
Forecasting the Sri Lankan Population with the Gompertz and Verhulst Logistic Growth Models

Sri Lanka Journal of Economic Research
Vol. 07(1): pp 1-12
Sri Lanka Forum of University
Economists
ISSN: 2345-9913



W.A.D.M. Welagedara¹, Lakshika S. Nawarathna² &
Ruwan D. Nawarathna²

Abstract

Population growth is one of the major problems in the world. The population forecast is predicted by each and every country from an early age. Further, the population growth and the size of the country are major factors which affect its economy and policies. Hence, it has emphasized the importance of predicting the future population of Sri Lanka. Therefore, this study mainly focuses on proposing population growth models to predict the population growth of Sri Lanka. The Verhulst logistic growth model and the Gompertz growth model are used to predict the population of Sri Lanka using population data from the census population and the mid-year population, the birth rate and the death rate obtained from the Department of Registrar General of Sri Lanka from 1990 to 2018. The explicit solutions for each model are derived by using mathematical techniques of differentiation and integration. The carrying capacity; i.e., the maximum number of the population an environment can support, indefinitely was calculated using the two models. The percentages of Root Means Square Error (RMSE), Mean Absolute Percentage Deviation (MAPD) and Symmetric Mean Absolute Percentage Error (SMAPE) were calculated to measure the prediction accuracy between the actual data and the predicted data of the proposed models. Then the Sri Lankan population from 2019 to 2048 was predicted using the proposed models. Results show the prediction accuracy of Gompertz model to be higher compared to the one of Verhulst logistic growth model. This study provides a deep insight into the population prediction in Sri Lanka, a country with limited resources, and in an area teeming with conflicts.

Keywords: *Carrying capacity, Gompertz model, Growth rate, Population prediction, Verhulst Logistic growth model*

¹ Postgraduate Institute of Science, University of Peradeniya, Sri Lanka.

² Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.

Introduction

The growth of the population and the size of the population of a country are major factors for its economy and its policy. Most of the developed countries usually change their policies according to their population predictions. Many countries consider the consequences of the human population growth for its social and economic development (Abayasekara 1985; Abeykoon 2011; Department of Census & Statistics, 2010). Population forecasting is essential for all long-term planning and for the provision of services of a country. Therefore, developing a model for forecasting population growth will help a country to estimate the future size and the structure of the population. It is also important to interpret and set parameters for Sri Lanka's future population development in line with the framework for global shared socio-economic pathways. Mathematical models play a major role to predict the population growth of the world (Eberhardt and Breiwick 2012; Terano 2017). Moreover, mathematical models take many forms to predict the population growth, such as dynamical systems, statistical models, models with differential equations etc. There are many factors associated with the population growth such as the food supplies, available land, technology, birth rate, death rate and emigration, prevailing conditions in the country like war, etc. (Fernando 1991). The population will fluctuate due to the above-mentioned factors but will grow exponentially.

Several researchers in numerous fields have recently become inquisitive about finding out quantitative models of population growth (Shair, Purcal and Parr 2017). Once it is modelled, it can be utilized for planning population programs, either for controlling growth or dispersing population for a balanced distribution. However, population predictions are an inevitable tool for decision-makers and planners. The government officials, particularly health, education, transport, environment, social welfare and housing, constantly seek predictions of future demographic parameters for planning purposes and resource allocation (Smith 1977). The most common model used to estimate growth in populations with overlapping generations is the Verhulst–Pearl logistic equation in which the population growth stops at the carrying capacity. In this study, some phenomenological growth models based only on the population information are deduced in an intuitive way. These models, for instance, Verhulst and Gompertz models are introduced in such a way that all the parameters involved have a physical interpretation (Ausloos 2013; Ribeiro 2016). The objectives of this study are to propose population growth models to predict the population growth of Sri Lanka, to predict the population by Verhulst and Gompertz models and to estimate the carrying capacity of Sri Lanka.

This article is organized as follows. In section 2, the theoretical aspects of the considered population models are described. In section 3, the analysis carried out by the real data set is presented. To sum up, in section 4, the obtained results are discussed.

The conclusion states the main findings with direct comparison against the one previous study conducted in Sri Lanka and against literature, published worldwide. It further discusses the limitations of the findings for readers' reference.

