An Improved Logistic Growth Population Model

An approach to help ecologists improving population models

Introduction

Since I returned from my Seed of Hope program in Philippines, the poor lives there have obsessed me all the time. Hoping to do something for the poor people, I have done quite a lot of study and agree with “The Borgen Project”, a nonprofit organization to fight extreme poverty,” Population growth and poverty present the classic “chicken or egg” dilemma.

Population control is crucial to prevent from further overpopulation in the Third World. Essentially, our society needs an appropriate tool or model to accurately evaluate and estimate the effectiveness of any control plan. The prevailing one is the logistic growth model. It must be constructive to the world if I could explore a way to enhance the estimation accuracy on this model or those inherited from it. Since China has the world's largest population (1.42 billion), followed by India (1.35 billion), they will be the good work examples of my exploration; we could also learn the effectiveness of their population control.

Research

Estimation accuracy matters for governmental control plans. It relies on good mathematical models which could, with parameters sophisticatedly tuned, fit well with the collected population data.

Growth or decay is a common occurrence in biology, and we express the rate of growth or decay as $dP(t)/dt$, where $P(t)$ is the size of population at time t. Due to environmental trends and fluctuation, no biologist believes that one equation would suit all growth processes. Some simple models that are expressed solely in terms of differential equations of the first order that ignore random fluctuations in the population. Such models are referred to as deterministic and they are written as

$$\frac{dP(t)}{dt} = G(t, P(t)) \qquad \text{---Eq.1}$$

,where P(t) is the size of population at time t and G is the function to be figured out to match ecological observation. There are two basic trends in simplifying G function and making the equation either autonomous differential or separable one:

$$\frac{dP(t)}{dt} = F\big(P(t)\big) \qquad \text{----Eq.2}$$

or

$$\frac{dP(t)}{dt} = F_t(t) \cdot F_p(P(t)) \qquad \text{----Eq.3}$$

, where F, F_t, F_p are the functions ecologists have been trying to derive from their observations.

Thomas Malthus (1766--1834) was one of the first to note that populations grow with a geometric pattern i.e., $F\big(P(t)\big) = rP(t)$ in Eq.2, while contemplating the fate of humankind (Dobrushkin, 2015). He asserted

$$\frac{dP(t)}{dt} = rP(t) \qquad \text{---Eq.4}$$

, where $P(t)$ is the size of population at time t and r is the rate of change. (Dobrushkin, 2015)

The Malthus model leads to exponential (unbounded) growth,$P(t) = P(0)e^{rt}$,which is clearly not viable in the long-term run. Pierre François Verhulst, a Belgian mathematician, assumed that all populations are prone to suffer natural inhibitions in their growths, with a tendency to reach a steady limit value as time increases. By modifying T. Malthus' $F\big(P(t)\big) = rP(t)$ to be $F\big(P(t)\big) = rP(t)(1 - \frac{P(t)}{K})$ used for Eq.2, P.F. Verhulst published his equation to model population growth under limited resource in 1838:

$$\frac{dP(t)}{dt} = rP(t)\left(1 - \frac{P(t)}{K}\right) \qquad \text{---Eq.5}$$

, where $P(t)$ is the size of population at time t, r is the rate of change and K is the maximum population size, called carrying capacity, which can be supported by the environment (Dobrushkin, 2015)

I resolved P.F. differential equation as below to get the solution:

$$\frac{dP(t)}{dt} = rP(t) \cdot \left(1 - \frac{P(t)}{K}\right)$$

$$\Rightarrow \frac{1}{P(t) \cdot \left(1 - \frac{P(t)}{K}\right)} dP(t) = rd(t)$$

$$\Rightarrow [\frac{1}{P(t)} + \frac{1}{K - P(t)}] dP(t) = rd(t)$$

$$\Rightarrow \int [\frac{1}{P(t)} + \frac{1}{K-P(t)}]dP(t) = \int rd(t) \ + C, \ where \ C \ is \ a \ constant$$

$$\Rightarrow ln[P(t)] - ln[K - P(t)] = rt + C$$

$$\Rightarrow ln\left[\frac{P(t)}{K-P(t)}\right] = rt + C$$

$$\Rightarrow \frac{P(t)}{K-P(t)} = e^{rt+C}$$

$$\Rightarrow \frac{K-P(t)}{P(t)} = e^{-(rt+C)}$$

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

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

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

$$\Rightarrow P(t) = \frac{K}{1+e^{-r(t-t_0)}} \quad where \ t_0 = -\frac{C}{r}$$

t_o is a constant determined by the initial condition. It is in line with the solution from professor Strang (Strang, 2019)

However, getting a solution of any modified logistic growth model involves sophisticated skills to resolve a variant from the original differential equation. I barely found successful attempts except the paper from Yao-Zheng (Zheng, 2006).

