

ASIAN JOURNAL OF PLANT BIOLOGY

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



# Testing the Normality of Residuals of the Modified Gompertz Model used for Modelling the Growth of Callus Cultures from *Glycine wightii* (Wight & Arn.) Verdc.

Shukor, M.S.<sup>1</sup>, Masdor, N.A.<sup>2</sup>, Halmi, M.I.E.<sup>3</sup>, Ahmad, S.A.<sup>4</sup> and Shukor, M.Y.<sup>1,4\*</sup>

<sup>1</sup>Snoc International Sdn Bhd, Lot 343, Jalan 7/16 Kawasan Perindustrian Nilai 7, Inland Port, 71800, Negeri Sembilan, Malaysia. <sup>2</sup>Biotechnology Research Centre, MARDI, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia

<sup>3</sup>Department of Chemical Engineering and Process, Faculty of Engineering and Build Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor.

<sup>4</sup>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: <u>yunus.upm@gmail.com</u> Tel: +603-89466722 Fax: +603-89430913

# HISTORY

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

#### **KEYWORDS**

*Glycine wightii* callus growth modified Gompertz model least square method normality test

# ABSTRACT

*Glycine wightii* species is native to Brazil and Africa. It is often known as an important climbing vine-like perennial soybean. It is in the family of Leguminosae, within the sub-family Papilionoideae, under the genus Glycine and with the sun-genus Bracteata. In a prior work, we modeled the growth of the bacterium *Glycine wightii* from published literatures to acquire crucial growth constants. We learned that the modified Gompertz 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 this assumption to be met, we perform statistical diagnosis tests such as the Kolmogorov-Smirnov, Wilks-Shapiro and D'agostino-Pearson tests.

## INTRODUCTION

Glycine wightii species is native to Brazil and Africa. It is often known as an important climbing vine-like perennial soybean [1]. It is in the family of Leguminosae, within the sub-family Papilionoideae, under the genus Glycine and with the sun-genus Bracteata. Tissue culture of in vitro cells, tissues and organs of Glycine wightii can yield efficient means in breeding genetics, understanding the physiology and biochemistry of legumes. In addition it can be utilized in the production of plant biomass, plant improvement, as a mean for studying protein synthesis, and production of secondary metabolites [2,3]. In vitro culture of Glycine wightii species has been documented from leaves [4] and cotyledons and hypocotyls [5]. The production of callus and cell culture can be an important tool to study plant regulation, biosynthesis and biochemistry [6]. One of the most important preliminary investigation of callus attributes is the growth characteristics [7]. Most often than not, callus growth curve is sigmoidal in characteristics. Frequently, plant scientists studying callus growth neglect the utilization of mathematical growth that are useful in obtaining important growth constants such as lag period, maximum specific growth rate and maximum growth or asymptote. All these constants are useful for further modelling.

We have utilized several growth models (manuscript in preparation) to model the growth of *Glycine wightii* callus from a published literature [7]. We discovered that the modified Gompertz model via nonlinear regression utilizing the least square method was the best model to describe the growth curve (manuscript in preparation). However, the use of statistical tests to choose the best model relies heavily on the residuals of the curve to be distributed normally. We performed 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

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [8] which helps to digitize scanned plots into table of data with good enough precision [9]. Data were acquired from the works of Silva et al. [7] from Figure 1 and then replotted (**Fig. 1**, with permission) (Shukor, M.S., Masdor, N.A., Shamaan, N.A., Wan Johari, W.L. and Shukor, M.Y 2015. Modelling the growth of callus cultures from *Glycine wightii* (Wight & Arn.) Verdc. Manuscript in preparation).



Fig. 1. Growth curves of *Glycine wightii* callus modelled using the modified Gompertz (MG) model.