Models for Forecasting Population

Under this section, two growth models which are used to predict the population of Sri Lanka would be discussed. The final model equations are obtained by solving the differential equations in order to predict the population of Sri Lanka utilizing predated population data from the census population and the calculated mid-year populations, the birth rate and the death rate obtained by the Department of Registrar General of Sri Lanka from 1990 to 2018. In order to check the accuracy and select the best model, three measures of the forecasted errors namely Root Mean Square Error (RMSE), Mean Absolute Percentage Deviation (MAPD) and Symmetric Mean Absolute Percentage Error (SAPE) are discussed.

The Verhulst Logistic Growth Model

The population growth model for the population P is $\frac{dP}{dt} = (b - d)P$, where b = birth rate, d = death rate, $(b-d)$ = per capita growth rate (r). The solution of the Malthus differential equation model by the separation of variable becomes $P(t) = P_0 e^{rt}$, where $P(t)$ = the population at time t , P_0 = the initial population, $r = (b - d)$. Assuming that the environment has an intrinsic carrying capacity, K , Verhulst (1838) proposed

$$\frac{dP}{dt} = (b - d)P \cdot f(P)$$

where $f(P)$ is the population growth function which should satisfy that $f(0) = 1$, that is the population grows exponentially with the growth rate $(b - d)$ when P is small (Bhowmick, Chattopadhyay and Bhattacharya, 2014). This is called the Malthus exponential growth model. The function should also satisfy that $f(K) = 0$, that is the population growth becomes zero when the population reached its maximum, which is called the carrying capacity K of the environment. Further, $f(P) < 0$ for $P > K$, that is the population should decrease when its size exceeds the carrying capacity of the environment. To satisfy all these conditions, the simplest population growth function ($f(P)$) can be defined as $f(P) = \left(1 - \frac{P}{K}\right)$.

Then the Verhulst' population growth model can be written as $\frac{dP}{dt} = rP \left(1 - \frac{P}{K}\right)$ where

$r = b - d$. To solve this model, the separation of variables method was used and $\int \frac{dP}{P(1 - \frac{P}{K})} = \int r dt$ was observed. Resolving into partial fractions $\int \frac{dP}{P} + \int \frac{dP}{K - P} = \int r dt$ was observed. Solving this differential equation and substituting the initial condition, t

$= 0$, $P = P_0$, finally Verhulst's population growth model was obtained to predict the population (P) at time as

$$P(t) = \frac{KP_0e^{rt}}{K + P_0(e^{rt} - 1)}$$

As $t \rightarrow \infty$, it is observed that $P(t) = K$, the maximum population is called the carrying capacity of the environment (Kucharavy and Guio 2008). The carrying capacity was calculated which has directly derived from the model.

$$K = \frac{P(t) \times P_0(e^{rt} - 1)}{P_0e^{rt} - P(t)}$$

The Gompertz Growth Model

The Gompertz function is a type of mathematical model for a time series and it is a sigmoid function which describes growth as being slowest at the start and end of a given time period. This is a special case of the generalized logistic function given by

$$\frac{dN}{dt} = rN \ln\left(\frac{K}{N}\right) \quad (2.1)$$

where $N=N(t)$ is the population at time t , r is the constant intrinsic growth of population, with $r>0$, K is the carrying capacity of the population (Bhowmick, Chattopadhyay and Bhattacharya 2014).

By solving the above first order differential equation (2.1), an equation for the population which varies by time can be obtained. In order to do this, a replacement of variables were performed, considering that,

$$v = \ln\left(\frac{N}{K}\right) \quad (2.2)$$

Then the exponential function on both sides of the equation (2.2) were applied and, obtained, $e^v = \frac{N}{K}$. Then, multiplying the equation (2.2) for K the following equation was obtained.

$$N(t) = K \cdot e^v \quad (2.3)$$

Using the chain rule, it is concluded that

$$\frac{dN}{dt} = K \cdot e^v \frac{dv}{dt} \quad (2.4)$$

Taking the equation (2.3) in equation (2.4)

$$\frac{dN}{dt} = N \frac{dv}{dt} \quad (2.5)$$

Of course, equation (2.2) can be written as

$$v = -\ln \frac{K}{N} \quad (2.6)$$

Also, replace the equation (2.1) to (2.6) and obtain a new relation.

$$\frac{dN}{dt} = -v \cdot r \cdot N \quad (2.7)$$

Then match the equations (2.5) and (2.7) to find

$$\frac{dv}{dt} + r \cdot v = 0 \quad (2.8)$$

Now the equation (2.8) is solved by the method of integrating factor, noting that it is given by $P(t) = e^{rt}$. Multiplying the equation (2.8) by the integrating factor, $\mu(t)$,

$$\frac{dv}{dt} \cdot e^{rt} + r \cdot v \cdot e^{rt} = 0 \Rightarrow \frac{d}{dt}(v \cdot e^{rt}) = 0 \Rightarrow (v \cdot e^{rt}) = c$$

where c is an arbitrary constant. Thus, the general solution of the equation (2.5) was obtained.