Assistant Professor Yao-Zheng from Northern Illinois University asserted she could leverage power exponent models which are widely used in scientific and engineering. By making $F\big(P(t)\big) = rP(t) \cdot (K - P(t))^{\alpha}$ used for the autonomous differential equation (Eq.2) of deterministic model, she built a variant from logistic growth model to be,

$$\frac{dP(t)}{dt} = rP(t) \cdot (K - P(t))^{\alpha} \qquad \text{---Eq.6}$$

where P(t) represents the population at time t, r is the intrinsic growth rate and K is the carrying capacity; α is a constant with which she tries to reflect complex circumstance. She claimed to have the solution of this equation in analytic form. However, even for integer α, she was not able to present a solution consisting of fundamental functions but only able to simplify the differential equation to another equation not calling for differential operation, as shown below. (*Figure.1*). It must be way more complicated for fractional α , which she does not even mention in her paper.

$$ln\frac{P(t)}{K-P(t)}+\frac{K}{K-P(t)}=rK^2(t-t_o)+\frac{K}{K-P_o}+ln\frac{P_o}{K-P_0}\quad for\ \alpha=2$$

$$\frac{1}{K^3}ln\frac{P(t)}{K-P(t)}+\frac{1}{2K(K-P(t))^2}+\frac{1}{K^2(K-P(t))}=r(t-t_o)+\frac{1}{K^3}ln\frac{P_o}{K-P_0}+\frac{1}{2K(K-P_0)^2}++\frac{1}{K^2(K-P_0)}\quad for\ \alpha=3$$

Figure.1 Yao-Zheng's solution expression for her power exponent model

To resolve the equation shown in *Figure 1*, it calls for a graphic calculator to approach the solution by plotting the left and right functions respectively and then finding the intersection point. We could only work out a point solution at a time and then list the point solutions one by one. It's difficult to infer the continuous fitting curve.

The difficulty to resolve differential equations detains a scientist from making progress in his research if he does not come from an advanced math background.

Aim

The aim is to make an approach to difficult differential equation of logistic growth model variants and help ecologists propose model improvement.

I will leverage a way, Point-by-Point method, which I developed on my own for resolving N-order differential equations before (Wu, 2020). Without the necessity to get a solution in an analytic form, it becomes easy to handle any variety of logistic growth models in form of $\frac{dP(t)}{dt}=G(t,P(t))$.

Introduction on Point-by-Point Method

From first principles, the derivative at $x = x_i$ is

$$y'(x_i) = \frac{dy(x)}{dx}|_{x=x_i} = \lim_{h\to 0}\frac{y(x_i+h)-y(x_i)}{h}$$

Equivalently, we have

$$y'(x_i) = \frac{y(x_i+h)-y(x_i)}{h}$$

where h is an infinitely small change as used in first principles formula.

After rearrangement, it tells us:

$$y(x_i+h) = y(x_i) + h\cdot y'(x_i)$$

When we apply it to resolve a differential equation,$y'(x) = G\big(y(x)\big)$,
where G is any given function,
it becomes

$$y(x_i+h) = y(x_i) + h\cdot G(y(x_i))$$

If $y(x_o)$ is known as initial condition, then setting $x_i = x_o$ and by

$y(x_i+h) = y(x_i) + h\cdot G\big(y(x_i)\big) \xrightarrow{x_i=x_0} y(x_0+h) = y(x_0) + h\cdot G(y(x_0))$

, we could know the value of $y(x_0+h)$ from the given $y(x_0)$

We can express any x in domain by $x = x_0 + kh$ where h is infinitely small increment as previously described. Below is the procedure to get $y(x)$.

For x_i to increase by h to be $x_i = x_o + h$, by

$y(x_i+h) = y(x_i) + h\cdot G(y(x_i)) \xrightarrow{x_i=x_0+h} y((x_0+h)+h) = y(x_0+h) + h\cdot G(y(x_0+h))$

$\rightarrow y(x_0+2h) = y(x_0+h) + h\cdot G(y(x_0+h))$

, we could know the value of $y(x_0+2h)$ from the given $y(x_0)$ and then learned $y(x_0+h)$

For x_i to keep increasing by another h to be $x_i = x_o + 2h$, by

$y(x_i+h) = y(x_i) + h\cdot G(y(x_i)) \xrightarrow{x_i=x_0+2h} y((x_0+2h)+h) = y(x_0+2h) + h\cdot G(y(x_0+2h))$

$\rightarrow y(x_0+3h) = y(x_0+2h) + h\cdot G(y(x_0+2h))$

, we could know the value of $y(x_0 + 3h)$ from the given $y(x_0)$ and the learned $y(x_0 + h)$ and $y(x_0 + 2h)$

………

For x_i to keep increasing by another h to reach $x_i = x_o + (k-1)h$, by

$$y(x_i + h) = y(x_i) + h \cdot G(y(x_i))$$

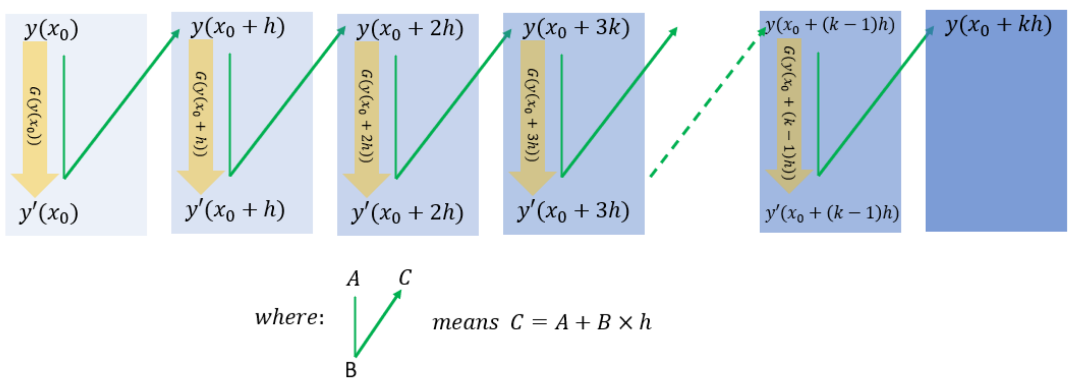
$$\xrightarrow{x_i = x_0 + (k-1)h} y((x_0 + (k-1)h) + h) = y(x_0 + (k-1)h) + h \cdot G(y(x_0 + (k-1)h))$$

$$\rightarrow y(x_0 + kh) = y(x_0 + (k-1)h) + h \cdot G(y(x_0 + (k-1)h))$$

, we could know the value of $y(x_0 + kh)$ from the given $y(x_0)$ and then learned $y(x_0 + h)$, $y(x_0 + 2h)$,…. , $y(x_0 + (k-1)h)$

