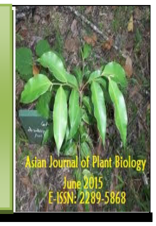
ASIAN JOURNAL OF PLANT BIOLOGY
http://journal.hibiscuspublisher.com/index.php/AJPB



Short Communication

Outlier Analysis of the Modified Gompertz Model used for Modelling the Growth of Callus Cultures from *Glycine wightii* (Wight & Arn.) Verdc.

Shukor, M.S.[1], Masdor, N.A.[2], Halmi, M.I.E.[3], Ahmad, S.A.[4] and Shukor, M.Y.[1,4]*

[1]Snoc International Sdn Bhd, Lot 343, Jalan 7/16 Kawasan Perindustrian Nilai 7, Inland Port, 71800, Negeri Sembilan, Malaysia.
[2]Biotechnology Research Centre, MARDI, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia
[3]Department of Chemical Engineering and Process, Faculty of Engineering and Build Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor.
[4]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: yunus.upm@gmail.com
Tel: +603-89466722
Fax: +603-89430913

HISTORY

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

KEYWORDS

Glycine wightii
callus growth
modified Gompertz model
least square method
Grubbs test

ABSTRACT

The production of callus and cell culture can be an important tool to study plant regulation, biosynthesis and biochemistry. Previously, we model the growth of growth of *Glycine wightii* published literature to obtain vital growth constants. We discovered that the modified Gompertz model was the best model based on statistical test such as root-mean-square error (RMSE), adjusted coefficient of determination (R^2), bias factor (BF), accuracy factor (AF) and corrected AICc (Akaike Information Criterion) to explain the callus growth. However, the use of statistical tests to choose the best model relies heavily on the residuals of the curve to be statistically robust. Usually, the residuals need be tested for the occurrence 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. The results showed that there was no outlier present, and the modified Gompertz model was adequate to model the callus growth.

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 the genetics of 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, 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.

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) [8]. 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 [9].

In this work, we demonstrated the utilization of the Grubbs's test to detect the presence of outlier in the residuals of the modified Gompert model used in modelling callus growth from the plant *Glycine wightii* (Wight & Arn.) Verdc.

METHODOLOGY

In order to process the data, the graphs were scanned and electronically processed using WebPlotDigitizer 2.5 [10] which helps to digitize scanned plots into table of data with good enough precision [11]. 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).

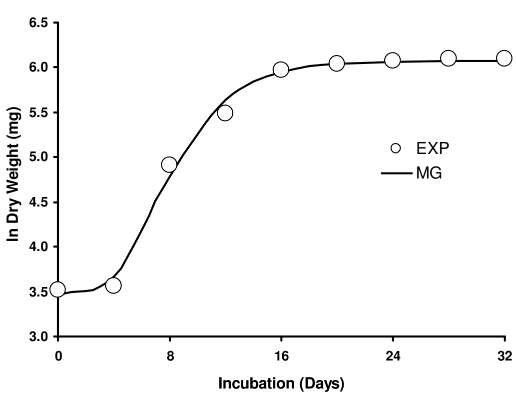


Fig. 1. Growth curves of *Glycine wightii* callus modelled using the modified Gompertz (MG) model.

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 [12]. 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**).

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

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

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

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

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

In the event that there is more than one outlier, then the ROUT method [13] 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 [14]. The residuals for the modified Gompertz model are shown in Table 1.

Table 1. Residuals for the modified Gompertz model utilized in the modelling of the callus growth of *Glycine wightii*.

Time (d)	Residuals
0	0.00
4	-0.01
8	0.05
12	-0.13
16	0.13
20	-0.03
24	-0.01
28	-0.01
32	0.00

When the Grubbs' test was applied, there was an absence of outlier for the residual data above. This indicates that the modified Gompertz was adequate to model the data. Either data distortions by a single data point 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 [12]. 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 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.

In conclusion, the Grubbs' test shows the absence of any outlier. The residuals of the modified Gompertz model were robust enough to be used in all future statistical tests such as normality, runs test, tests for homoscedasticity and presence of autocorrelation. This test is important because data distortions by a single data can lead to gross error in the fitting of a nonlinear curve.

ACKNOWLEDGEMENT

Snoc International Sdn Bhd. financially supported this project.

REFERENCES

[1] 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 & Arn.) Verdc. Ciênc E Agrotecnologia. 2005;29(6):1161–6.

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

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

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

[11] 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.

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

[13] 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.

[14] 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.