HIBISCUS PUBLISHER

JOURNAL ENVIRONMENTAL BIOREMEDIATION & TOXICOLOGY

http://journal.hibiscuspublisher.com/index.php/JEBAT

Short Communication

Outlier Analysis of the Buchanan-three-phase Model used in Fitting the Growth of *Paracoccus* sp. SKG on Acetonitrile

Gunasekaran, B.[1*], Shukor, M.S.[2], Masdor, N.A.[3], Shamaan, N.A.[4], and Shukor, M.Y.[1,2]

[1]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, 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.

*Corresponding author:
Dr. Baskaran Gunasekaran
Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.
Email: berganza1985@yahoo.com

HISTORY

Received: 3rd May 2015
Received in revised form: 15th June 2015
Accepted: 7th of July 2015

KEYWORDS

acetonitrile-degrading
Buchanan-three-phase
Paracoccus sp. SKG
ordinary least squares method
Grubbs test

ABSTRACT

Organonitriles are widely used in industry such as the synthesis of plastics, rubber, herbicides, pharmaceuticals, drug intermediates, and pesticides. They are carcinogenic and mutagenic. Bioremediation of acetonitrile, an organonitrile, has been touted as a more economical and feasible method compared to physical and chemical approaches. In this work, we model the growth of growth of *Paracoccus* sp. SKG on acetonitrile from published literature to obtain vital growth constants. We discovered that the Buchanan-three-phase model via nonlinear regression utilizing the least square method was the very best model to explain 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.

INTRODUCTION

Organonitriles are classified as priority pollutants. The global industrial consumption of acetonitrile alone is more than 4×10^4 tonne in 2001 [1,2]. Consequently, wastewaters from the various usages of organonitriles often contain high contents of organonitrile compounds. Organonitriles are carcinogenic and mutagenic. They are widely used in industry such as the synthesis of plastics, rubber, herbicides, pharmaceuticals, drug intermediates, and pesticides. In addition, acetonitrile, an organonitrile, is extensively utilized in laboratories as a solvent and extractant for HPLC (High Performance Liquid chromatography). Bioremediation of acetonitrile has been touted as a more economical and feasible method compared to physical and chemical approaches. Santoshkumar et al [3] has isolated a bacterial strain that could grow on acetonitrile. The growth profile of the strain showed inhibition of growth at elevated concentrations of acetonitrile. Modelling of the growth curves can yield important parameters that could be used for further secondary modelling exercise such as the inhibitory effect of substrate on growth.

Previously, we have utilized several growth models to model the growth of *Paracoccus* sp. SKG on acetonitrile using published available data from [3]. We discovered that the Buchanan-three-phase model via nonlinear regression utilizing the least square method was the best model to describe the growth curve (published elsewhere). 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.

METHODOLOGY

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [4] which helps to digitize scanned plots into table of data with good enough precision [5]. Data were acquired from the works of Santoshkuma et al. [3] from Figure 4 and then replotted, and then assessed using several growth models where the Buchanan-three-phase model was found to be the best (**Fig. 1,** with permission) (Shukor, M.S., Masdor, N.A., Shamaan, N.A., Ahmad, S.A.,

Roslan, M.A.H. and Shukor, M.Y. 2015. The growth of *Paracoccus* sp. SKG on acetonitrile is best Modelled using the Buchanan Three Phase Model. Manuscript in preparation).

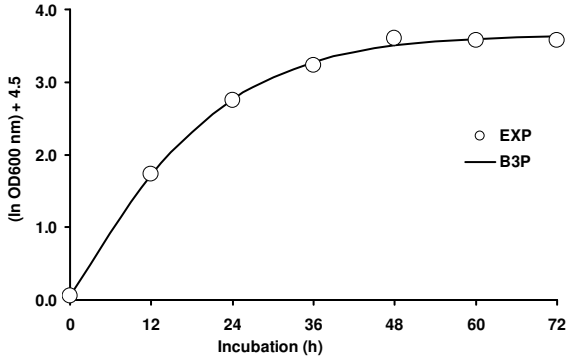

Fig. 1. Growth curves of *Paracoccus* sp. SKG on acetonitrile fitted by the Buchanan-three-phase model.

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 [6]. 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**).

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

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

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

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

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

In the event that there is more than one outlier, then the ROUT method [7] will be used. The method is based on the False Discovery Rate (FDR). Q, a chance of (falsely) identifying one or more outliers will have to be set manually. It is the maximum desired FDR. In the absence of outliers, Q is very similar to alpha. Assuming all data come from a Gaussian distribution.

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 [8]. The residuals for the Buchanan-three-phase model are shown in Table 1.

Table 1. Residuals for the Buchanan-three-phase model utilized in the modelling of the growth of *Paracoccus* sp. SKG on acetonitrile.

Time (h)	Residuals
0	0.00
12	0.01
24	-0.01
36	-0.04
48	0.10
60	-0.01
72	-0.05

Grubbs' test identifies one outlier at any given time. This outlier is expunged from the dataset and the analysis is iterated until 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 there was no outlier.

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) [9]. 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 [10].

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] Chapatwala KD, Nawaz MS, Richardson JD, Wolfram JH. Isolation and characterization of acetonitrile utilizing bacteria. J Ind Microbiol. 1990;5(2-3):65–70.

[2] Li C, Li Y, Cheng X, Feng L, Xi C, Zhang Y. Immobilization of *Rhodococcus rhodochrous* BX2 (an acetonitrile-degrading

bacterium) with biofilm-forming bacteria for wastewater treatment. Bioresour Technol. 2013;131:390–6.

[3] Santoshkumar M, Veeranagouda Y, Lee K, Karegoudar TB. Utilization of aliphatic nitrile by *Paracoccus* sp. SKG isolated from chemical waste samples. Int Biodeterior Biodegrad. 2011;65(1):153–9.

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

[5] 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.

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

[7] Motulsky HJ, Brown RE. Detecting outliers when fitting data with nonlinear regression - A new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics. 2006;7.

[8] 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.

[9] Barnett V, Lewis T. Outliers in Statistical Data. 3rd ed. Chichester ; New York: Wiley; 1994. 604 p.

[10] Rosner B. Fundamentals of biostatistics. 7th ed. Boston: Brooks/Cole; 2011.