Being able to know $y(x_0 + kh)$ for a very small h and any integer k equivalently means that we could learn $y(x), \forall x \; in \; domain$ because we already express x as $x = x_0 + kh$

Below is a picture I depicted to visualize the process, starting from initial condition.



The demonstrated process could be generalized to resolve any *N-order differential* equation.

With the given $y^{(0)}(x_0), y^{(1)}(x_0), y^{(2)}(x_0), \cdots, y^{(n-1)}(x_0)$ and well-defined function G, below is the generalized procedure of the Point-by-Point method,

Point-by-Point Method

Given differential equation,

$$y^{(n)}(x) = G(y^{(0)}(x), y^{(1)}(x), y^{(2)}(x), \cdots, y^{(n-1)}(x))$$

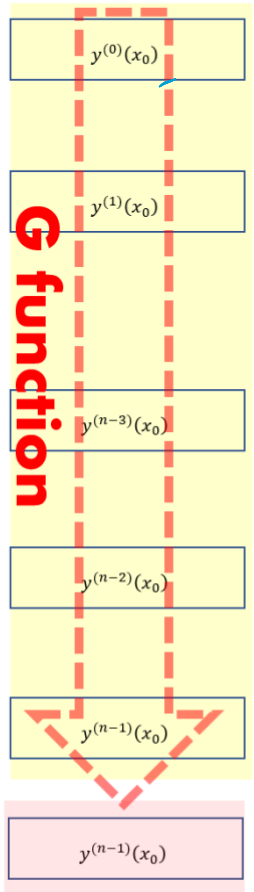
, and initial condition, $y^{(0)}(x_0), y^{(1)}(x_0), y^{(2)}(x_0), \cdots, y^{(n-1)}(x_0)$,

, below is the procedure to get the solution which consists of the target

$y(x)$ and those $y(x_i), \forall x_i \in (x_0, x)$

Step#1: Get $y^{(n)}(x_0)$ by

$$y^{(n)}(x_0) = G(y^{(0)}(x_0), y^{(1)}(x_0), y^{(2)}(x_0), \cdots, y^{(n-1)}(x_0))$$



Step#2: Get $y^{(0)}(x_0+h), y^{(1)}(x_0+h), y^{(2)}(x_0+h), \cdots, y^{(n-1)}(x_0+h)$ by

$$y^{(0)}(x_0+h) = y^{(0)}(x_0) + y^{(1)}(x_0) \times h$$

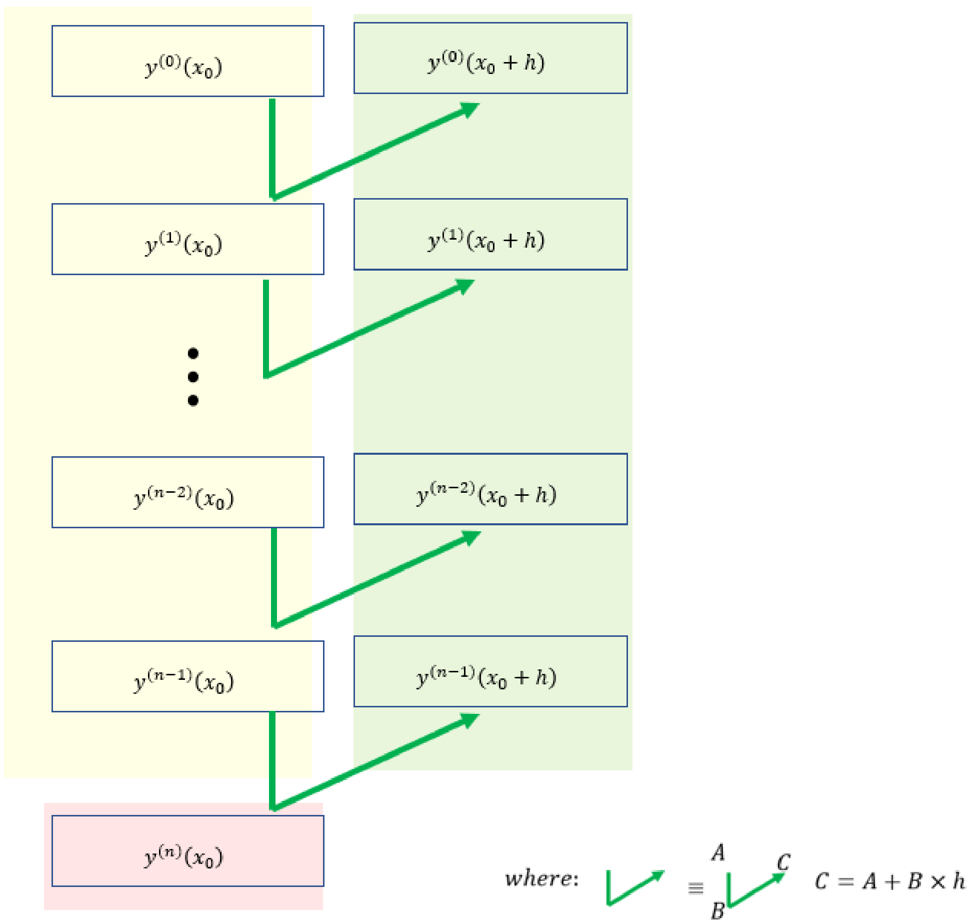
$$y^{(1)}(x_0+h) = y^{(1)}(x_0) + y^{(2)}(x_0) \times h$$