$$v = ce^{-rt} \quad (2.9)$$

Then, that equals the equations (2.2) and (2.9),

$$\ln \left(\frac{N}{K} \right) = ce^{-rt} \Rightarrow \frac{N}{K} = e^{ce^{-rt}}$$

which means that the function $N(t)$, which represents the population is given by

$$N(t) = K \cdot e^{ce^{-rt}} \quad (2.10)$$

Once solved the Gompertz equation, considering an initial condition that points out which one was the population at the beginning of the analysis in a way that an Initial Value Problem, i.e., one differential equation associated to an initial condition. Let then, the initial condition $N(0) = n_0$. Applying this condition to the Equation (2.10),

$$n_0 = K \cdot e^{ce^{-r \cdot 0}} \Rightarrow e^c = \frac{n_0}{K} \Rightarrow c = \ln \left(\frac{n_0}{K} \right) \quad (2.11)$$

Therefore, the initial value problem which associated differential equation is the Gompertz equation, that has been attained as the solution to the time function, $N(t) = K e^{e^{-rt} \ln \left(\frac{n_0}{K} \right)}$.

In possession of the Gompertz equation and its analytical solution, considering a given population, this study can be performed based on the population of a country. Considering that, when the time grows, the population, $N(t)$, tends to increase. It is clear that it tends to approximate to the carrying capacity, K , of the population (Sergio 2012; Tjorve 2017). The stochastic version of the Gompertz model presented here assumes that the stochasticity arises solely from environmental (process) noise. i.e.,

$\ln [N_{t+1}/N_t] = r + F$, where $F \sim \text{Normal}(0, \sigma^2)$ or equivalently $\ln[N_{t+1}/N_t] \sim \text{Normal}(r, \sigma^2)$.

Under the assumption, there is a constant linear decrease in the growth rate (r) as population size increases. It is easy to adapt this for density dependence.

$$\ln[N_{t+1}/N_t] \sim \text{Normal}(r_{\max} + bN_t, \sigma^2)$$

This is the Stochastic Gompertz (Logistic) Model where r_{\max} and b are estimated parameters and b measures the magnitude of intraspecific competition. This allows us to predict next year's population size, but it is always dependent on the previous year's abundance.

$$\ln[N_{t+1}] = \ln[N_t] + r_{\max} + bN_t + F.$$

This equation is the same for a linear model of which the parameters can be estimated using linear regression.

The response variable of the model is $\ln[N_{t+1}/N_t]$, whereas the predictor variable is N_t . This can be performed using any mathematical/statistical software by plotting the natural logarithm of our observed instantaneous growth rates, $\ln[n_{t+1}/n_t]$, against observed population sizes n_t . Then, a linear trend was fitted to this plot and the equation was displayed. The Y-intercept is the estimate of r_{\max} and the slope is the estimate of b . If the estimate of b is negative, there is a 'negative density dependence' meaning that the growth rate decreases with the increasing abundance. For the Gompertz model, r_{\max} is the growth rate when the population size equals 1. Interestingly, under the Gompertz model, the growth rate goes to infinity as the population size goes to 0. Estimates of r_{\max} and b by regressing $\ln[N_{t+1}/N_t]$ against $\ln[N_t]$ was obtained. Under the Gompertz model, our estimate of carrying capacity K is $K = \text{Exp}\left[-\frac{r_{\max}}{b}\right]$.

Model Validation

A method of describing the uncertainty in population forecast is based on the calculation of forecast error measures. The three measures namely, Root Means Square Error (RMSE), Mean Absolute Percentage Deviation (MAPD), and the Symmetric Mean Absolute Percent Error (SMAPE) were used for the validation of the fitted models (Chai and Draxler 2014; Gary 2017; Willmott and Matsuura 2005).

Root Means Square Error (RMSE)

$$\text{Percentage RMSE} = \left(\sqrt{\frac{\sum_{i=1}^n (V_{\text{assign}}^2 - V_{\text{observe}}^2)^2}{n}} \right) * \left(\frac{n * 100}{\sum_{i=1}^n V_{\text{observe}}^2} \right)$$

Mean Absolute Percentage Deviation (MAPD)

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

where A_t, F_t, n are the values of actual population, predicted population and the number of the considered observations respectively.

