sizes of between 12 and 100 (b) using non-intervention and intervention regression models, (c) utilizing directional and nondirectional tests (d) with three levels of α, and (e) with 19 levels of autocorrelation among the errors. In addition, both directional and nondirectional tests produce no satisfactory results with respect to Type I error. The increase of the ratio of degrees of freedom to sample size to as high as .98 could also not remedy the situation. Hence, the Durbin-Watson method would be the method of choice to assess autocorrelation.

In conclusion, various tests for the residuals used in this work has indicated that the use of the Baranyi-Roberts model in fitting of the growth curve of an algae shows adequate statistics strength based on the diagnostics test of the residuals. Many publications negate statistical diagnosis of the model they used and the data may be nonrandom- an important requirement for all of the parametric statistical evaluation methods. In the event that the diagnostic tests shows that the residuals indicate a trend, then various treatments such as nonparametric analysis or changing to a different model should remedy the problem.

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 p.

[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] Razali NM, Wah YB. Power comparisons of Shapiro–Wilk, Kolmogorov– Smirnov, Lilliefors and Anderson–Darling tests. J Stat Model Anal. 2011;2:21–23.

[8] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Econ Lett. 1981;7(4):313–318.

[9] Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames Iowa: Iowa State University Press; 1980.

[10] 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.

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

[12] 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.

[13] Draper NR, Smith H. Applied Regression Analysis. Wiley, New York; 1981.

[14] Huitema BE, McKean JW, Zhao J. The runs test for autocorrelated errors: unacceptable properties. J Educ Behav Stat. 1996;21(4):390–404.