$$y^{(2)}(x_0+h) = y^{(2)}(x_0) + y^{(3)}(x_0) \times h$$

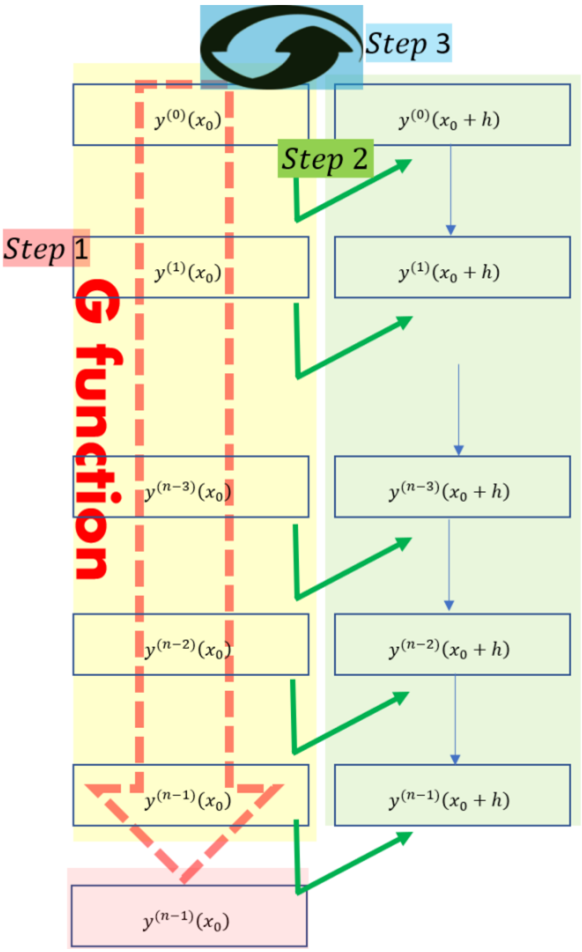
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$y^{(n-2)}(x_0+h) = y^{(n-2)}(x_0) + y^{(n-1)}(x_0) \times h$$

$$y^{(n-1)}(x_0+h) = y^{(n-1)}(x_0) + y^{(n)}(x_0) \times h$$



Step#3: Until $x = x_0$, assign x_0 with the new value $(x_0 + h)$. Go to Step#1



Step 3
$y^{(0)}(x_0)$
$y^{(0)}(x_0 + h)$
Step 2
Step 1
$y^{(1)}(x_0)$
$y^{(1)}(x_0 + h)$
G function
$y^{(n-3)}(x_0)$
$y^{(n-3)}(x_0 + h)$
$y^{(n-2)}(x_0)$
$y^{(n-2)}(x_0 + h)$
$y^{(n-1)}(x_0)$
$y^{(n-1)}(x_0 + h)$
$y^{(n-1)}(x_0)$

Let us take the logistic growth equation as the work example:

$$P'(t) = r \cdot P(t) \cdot \left(1 - \frac{P(t)}{K}\right)$$

With $P(0)$ given as the initial condition, we know:

$$P'(h) = r \cdot P(0) \cdot \left(1 - \frac{P(0)}{K}\right)$$

Then we can proceed to know $P(h)$ by

$$P(h) = P(0) + h \cdot P'(0) = P(0) + r \cdot P(0) \cdot (1 - \tfrac{P(0)}{K}).$$

With $P(h)$ acquired, we learn by the given differential equation:

$$P'(h) = r \cdot P(h) \cdot \left(1 - \frac{P(h)}{K}\right)$$

Then we can proceed to know $P(2h)$ by

$$P(2h) = P(h) + h \cdot P'(h) = P(h) + r \cdot P(h) \cdot (1 - \tfrac{P(h)}{K}).$$

With $P(2h)$ acquired, we learn by the given differential equation:

$$P'(2h) = r \cdot P(2h) \cdot \left(1 - \frac{P(2h)}{K}\right)$$

Then we can proceed to know $P(3h)$ by

$$P(3h) = P(2h) + h \cdot P'(2h) = P(2h) + r \cdot P(2h) \cdot (1 - \tfrac{P(2h)}{K}).$$

The way could be iteratively applied to the target time $t = kh$ to learn $P(t)$.

Proposal for an improved logistic growth model

To address the practical need for better accuracy and also demonstrate how my methodology works, I would like to provide a new variant from original logistic growth model by making $F\big(P(t)\big) = r \cdot P(t) \cdot (1 - \frac{P(t)+\alpha \cdot P^2(t)}{K})$ used for the autonomous differential equation (Eq.2) of deterministic model, :

$$\frac{dP(t)}{dt} = r \cdot P(t) \cdot (1 - \frac{P(t)+\alpha \cdot P^2(t)}{K}) \qquad \text{---Eq.7}$$

where, as defined in typical logistics growth model, $P(t)$ is the population function of time, r is the intrinsic growth rate, K is the carrying capacity. The new coefficient α is to reflect complex ecologist circumstance beyond P.F. Verhulst's model.

