BULLETIN OF ENVIRONMENTAL SCIENCE AND MANAGEMENT
http://journal.hibiscuspublisher.com/index.php/BESM

Testing the Normality of Residuals on Regression Model for the Growth of *Moraxella* sp. B on Monobromoacetic Acid

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

[1]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.
[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 Food Science and Nutrition, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
[5]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.

*Corresponding author:
Dr. Mohd. Khalizan Sabullah
Faculty of Food Science and Nutrition, Universiti Malaysia Sabah, Jalan UMS,
88400 Kota Kinabalu, Sabah, Malaysia
Email: ejanz_sart2@yahoo.com
Tel: +6088) 320000 Ext: 8875
Fax : +6088 320259

HISTORY

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

KEYWORDS

monobromoacetic acid-degrading
Buchanan-three-phase
Moraxella sp. B
ordinary least squares method
normality test

ABSTRACT

Bioremediation of monobromoacetic acid, a haloacetic acid, continues to be recommended as a cheaper and achievable method in comparison to physical and chemical techniques. In a prior work, we model the growth of the bacterium *Moraxella* sp. B on monobromoacetic acid from published literature to acquire crucial growth constants. We learned that the Buchanan-three-phase model via nonlinear regression using the least square method was the most effective model to describe the growth curve. Nevertheless, 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

Monobromoacetic acid (MBA) has been utilized as a chemical intermediate for production of numerous chemicals with application in agriculture and pharmacy [1]. Their pollution and water bodies and soils have prompt several workers to isolate MAB-degrading microorganisms as a tool for bioremediation. Aerobic degradation, either cometabolically [2] or as a sole carbon and energy source [1,3–8] have been reported. The strain *Moraxella* sp. B [9] possesses two haloacid dehalogenases that are active with MCA and MBA [10]. Torz et al [7] study the growth of this bacterium on monohaloacetates including MBA. The growth curve displayed sigmoidal attributes; nevertheless, the authors did not take advantage to the presence of various primary growth models to obtain important growth constants, which they can use in further modelling. It is expected that modelling of the growth curves can yield important parameters suitable to 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 *Moraxella* sp. B on monobromoacetic acid. 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 [11].

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

Data from Torz et al. [7] from Figure 1 was scanned and electronically processed using WebPlotDigitizer 2.5 [12] that helps to digitize scanned plots directly into table of data with good adequate precision [13]. The data were then replotted (**Fig. 1,** with permission) [11].

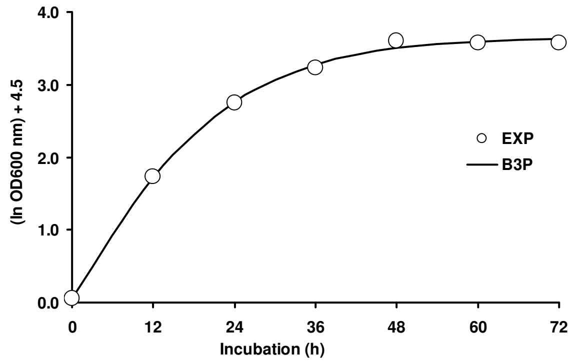

Fig. 1. Growth curves of *Moraxella* sp. B on monobromoacetic acid fitted by the Buchanan-three-phase model.

Normality test

Three normality tests- Kolmogorov-Smirnov [14,15], Wilks-Shapiro [16] and the D'Agostino-Pearson omnibus K2 test [17] were utilized to the residuals from the Buchanan-three-phase model. Two approaches to look for normality are via graphical and numerical means. Graphical techniques including the normal quantile-quantile (Q-Q) plots, histograms or box plots are classified as the easiest and simplest way to evaluate normality of data.

The depth of the mathematical basis of these normality test statistics is substantial and is obtainable in from the literature [18]. The normality tests were carried out using the GraphPad Prism® 6 (Version 6.0, GraphPad Software, Inc., USA). Residuals can be used in assessing the health of a curve from a model. Mathematically, residual for the i^{th} observation in a given data set can be defined as follows (Eqn. **1**);

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

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

A precise determination of the fit of a statistical model is often determined tests which use residuals. Residuals are the difference between a predicted and observed quantity utilizing a particular mathematical model. The general rule would be that the larger the difference between the predicted and observed values, the worse the model. Plot of residuals (observed-predicted) were examined and the normality analysis utilizing three tests demonstrated that the data were normal suggesting the nonlinear regression model utilizing the Buchanan-three-phase via least square method was appropriate (**Table** 1). The residuals plot does no indicate data that supported non normal distribution (**Fig**. 2).

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

Normality test	Analysis
D'agostino & Pearson omnibus	
K2	0.1053
p value	0.9487
passed normality test (alpha=0.05)?	yes
p value summary	ns
Shapiro-Wilk	
w	0.9431
p value	0.4980
passed normality test (alpha=0.05)?	yes
p value summary	ns
Kolmogorov–Smirnov (K—S)	
KS distance	0.1854
p value	> 0.1000
passed normality test (alpha=0.05)?	yes
p value summary	ns

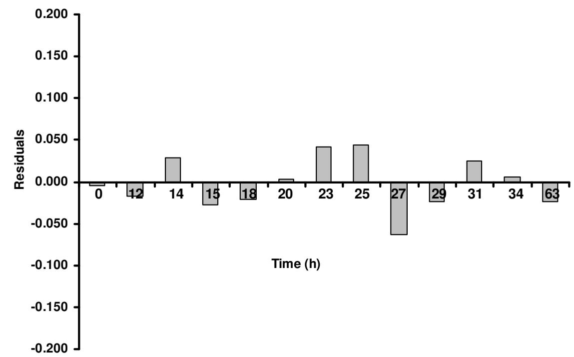

Fig. 2. Residual plot for the Buchanan- three phase model.