## Normality test

Three normality tests- Kolmogorov-Smirnov [10,11], Wilks-Shapiro [12] and the D'Agostino-Pearson omnibus K2 test [13] were utilized to the residuals from the modified Gompertz 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 [14]. 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);

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

where  $y_i$  denotes the *i*<sup>th</sup> response from a given data set while  $x_i$  is the vector of explanatory variables to each set at the *i*<sup>th</sup> 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. A 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 modified Gompertz via least square method was appropriate (**Table** 1). The residuals plot does not indicate data that supported non normal distribution (**Fig.** 2).

 Table 1. Numerical normality test for the residual from the modified Gompert model.

Normality test	Analysis
D'agostino & Pearson omnibus	
K2	1.212
p value	0.5455
passed normality test (alpha=0.05)?	yes
p value summary	ns
Shapiro-Wilk	
W	0.9049
p value	0.2819
passed normality test (alpha=0.05)?	yes
p value summary	ns
Kolmogorov-Smirnov (K-S)	
KS distance	0.2401
p value	> 0.100
passed normality test (alpha=0.05)?	yes
p value summary	ns
ns: not significant	



Fig. 2. Residual plot for the modified Gompertz model.

The result indicates that the data conformed to normality (Fig. 3) based on the normal probability Q-Q plot of residuals for the modified Gompertz 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 modified Gompertz 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 [10,11]. In the Wilks-Shapiro test, a W<sup>2</sup> 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<sup>2</sup> is high, then the agreement is rejected [12]. 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 [13].

In conclusion, normality tests for the residuals used in this work has indicated that the use of the modified Gompertz model in fitting of the callus growth curve of Glycine wightii 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

Snoc International Sdn Bhd. financially supported this project.

#### REFERENCES

- Tokita N, Shimojo M, Masuda Y. Amino acid profiles of tropical legumes, cooper (*Glycine wightii*), Tinaroo (*Neonotonia wightii*) and Siratro (*Macroptilium atropurpureum*), at pre-blooming and blooming stages. Asian-Australas J Anim Sci. 2006;19(5):651–4.
- [2] Ibrahim R, Hussein S, Noordin N, Azlan E, Manan MA, Adrian H, et al. Advanced cell culture technology for essential oil production and microarray studies leading to discovery of genes for fragrance compounds in *Michelia alba* (Cempaka putih). Acta Hortic. 2008;765:95–100.

- [3] Hussein S, Ling APK, Ng TH, Ibrahim R, Paek KY. Adventitious roots induction of recalcitrant tropical woody plant, *Eurycoma longifolia*. Romanian Biotechnol Lett. 2012;17(1):7026–35.
- [4] Hammatt N, Nelson RS, Davey MR. Plant regeneration from seedling explants of perennial Glycine species. Plant Cell Tissue Organ Cult. 1987;11(1):3–11.
- [5] Pandey P, Bansal YK. Plant regeneration from leaf and hypocotyl explants of *Glycine wightii* (W. and A.) Verdc. var longicauda. Jpn J Breed Tokio. 1992;42:1–5.
- [6] Kiong ALP, Then CN, Hussein S. Callus induction from leaf explants of *Ficus deltoidea* Jack. Int J Agric Res. 2007;2(5):468– 75.
- [7] Silva ALC da, Caruso CS, Moreira R de A, Horta ACG. Growth characteristics and dynamics of protein synthesis in callus cultures from *Glycine wightii* (Wight & amp; Arn.) Verdc. Ciênc E Agrotecnologia. 2005;29(6):1161–6.
- [8] Rohatgi, A. WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer/app/ Accessed June 2 2014.:
- [9] 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.
- [10] Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G Dell' Ist Ital Degli Attuari. 1933;4:83–91.
- [11] Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat. 1948;19:279–81.
- [12] Royston P. Wilks-Shapiro algorithm. Appl Stat. 1995;44(4):R94.
- [13] D'Agostino RB. Tests for Normal Distribution. In: D'Agostino RB, Stephens MA, editors. Goodness-Of-Fit Techniques. Marcel Dekker; 1986.
- [14] 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.