My model is not a wild guess but a logical strategy to add an additional term. The typical logistic growth model is just one special case of my proposal. If accuracy is enhanced with the α value different from 0, it shows I have created an improved model. If optimal α turns out to be zero, it just proves the logistic growth model truly endures the trials. Both cases are meaningful. Besides, if my methodology works, it is applicable for further improvement by trying out additional terms into the model, i.e.

$$\frac{dP(t)}{dt} = r \cdot P(t) \cdot (1 - \frac{P(t)+\alpha \cdot P^2(t)+\beta \cdot P^3(t)+\cdots}{K}) \qquad \text{---Eq.8}$$

Algorithm Implementation

Given initial point $P(t_0), r, K$ and α, below are the steps to carry out my Point-to-Point method to resolve my new variant of logistic growth model, $\frac{dP(t)}{dt} = r \cdot P(t) \cdot (1 - \frac{P(t)+\alpha \cdot P^2(t)}{K})$

Step 0: Choose a very small h
Step 1: Set $n = 0$

Step 1:Get $P'(t_0 + nh)$ by $P'(t_o + nh) = rP(t_0 + nh)(1 - \frac{P(t_0+nh)+\alpha P^2(t_0+nh)}{K})$

Step 2:Get $P(t_0 + (n+1)h)$ by $P(t_0 + (n+1)h) = P(t_0 + nh) + h \cdot P'(t_0 + nh)$
Step 3: Increase n by 1 and go to Step 1 till satisfactory coverage.
Step 4. Pick the population data corresponding to each year and get Root Mean Square Error (RMSE) calculated.

The condition of Point-to-Point method requires the time interval ,h, to be very small. I chopped a year into divisions as small as possible until finer division makes no noticeable difference. After building the curve, I conducted down-sampling to collect the estimation data that corresponds to each integer year. Estimation accuracy is evaluated through Root Mean Square Error (RMSE) which is a frequently adopted measure of the differences between estimated and observed values. The less RMSE, the better estimation.

The " r, K, α and the initial point $P(t_0)$" will be searched in four dimensions to get the minimum RMSE.

Data Analysis

In this paper, I used "year 1960" as the start point. I also built the curve backward to have a better traceability.

(A) China:

With the population data of China from World Bank Group, here is the result comparison between the original logistic growth model from Pierre-Francois Verhulst, the improved logistic growth model from Yao-Zheng and the modified logistic growth model I propose.

1. For the logistics growth model proposed by Pierre-Francois Verhulst,

$$\frac{dP}{dt} = rP(1 - \frac{P}{K})$$

, where $P(t)$ is the population function of time, r is the intrinsic growth rate and K is the carrying capacity,
the minimum RMSE is 1.263%

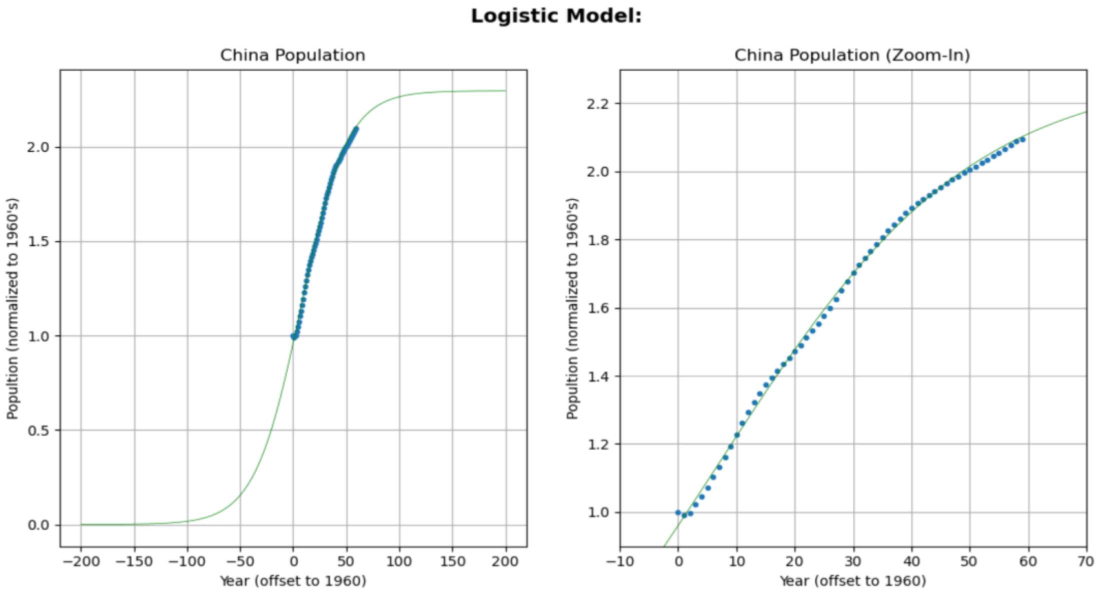


Figure.3 Curve-fitting of China population data with the typical logistic growth equation using simulation program coded by Troy Wu.