Symmetric Mean Absolute Percentage Error (SMAPE)

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{|A_t| + |F_t|}$$

where A_t, F_t, n represent the value of actual population, the value of the predicted population and number of the considered data.

Empirical Analysis and Forecasts

The analysis is carried out with the census of Sri Lankan population data from 1990 to 2018. The parameter carrying capacity of the Verhulst Logistic growth model and the Gompertz growth model have been estimated considering the data from 1990 to 2010 using the two separate equations derived above.

Table 1 shows the estimated parameters for both Verhulst Logistic growth and Gompertz growth models. The carrying capacity of the Verhulst Logistic growth model is higher than that of Gompertz growth model (Bhowmick, Chattopadhyay and Bhattacharya 2014).

Table 1. Estimated parameters of both Verhulst Logistic growth model and Gompertz growth model

Model Name	Carrying Capacity	Initial Population	Growth rate
Verhulst Logistic growth model	59178427.45	17015000	0.0142
Gompertz growth model	26107639.12	17015000	0.0142

The population was predicted by the Verhulst Logistic growth model and the Gompertz growth model from 2011 to 2018 and were compared with the actual values. Figure 1 shows the actual and predicted population of Sri Lanka from using both models. A convergence can be observed in the two populations. Further, there was a drop of the actual population while the predicted values were gradually increased in 2012. The

increasing rate of predicted population in Verhulst Logistic growth model is higher than that of the Gompertz model.

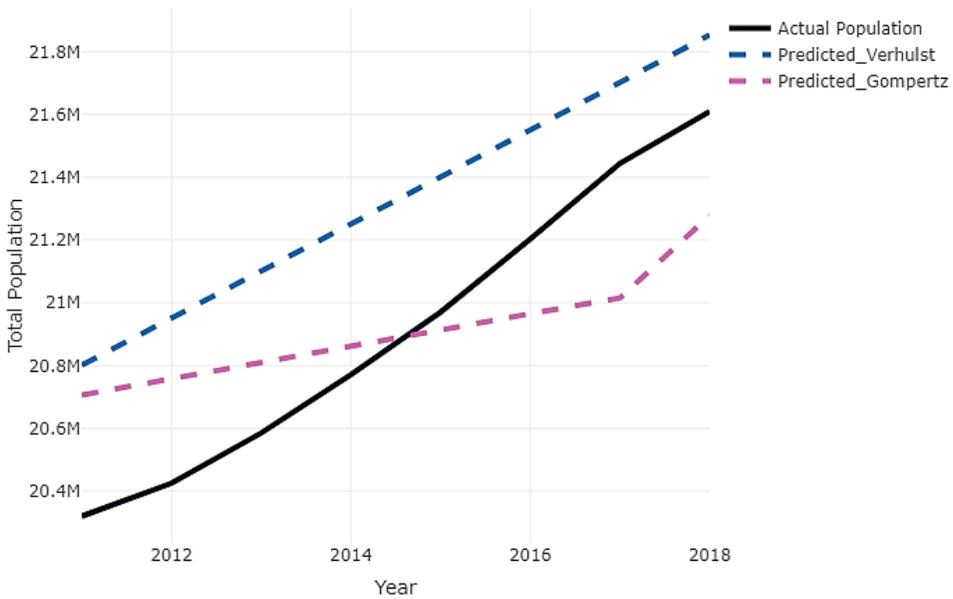


Figure 1. Actual and predicted population of Sri Lanka using both the Verhulst model and Gompertz growth model.

In order to get an idea of the accuracy of the Verhulst Logistic model and the Gompertz growth model in the case of population forecasts, these models have been identified, estimated, and used for forecasting for various years. The forecasts were made and compared with the actual population, and subsequently, the various error measures were calculated.

Percentage of errors and their accuracy for the fitted models are calculated and are tabulated in Table 2. These values are beneficial in model validation and to finalize which model is the most appropriate for our future predictions (Sargent 2011).

Table 2. Percentage of error and their accuracies for the fitted models.

Model	RMSE (%)		MAPD (%)		SMAPE (%)	
	Value	Accuracy	Value	Accuracy	Value	Accuracy
Verhulst Logistic	3.902	96.098	1.817	98.183	0.899	99.101
Gompertz	2.428	97.572	1.062	98.938	0.532	99.468

Under these obtained error values and the model accuracies, the Gompertz growth model can be more appropriate to forecast the population with a higher accuracy than the Verhulst Logistics growth model (Pierre, 2014).

