An Enhanced Logistic Growth Model for Population Estimation: Case Studies of China and India

Author: Yuchung (Troy) Wu

Date: March 2021

Introduction

Since returning from Seed of Hope program, I have been haunted with the miserable lives that people live in an overpopulated nation. I am eager to do something for them right away.

Population control is pressing to prevent from further over-competition on limited resources on the planet. Essentially, our society needs an appropriate tool or model to accurately evaluate and estimate the effectiveness of any control plan. The prevailing one is Logistic Growth Model. It will be constructive to the world if I could explore a way to enhance the estimation accuracy on the 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 model which could, with parameters sophisticatedly tuned, fit well the collected population data.

Pierre François Verhulst, a Belgian mathematician, published his equation to model population growth under limited resource in 1838:

$$\frac{dN}{dt} = rN - \alpha N^2$$

where *N* represents number of individuals at time *t*, *r* is the intrinsic growth rate, and α is the density-dependent crowding effect (also known as intraspecific competition). In this equation, the population equilibrium referred to as the carrying capacity, *K*, is

$$K = \frac{r}{\alpha} \quad .$$

His equation was popularized by Raymond Pearl and Lowell Reed to make the well-known one now,

$$\frac{dN}{dt} = rN(1 - \frac{N}{K})$$

where N represents number of individuals at time t, r the intrinsic growth rate and K the maximum population size that can be supported by the environment (carrying capacity) (Cramer, J.S. 2002*)*.

In this exploration, I adopt P(t) to replace N to reflect the population function of time, as prof. Gilbert Strang from Massachusetts Institute of Technology (MIT) and popular Wikipedia did.

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

This differential equation is resolved to have solution,

$$P(t) = \frac{K}{1+e^{-r(t-t_o)}},$$

where t_o is a constant determined by the initial condition. (Strang, G. Herman, E.J. (2019))

However, getting a solution of any modified Logistic Growth Model often involves sophisticated skills to resolve a variant from the original differential equation. I have hardly found successful attempts except the following paper from Yao-Zheng (Zeng, Y. (2006)).

Assistant Professor Yao-Zheng from Northern Illinois University asserts she could leverage power exponent models which are widely used in scientific and engineering. She builds a variant from Logistic Growth model to be,

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

where P(t) represents the population at time t, r the intrinsic growth rate and

K the carrying capacity; α is a constant with which she tries to reflect complex circumstance . She claims to have the solution of this equation in analytic form. However, even for α as integers, she is not able to present a solution consisting of fundamental functions but only able to simply the differential equation to another equation not calling for differential operation, as shown below,. (*Figure.1*). It must be way more complicated for the cases of fractional α , which she doesn't 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 Solutions of her Power Exponent Model,* $\frac{dP(t)}{dt}=rP(t)\cdot(K-P(t))^{\alpha}$

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 unique point, if any, each time. We could only list the point solution in table format as Y. Zheng does in her report. (Figure.2). It turns out we cannot infer the estimated curve which fits the collected data.

Table 1

Year	Census	Predict	Error	Year	Census	Predict	Error
1949	14.16	14.1600	0.00%	1977	30.29	30.0323	-0.85%
1950	14.48	14.6619	1.26%	1978	30.62	30.5621	-0.19%
1951	14.81	15.1740	2.46%	1979	30.95	31.0810	0.42%
1952	14.96	15.6958	4.92%	1980	31.26	31.5885	1.05%
1953	15.38	16.2270	5.51%	1981	31.79	32.0840	0.92%
1954	16.09	16.7671	4.21%	1982	32.17	32.5669	1.23%

Figure.2 The table in which Y. Zheng put her point solutions.

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: To make easy the approach to resolve the logistic growth model variants

By proposing a new model and introducing the procedure to resolve the corresponding differential equation, this report records a novel methodology which gets rid of mathematical barrier and help all the ecologists freely propose new modifications on logistics models. Accordingly, an ecologist could verify the estimation accuracy easily and observe the fitting curve to get the scientific sense at the same time.