2. For the improved logistic growth model from Yao-Zheng,

$$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$$

, where $P(t)$ is the population function of time, r is the intrinsic growth rate, K is the carrying capacity and α is the coefficient to reflect complex environment,

the minimum RMSE is 1.244%

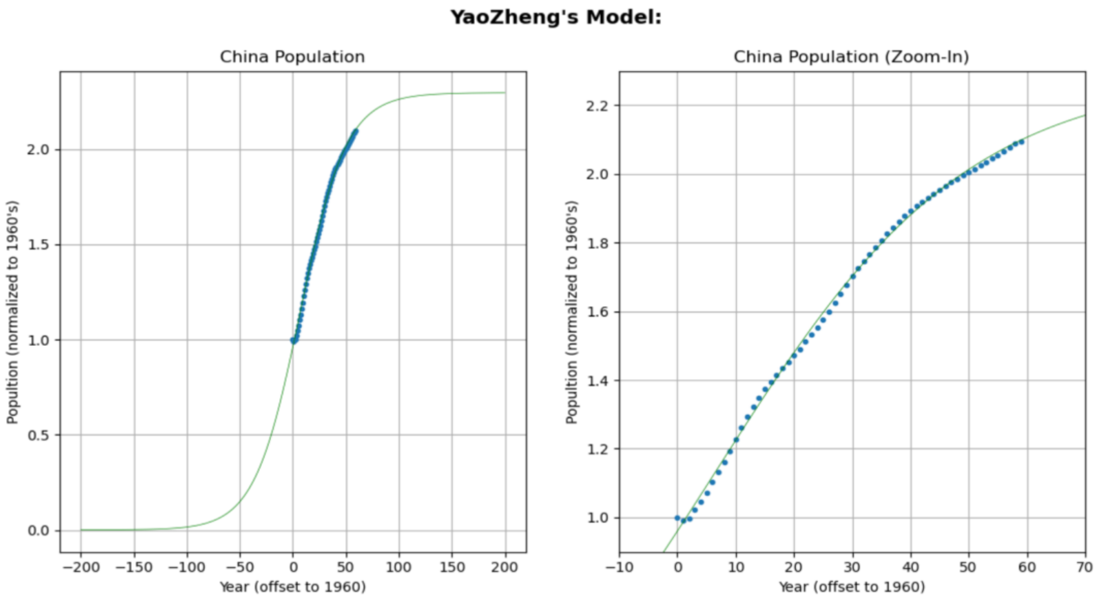


Figure.4 Curve-fitting of China population data with YaoZheng's equation using simulation program coded by Troy Wu.

3. For the improved logistic growth model that I propose,

   $$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$$

   , where $P(t)$ is the population function of time, r is the intrinsic growth rate, K is the carrying capacity and α is the coefficient to reflect complex environment,

   the minimum RMSE is 1.216%

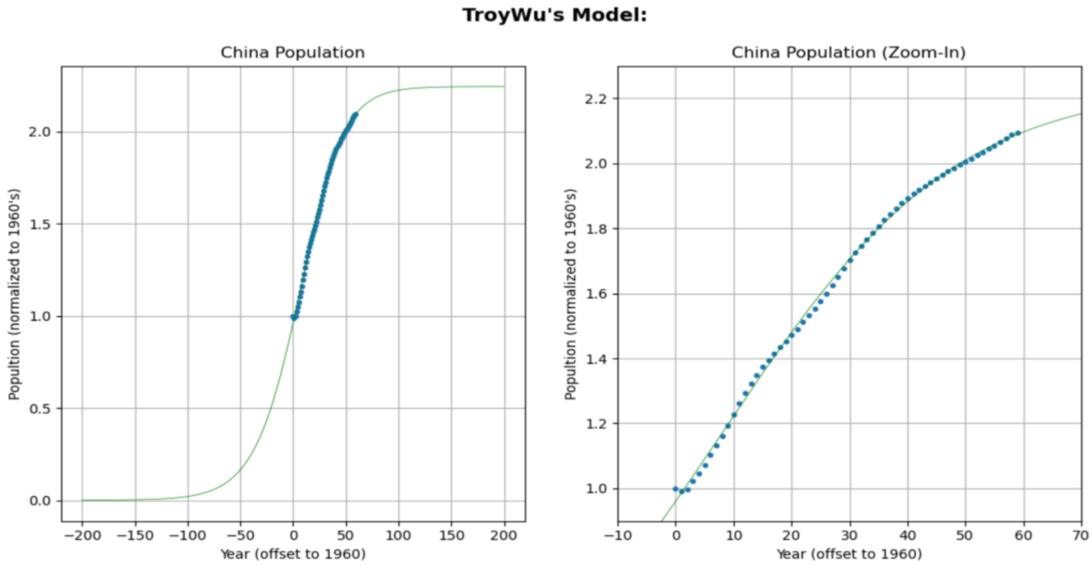


Figure.5 Curve-fitting of China population data with Troy Wu's equation using simulation program coded by Troy Wu.

Model Source	Equation	Carrying Capacity K	Growth Rate r	Parameter of New Model	Minimum RMSE
Typical	$\frac{dP}{dt} = rP(1 - \frac{P}{K})$	2.295	0.046	N/A	1.263%
Yao-Zheng	$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$	2.295	0.047	$\alpha = 1.02$	1.244%
Troy Wu	$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$	2.795	0.043	$\alpha = 0.11$	1.216%

Table 1: Comparison between various models for China

(B) India:

With the population data of India from World Bank Group, here is the result comparison between the original logistic growth model from Pierre-Francois Verhulst , the improved logistic growth model from Yao-Zheng and the modified logistic growth model I propose.