Using both the Verhulst Logistic growth model and the Gompertz growth model, taking the initial population as the population of the year 2018 and the growth rate for the year 2018, the population from 2019 to 2048 were forecasted and were plotted as in Figure 2.

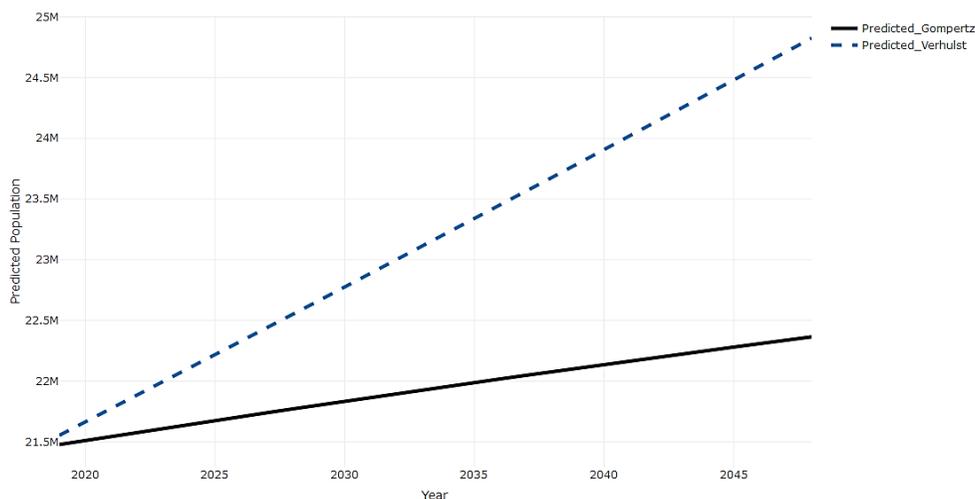


Figure 2. Predicted population of Sri Lanka by Verhulst and Gompertz models from 2019 to 2048.

The two populations predicted from both growth models are diverging yearly. When year 2048 is considered, the population predicted from the Verhulst growth model is higher than that of the Gompertz model. Further, the angle of inclination of each graph is higher in the Verhulst model compared to the Gompertz model which represents the rate of growing population. For the growth of the population of a country, not only the birth rate and death rate would affect but also the emigration. Therefore, a theoretical sense from these predictions can be obtained and hence can be used for the future planning of the country.

Discussion and Conclusion

In this study, Verhulst Logistic growth model and the Gompertz growth model were proposed for predicting population growth of Sri Lanka and to examine the pattern of the country's population growth in the long run (Zabadi, Assaf and Kanan 2017). Next, two growth models were compared to find the most appropriate growth model to forecast the population (Chai and Draxler 2014; Gary, 2017; Willmott and Matsuura, 2005). All the predictions were made assuming that the prevailing conditions in the country would be affecting the population growth, such as political, economic, social impact, poverty, environment, and war which remained unchanged during the period.

Subsequently, predictions for the future population of Sri Lanka were calculated from 2019 to 2048 using the proposed models. A huge deviation among the two plots was visible and the predicted values of the data set of the Verhulst model were higher than that of the Gompertz model. It is observed that the main reason for this was the different carrying capacity values. Therefore, in order to improve the accuracy of theoretical carrying capacity value, a dataset with many observations should be used. When the model accuracies are compared, the Gompertz growth model has a higher accuracy of all the three validation methods than the Verhulst Logistic growth model (Berger, 1980). Therefore, the proposed Gompertz growth model is more appropriate for predicting the human population. The proposed model provides a good fit to the population data which was obtained from the Department of Census & Statistics, Sri Lanka. Since the carrying capacity parameter of the proposed model is well accurate theoretically, the proposed model can be used to predict the carrying capacity for a given year.

Acknowledgement

The authors are grateful to the Department of Census & Statistics, Sri Lanka for providing the dataset.

Conflict of Interest

The author declares that there is no conflict of interest.