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

Introduction on Point-by-Point Method

According to first principle,

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

, we have equivalently

$$y'(x_i) = \frac{y(x_i+h)-y(x_i)}{h}, where\ h \to 0, \forall x_i \in domian$$

After rearrangement, it tells us :

$y(x_i + h) = y(x_i) + h \cdot y'(x_i)$, where $h \to 0$

When we apply it to resolve a differential equation,

$y'(x) = G\big(y(x)\big)$, ,where G is any given function

, it becomes

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

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

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

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

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

For x_i to increase by h, i.e. $x_i = x_o + h$, by

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

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

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

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

$y(x_0 + h)$

For x_i to increase by another h, i.e. $x_i = x_o + 2h$, by

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

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

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

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

$y(x_0 + h)$, $y(x_0 + 2h)$

………

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

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

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

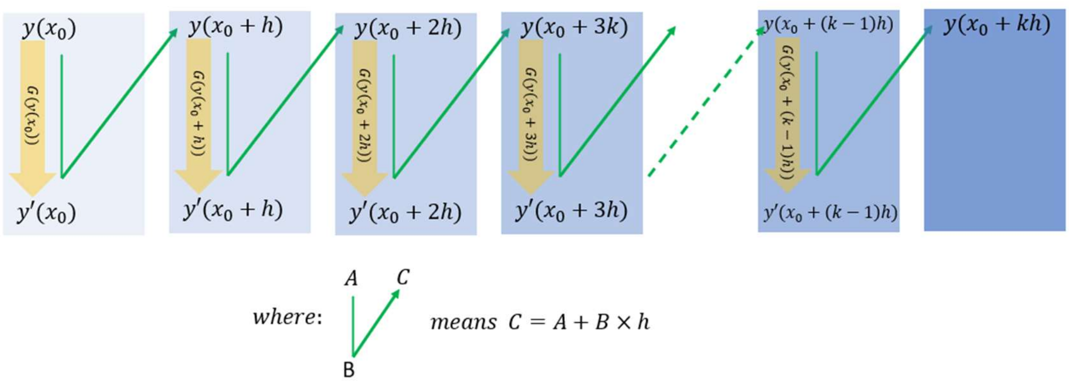
$\rightarrow y(x_0 + kh) = y(x_0 + (k-1)h) + h \cdot G(y(x_0 + (k-1)h))$

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

$y(x_0 + h),\ y(x_0 + 2h) \ldots .\ y(x_0 + (k-1)h)$

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

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



The demonstrated process could be generalized to resolve any *N-order differential* equation which is in the form of

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

, where G(x) is any well-defined function,

(e.g.,$y^{(2)}(x) = \sqrt{\sin\,(y(x) + b(y'(x))^2}$)

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 to resolve $y^{(n)}(x) = G\left(y^{(0)}(x), y^{(1)}(x), \ldots, y^{(n-1)}(x)\right)$

(Wu, Troy (2020))

Given differential equation

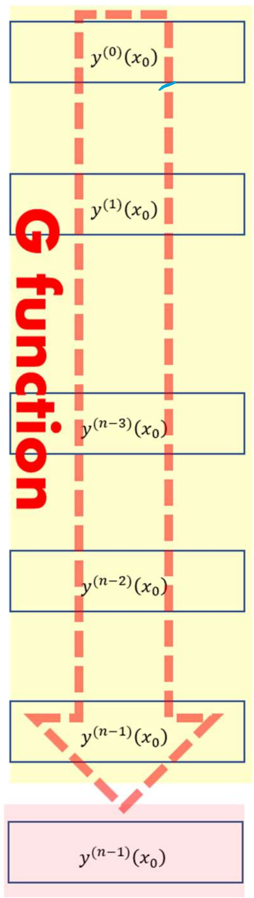
$$y^{(n)}(x) = G(y^{(0)}(x), y^{(1)}(x), y^{(2)}(x), \cdots, y^{(n-1)}(x))$$

with 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 $y(x)$ value together with 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