1. For the logistic growth model proposed by Pierre-Francois Verhulst,

$$\frac{dP}{dt} = rP(1 - \frac{P}{K})$$

, where $P(t)$ is the population function of time, r is the intrinsic growth rate and K is the carrying capacity,

the minimum RMSE is 1.632%

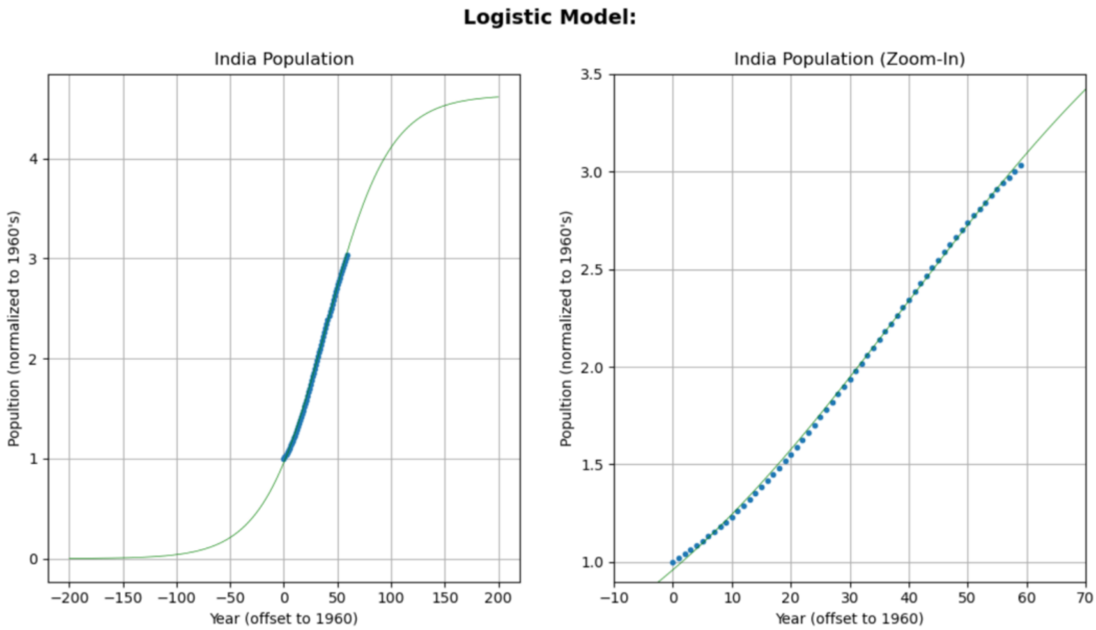


Figure.6 Curve-fitting of India population data with the typical logistic growth equation using simulation program coded by Troy Wu.

2. For the improved logistic growth model from Yao-Zheng,

$$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$$

, where $P(t)$ is the population function of time, r is the intrinsic growth rate, K is the carrying capacity and α is the coefficient to reflect complex environment,

the minimum RMSE is 1.625%

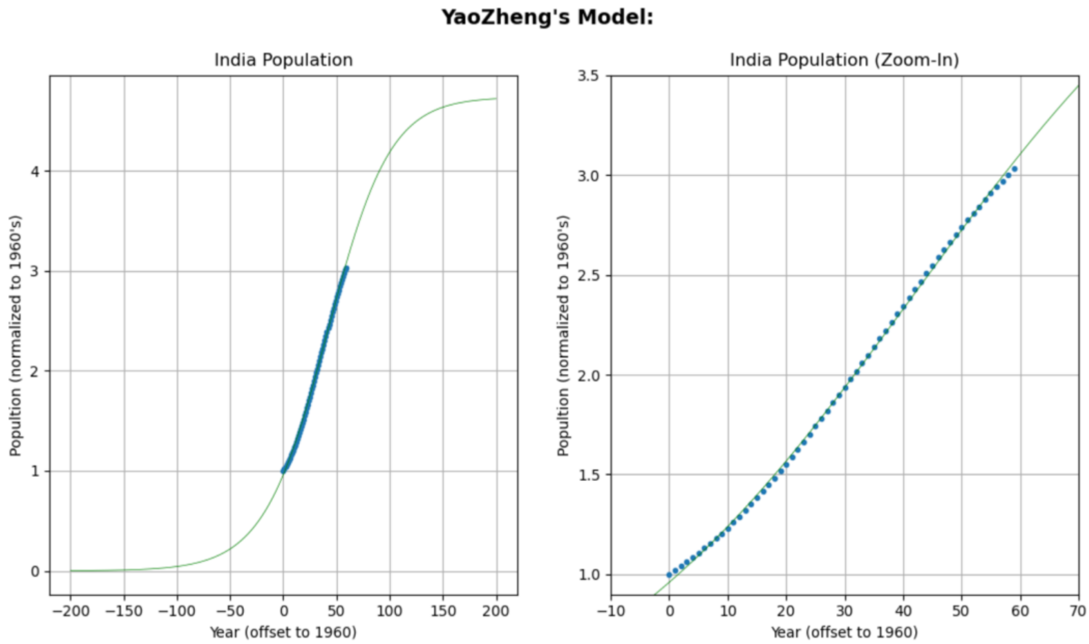


Figure.7 Curve-fitting of India population data with YaoZheng's equation using simulation program coded by Troy Wu.