The result indicates that the data conformed to normality (**Fig**. 3) based on the normal probability Q-Q plot of residuals for the Buchanan-three-phase model which was almost in a straight line appearing to show no underlying pattern. An overlay of the histogram to the calculated normal distribution curve (**Fig**. 4) indicated that the residuals were truly random and the model used was appropriately fitted.

Graphical diagnostic of residuals normality

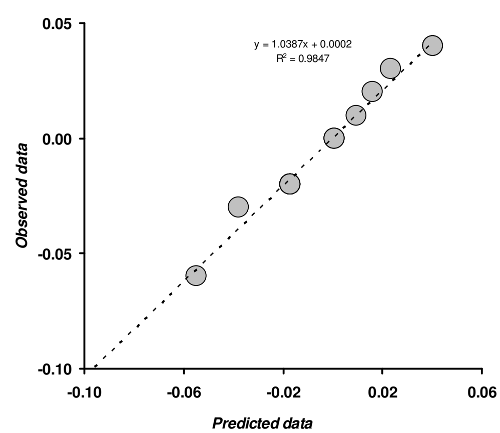

Fig 3. Normal Q-Q plot for the observed sample against theoretical quantiles.

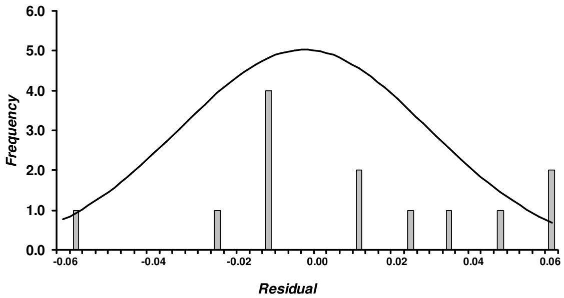

Fig. 4. Histogram of residual for the Buchanan-three-phase model overlaid with a normal distribution.

It is important to note that the number of bins and samples examined determined the shape of the distribution. 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 [14,15].

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 [16]. 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 [17].

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 *Moraxella* sp. B on monobromoacetic acid was adequate. It is observed that numerous publications failed to elaborate more on the use of statistical diagnosis of the residuals from the model employed. This might leads to data violating the Gaussian or normal distribution. This presumption is a vital necessity for many of the parametric statistical assessment methods utilized in non linear regression. Methods including the Pearson's correlation coefficient either normal or adjusted, root mean square analaysis, F-test and t-test depend on the residuals to be normally distributed. The use of adequate and tested assumptions could avoid errors of the Type I and II. In addition, in the event the dignostic tests implies that the residuals broken a number of the assumptions various nonparametric treatments could possibly be used or changing to a different model can in practice correct the situation.

ACKNOWLEDGEMENT

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

REFERENCES

[1] Hamid AAA, Tengku Abdul Hamid TH, Wahab RA, Huyop F. Identification of functional residues essential for dehalogenation by the non-stereospecific α-haloalkanoic acid dehalogenase from *Rhizobium* sp. RC1. J Basic Microbiol. 2015;55(3):324–30.

[2] Weightman AL, Weightman AJ, Slater JH. Microbial dehalogenation of trichloroacetic acid. World J Microbiol Biotechnol. 1992;8(5):512–8.

[3] Yu P, Welander T. Growth of an aerobic bacterium with trichloroacetic acid as the sole source of energy and carbon. Appl Microbiol Biotechnol. 1995;42(5):769–74.

[4] Olaniran AO, Babalola GO, Okoh AI. Aerobic dehalogenation potentials of four bacterial species isolated from soil and sewage sludge. Chemosphere. 2001;45(1):45–50.

[5] McRae BM, Lapara TM, Hozalski RM. Biodegradation of haloacetic acids by bacterial enrichment cultures. Chemosphere. 2004;55(6):915–25.

[6] Olaniran AO, Pillay D, Pillay B. Haloalkane and haloacid dehalogenases from aerobic bacterial isolates indigenous to contaminated sites in Africa demonstrate diverse substrate specificities. Chemosphere. 2004;55(1):27–33.

[7] Torz MS, Yankov DS, Beschkov DN. Biodegradation of mono- and dihaloacetic acids by *Moraxella* sp. B. Comptes Rendus L'Academie Bulg Sci. 2006;59(3):295–300.

[8] Alomar D, Abdul Hamid AA, Khosrowabadi E, Gicana RG, Lamis RJ, Huyop F, et al. Molecular characterization of monochloroacetate-degrading *Arthrobacter* sp. Strain d2 isolated from Universiti Teknologi Malaysia agricultural area. Bioremediation J. 2014;18(1):12–9.

[9] Tonomura K, Kawasaki H, Tone N, Yahara H. Plasmid encoding mercury reductase and haloacetate halidohydrolase in drug-resistant bacteria. J Pharmacobiodyn. 1981;4(4).

[10] Liu J-Q, Kurihara T, Ichiyama S, Miyagi M, Tsunasawa S, Kawasaki H, et al. Reaction mechanism of fluoroacetate dehalogenase from *Moraxella* sp. B. J Biol Chem. 1998;273(47):30897–902.

[11] Shukor MS. Modelling the growth of Moraxella sp. B on Monobromoacetic acid (MBA). Bull Environ Sci Manag. 2015;3(1):1–6.

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

[13] 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.

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

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

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

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

[18] 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.