References

- Abayasekara, A.W.A.D.G. 1985. Population Growth and Redistribution in Sri Lanka, 1901-1981. Vol 3 of Population Redistribution and Development in South Asia.pp.139-154.
- Abeykoon, A.T.P.L. 2011. The demographic transition: Opportunities and Challenges-The case of Sri Lanka. Institute for Health Policy. 27th Asian Parliamentarians Meeting on Population and Development.
- Annual Health Bulletin. 2003. Sri Lanka Demographic and Health Survey 1993 & 2000.
- Ausloos, M. 2013. Gompertz and Verhulst frameworks for growth and decay description. *Int. J. Comput. Anticip. Syst.* 30 (2014) 15- 36.
- Berger, R.D. 1980. Comparison of the Gompertz and Logistic Equations to Describe Plant Disease Progress. Journal Series Paper 2471 of the Florida Agricultural Experiment Station. Vol 71, No.7, 716-719.
- Bhowmick, A.R., Chattopadhyay, G. and Bhattacharya, S. 2014. Simultaneous identification of growth law and estimation of its rate parameter for biological growth data: a new approach. *Journal of Biological Physics.* 40(1): 71–95.

- Chai, T. and Draxler, R.R. 2014. Root Mean Square Error (RMSE) or mean absolute error (MAE)-Arguments against avoiding RMSE in the literature. *Geosci. Model Dev*, 7, 1247-1250, <https://doi.org/10.5194/gmd-7-1247-2014>, 2014.
- Department of Census & Statistics. 2010. *Bulletin of International Migration Statistics of Sri Lanka*. Ministry of Finance & Planning. 2012-6832.
- Eberhardt, I.L., and Breiwick, J.M. 2012. *Models for Population Curves*. International Scholarly research Network. vol. 2012, Article ID 815016, 7 pages.
- Fernando, N. 1991. Demographic situation of Sri Lanka. *Asian Population and Development Association*, 77-86
- Gary, B. 2017. Mean absolute error and root mean square error: which is the better metric for assessing model performance? . *Geophysical Research Abstracts*. Vol 19, EGU2017-3574.
- Hategekimana, D., Joseph K., Mung'atu, Ndengo, M. 2016. Determining Population projections using Leslie, Lee Carter and ARIMA models: A case study of Rwandan Population Dynamics. *International Journal of Social Science and Humanities Research*. Vol. 4, Issue 2, pp: (144-159).
- Hussein, S., Halmi, M.I.E. and Ling, A.P.K. 2017. The modified gompertz model demonstrates a variable growth rate between two centellaasiatica phenotypes. *Journal of Biochemistry, Microbiology and Biotechnology*. 5, 18-20.
- Kan, H.K. 2012. *A Bayesian Mortality Forecasting Framework for Population and Portfolio Mortality*. MSc Thesis, University of Amsterdam.
- Kucharavy, D. and Guio, R.D. 2008. Logistic substitution model and technological forecasting. *TRIZ Future 2008 - Synthesis in Innovation*. 65-73.
- Pierre, N. 2014. A simple selection test between the Gompertz and Logistic Growth models. *Technological Forecasting and Social Change*, Elsevier, vol. 88(C), pages 98-105.
- Ribeiro, F.L. 2016. An attempt to unify some population growth models from first principles. *Revista Brasileira de Ensino de Física*, vol. 39, nº 1, e1311 (2017).
- Rodriguez, G. 2007. *Poisson Models for Count Data*. Academic press; Chapter 4.
- Sargent, R.G. 2011. Verification and validation of simulation models. *IEEE*. 978-1-4577-2109-0
- Sergio, D.J. 2012. Gompertz Model: Resolution and Analysis for Tumors. *Journal of Mathematical Modeling and Application*. Vol 1, No.7.
- Shair, S., Purcal, S., and Parr, N. 2017. Evaluating extensions to Coherent Mortality Forecasting Models. *Risks* 2017,5-16.

- Smith, D.A. 1977. Human Population Growth: Stability or Explosion?. *Mathematics Magazine*, Vol 50, No: 4
- Terano, H.J.R. 2017. Analysis of mathematical models of population dynamics applied to Philippine population growth. *Far East Journal of Mathematical Sciences (FJMS)*. Vol 103, No:3, 2018, Pages 561-571.
- Tjorve, K.M.C. and Tjorve E. 2017. The use of Gompertz models in growth analyses and new Gompertz-model approach: An addition to the Unified-Richards family. *Inland Norway University of Applied sciences. PLOS ONE* 12(6): e0178691. <https://doi.org/10.1371/journal.pone.0178691>
- Willmott, C.J., Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*. 30, 79-82.
- Zabadi, A.M, Assaf, R., Kanan, M. 2017. A Mathematical and statistical approach for predicting the population growth. *World Wide Journal of Multidisciplinary Research and Development*. 3(7), 50-59.