

## Short Communication

# Test of Randomness of Residuals for the Buchanan-three-phase Model used in the Fitting the Growth of *Paracoccus* sp. SKG on Acetonitrile

Gunasekaran, B.<sup>1\*</sup>, Shukor, M.S.<sup>2</sup>, Masdor, N.A.<sup>3</sup>, Shamaan, N.A.<sup>4</sup>, and Shukor, M.Y.<sup>1,2</sup>

<sup>1</sup>Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.

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

<sup>3</sup>Biotechnology Research Centre, MARDI, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia.

<sup>4</sup>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](mailto:berganza1985@yahoo.com)

## HISTORY

Received: 1<sup>st</sup> May 2015  
Received in revised form: 17<sup>th</sup> of June 2015  
Accepted: 5<sup>th</sup> of July 2015

## KEYWORDS

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

## ABSTRACT

Organonitriles are classified as priority pollutants. 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 *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. An important consideration that has not been highlighted enough is that the residuals of the model must be random. In order for randomness to be met, we perform the Wald-Wolfowitz runs test. The results showed that the number of runs was 5, the expected number of runs under the assumption of randomness was 3.857, indicating the series of residuals had adequate runs. The p-value obtained 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.

## INTRODUCTION

Organonitriles are classified as priority pollutants. The global industrial consumption of acetonitrile alone is more than  $4 \times 10^4$  tonne in 2001 [1,2]. They are carcinogenic and mutagenic, and 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). 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.

Previously, we have utilized several growth models to model the growth of *paracoccus* sp. skg on acetonitrile using published available data from (Santoshkumar et al. 2011). 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). 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 [4–6]. 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 [7] statistical diagnosis tests. 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 [8]. 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 [7].

## METHODOLOGY

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [9] which helps to digitize scanned plots into table of data with good enough precision [10]. 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).

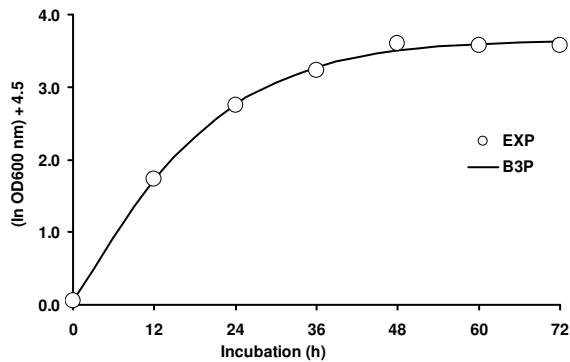


Fig. 1. Growth curves of *Paracoccus* sp. SKG on acetonitrile fitted by the Buchanan-three-phase model.

## Runs test

The runs test [8] 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 [7]. 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

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

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;

$$\bar{R} = \frac{2n_1n_2 + 1}{n_1 + n_2} \quad (2)$$

$$s^2R = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \quad (3)$$

As an example

Test statistic:  $Z = 3.0$

Significance level:  $\alpha = 0.05$

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

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 5, the expected number of runs under the assumption of randomness was 3.857, 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 | 5                 |
| below mean   | 5                 |
| above mean   | 2                 |
| no of runs   | 7                 |
| E(R)         | 3.857             |
| var(R)       | 0.884             |
| stdev(R)     | 0.940             |
| Z-value      | 1.215             |
| p-value      | 0.888             |

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 [11]. 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  $\alpha$ , 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 Buchanan-three-phase model in fitting of the growth curve of *Paracoccus* sp. SKG shows adequate statistics strength based on the diagnostics 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 show 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] 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] Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Stat Model Anal. 2011;2:21-3.
- [5] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Econ Lett. 1981;7(4):313-8.
- [6] Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames Iowa: Iowa State University Press; 1980.
- [7] 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.
- [8] Draper NR, Smith H. Applied Regression Analysis. Wiley, New York; 1981.
- [9] Rohatgi, A. WebPlotDigitizer. <http://arohatgi.info/WebPlotDigitizer/app/> Accessed June 2 2014.;
- [10] 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.
- [11] Huitema BE, McKean JW, Zhao J. The runs test for autocorrelated errors: unacceptable properties. J Educ Behav Stat. 1996;21(4):390-404.