HIBISCUS PUBLISHER

JOURNAL ENVIRONMENTAL BIOREMEDIATION & TOXICOLOGY

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

JOURNAL OF ENVIRONMENTAL BIOREMEDIATION AND TOXICOLOGY

Testing the Normality of Residuals on Regression Model for the Growth of *Paracoccus* sp. SKG on Acetonitrile

Halmi, M.I.E[1*], Shukor, M.S.[2], Masdor, N.A.[3], Shamaan, N.A.[4], Sabullah, M.K.[5], and Shukor, M.Y.[2,6*]

[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]Faculty of Food Science and Nutrition, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
[6]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.

*Corresponding author:
Dr. Mohd Izuan Effendi Bin Halmi
Department of Chemical and Process Engineering
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor
MALAYSIA
Email: zuanfendi@ukm.edu.my /zuanfendi88@gmail.com
Tel: +603 89216428
Fax: +603 89118345

HISTORY

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

KEYWORDS

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

ABSTRACT

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 conformation to normal distribution. In order for these assumptions to be met, we perform statistical diagnosis tests such as the Kolmogorov-Smirnov, Wilks-Shapiro and D'agostino-Pearson tests.

INTRODUCTION

Acetonitrile, an organonitrile, is extensively utilized in laboratories as a solvent and extractant for HPLC (High Performance Liquid chromatography). 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. 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. 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. 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.

METHODOLOGY

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).

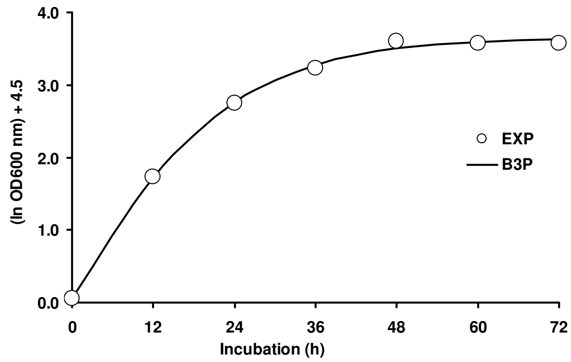

Fig. 1. Growth curves of *Paracoccus* sp. SKG on acetonitrile fitted by the Buchanan-three-phase model.

Normality test

Residuals from the Buchanan-three-phase model were subjected to three normality tests- Kolmogorov-Smirnov [6,7], Wilks-Shapiro [8] and the D'Agostino-Pearson omnibus K2 test [9]. 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 [10]. 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**);

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

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 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 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 normal with the exception of the D'agostino & Pearson omnibus normality test that indicated that the data was too small for the test. In practice, the data is considered to be normally distributed based on the majority of the tests (**Table** 1). The residuals plot does no indicate data that supported non normal distribution (**Fig**. 2).

The normal probability Q-Q plot of residuals for Buchanan-three-phase model was almost in a straight line and appears to show no underlying pattern (**Fig**. 3). The resulting histogram overlaid with the resulting normal distribution curve (**Fig**. 4) indicates the residuals were truly random and the model used was appropriately fitted.

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

Normality test	Analysis
D'agostino & Pearson Omnibus Normality Test	
K2	n too small
P Value	
Passed Normality Test (Alpha=0.05)?	
P Value Summary	
Shapiro-Wilk Normality Test	
W	0.8759
P Value	0.2091
Passed Normality Test (Alpha=0.05)?	Yes
P Value Summary	ns
Ks Normality Test	
Ks Distance	0.2681
P Value	> 0.1000
Passed Normality Test (Alpha=0.05)?	Yes
P Value Summary	ns

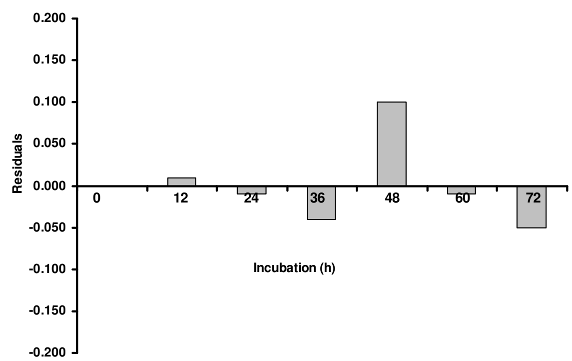

Fig. 2. Residual plot for the Buchanan- three phase model.

Graphical diagnostic of residuals normality

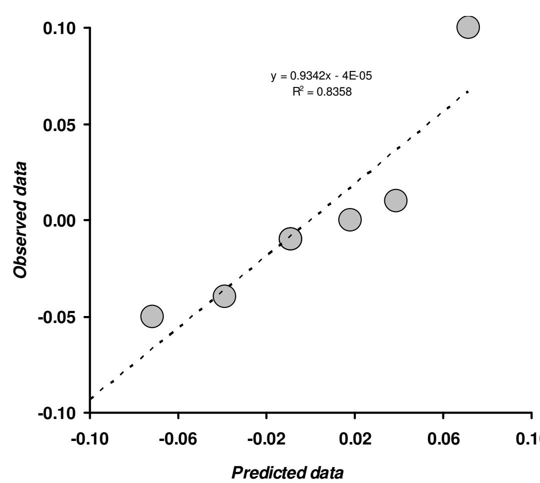

Fig 3. Normal Q-Q plot for the observed sample against theoretical quantiles.

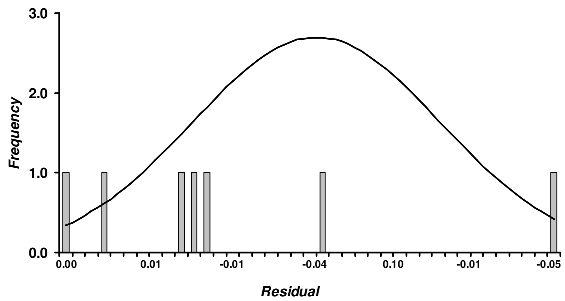

Fig. 4. Histogram of residual for the Buchanan-three-phase model overlaid with a normal distribution.

Number of bins and samples examined determined the shape of the distribution. In the Wilks-Shapiro test, a W^2 statistic is calculated based on the expected values of the order statistics between identically-distributed random variables and their independent covariance and the standard normal distribution, respectively. If the test statistics value-W^2 is high, then the agreement is rejected [8]. 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 [6,7].

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 [9].

In conclusion, normality tests for the residuals used in this work has indicated that the use of the Buchanan-three-phase model in fitting of the growth curve of *Paracoccus* sp. SKG on acetonitrile was adequate. 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] 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] Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G Dell' Ist Ital Degli Attuari. 1933;4:83–91.

[7] Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat. 1948;19:279–81.

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

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

[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):365–74.