3. For the improved logistic growth model that I propose,

$$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$$

,where $P(t)$ is the population function of time, r is the intrinsic growth rate, K is the carrying capacity and α is the coefficient to reflect complex environment,

the minimum RMSE is 1.430%

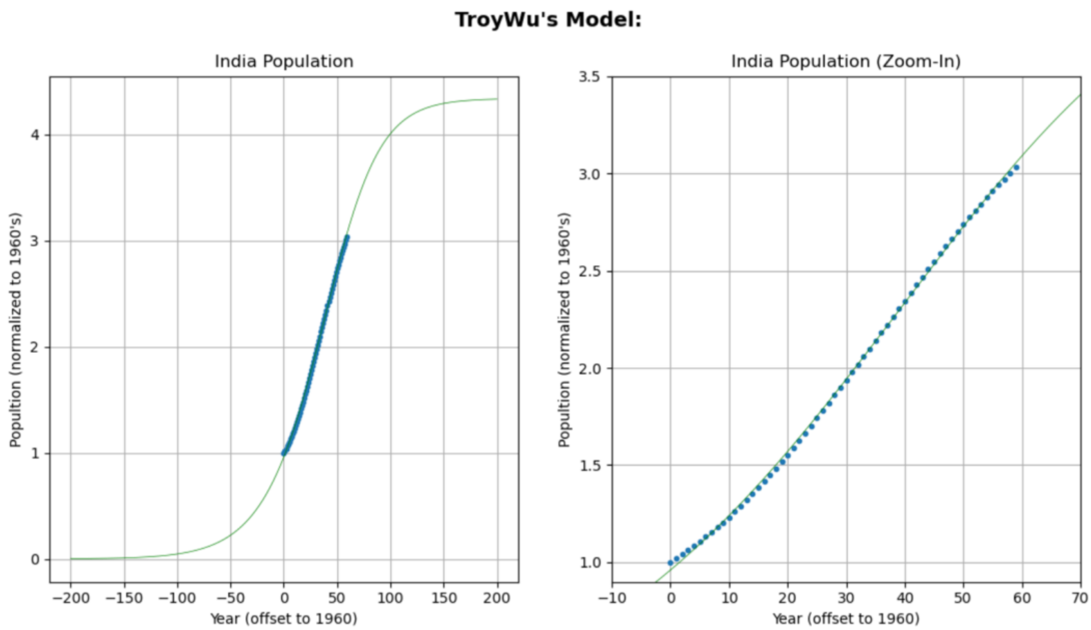


Figure.8 Curve-fitting of India population data with Troy Wu's equation using

simulation program coded by Troy Wu.

Model Source	Equation	Carrying Capacity K	Growth Rate r	Parameter of New Model	Minimum RMSE
Typical	$\frac{dP}{dt} = rP(1 - \frac{P}{K})$	4.663	0.034	N/A	1.632%
Yao-Zheng	$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$	4.733	0.033	$\alpha = 0.97$	1.625%
Troy Wu	$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$	6.033	0.032	$\alpha = 0.09$	1.430%

Table 2: Comparison between various models for India

It demonstrates that my algorithm to resolve differential equations works for all the three models under discuss. Besides, both Table 1 and Table 2 show my proposed model has the least estimation error.

Conclusion

Among the original logistic growth model, the modified logistic growth models respectively from Yao-Zheng and me, my proposal has the least root mean square error to reflect the best accuracy. However, my proposing a model to beat others is not the only purpose of this exploration. Instead, I intend to highlight the merit of my algorithm that makes the model-proposing and the model-verification become easy for scientists. Also, it is demonstrated that, by adding additional terms to the differential equation, we can easily improve an existing model after the math-handling is made so simple.

Furthermore, it shows another advantage of my method to tell where the sample data stands along the fitting curve. Accordingly, we could observe China's population is way closed than India's to the carrying capacity which could be interpreted as the control target set by a civilized government. It indicates China has been conducting more effective population control, compared with India.

The most difficult part of this exploration is that all my assertions are so uncertain before simulation verification which I must count on my own coding to work out. A logical error or an inevitable typo in coding often resulted in misleading simulation and trapped me for days. I came to realize that confidence and attention on every step are essential for going through a mist;

I also comprehended a project without any teammate calls for double efforts.

I used to think that difficult Calculus is hardly applicable in real life. Thanks to my teacher's guidance through this internal assessment, I have got an opportunity to combine mathematical knowledge and computer skill to make a useful tool for handling life problems. It is definitely a wonderful experience for me to act responsibly in an increasingly interconnected but uncertain world.

Bibliography

1. Meadows, D. (1986). Poverty Causes Population Growth Causes Poverty, from 7 Facts About Overpopulation and Poverty, Borgen Project https://borgenproject.org/7-facts-about-overpopulation-and-poverty/
2. Dobrushkin, V. (2015) MATLAB Tutorial I: the First Course in Differential Equations, part 1.2: Population models , Applied Differential Equations. The Primary Course by CRC Press, 2015 http://www.cfm.brown.edu/people/dobrush/am33/Matlab/ch2/population1.html
3. Strang, G. Herman, E.J. (2019). *32 The logistic Equation. Calculus Volume 2*. Rice University. Open Stax. Retrieved from https://opentextbc.ca/calculusv2openstax/chapter/the-logistic-equation/
4. Zheng, Y. (n.d.). *The improvement and verification for the logistic population growth equation*. Northern Illinois University. SSRN. Retrieved

from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=887542

5. Wu, T. (2020). *Finding actual solutions for nonlinear Equations* https://troywuofficial.com/?p=266