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

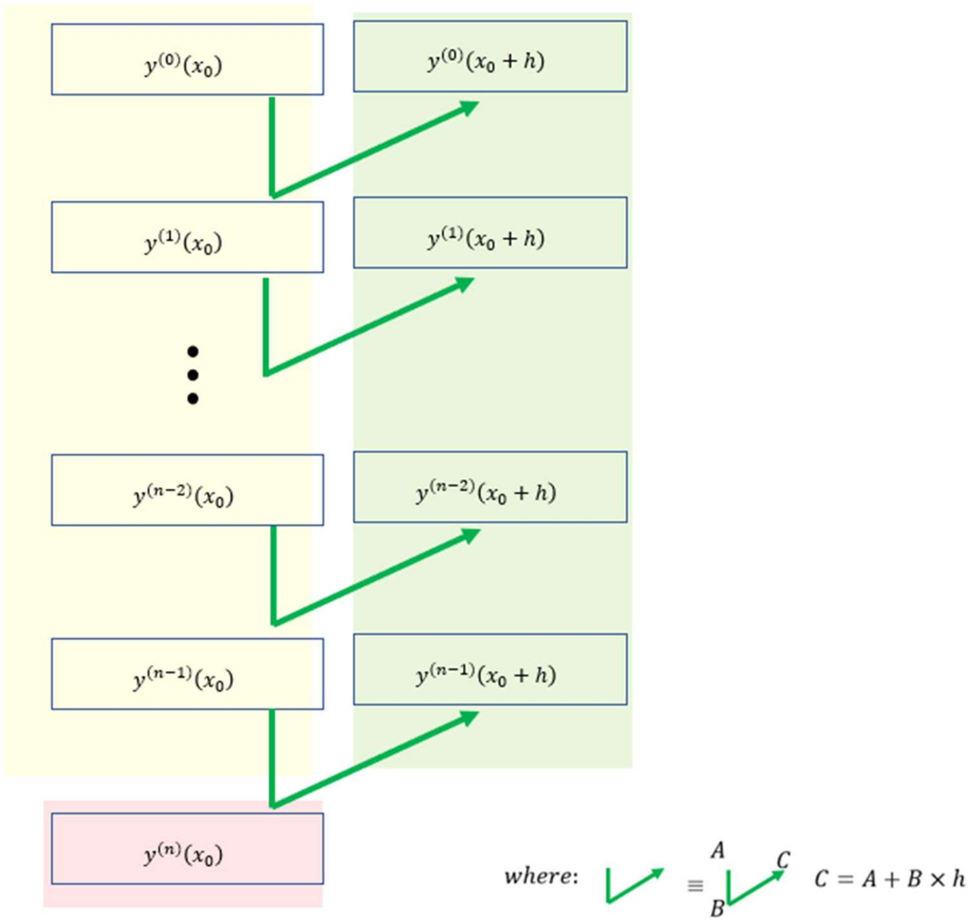
$$y^{(1)}(x_0+h) = y^{(1)}(x_0) + y^{(2)}(x_0) \times h$$

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

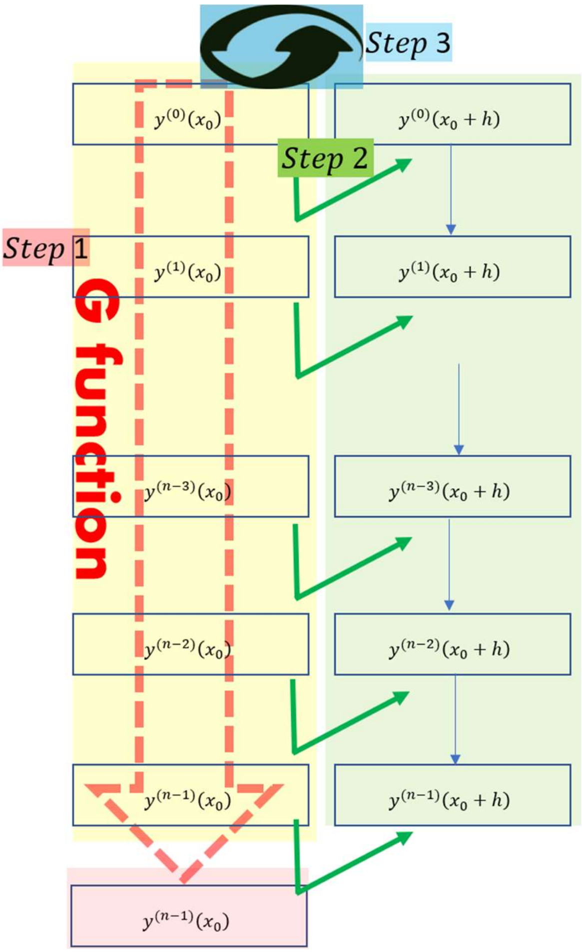
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

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

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



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



Take the logistic growth equation to the work example:

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

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 $t = kh, \forall t$ 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 Model:

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

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

It is next to impossible for my model to work worse because the typical Logistic Growth is just one special case in my model. If better accuracy is acquired with the α value different from 0, it shows I have hit an improved model. If optimal α turns out to be zero, it just proves the logistic growth mode truly endure the trials. Both cases are meaningful. Besides, if my model and methodology work, it is applicable for further improvement by trying out additional terms into the model , i.e.

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

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: 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 chop one year into divisions as small as possible until further division makes no difference. After building the curve, I conduct down-sampling to collect the estimation data that corresponds to each integer year. The sampled data is compared with the practical one through Root Mean Square Error (RMSE) which is a frequently used measure of the differences between estimated and observed values. The less RMSE, the better estimation.

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

Data Analysis

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

(A) China:

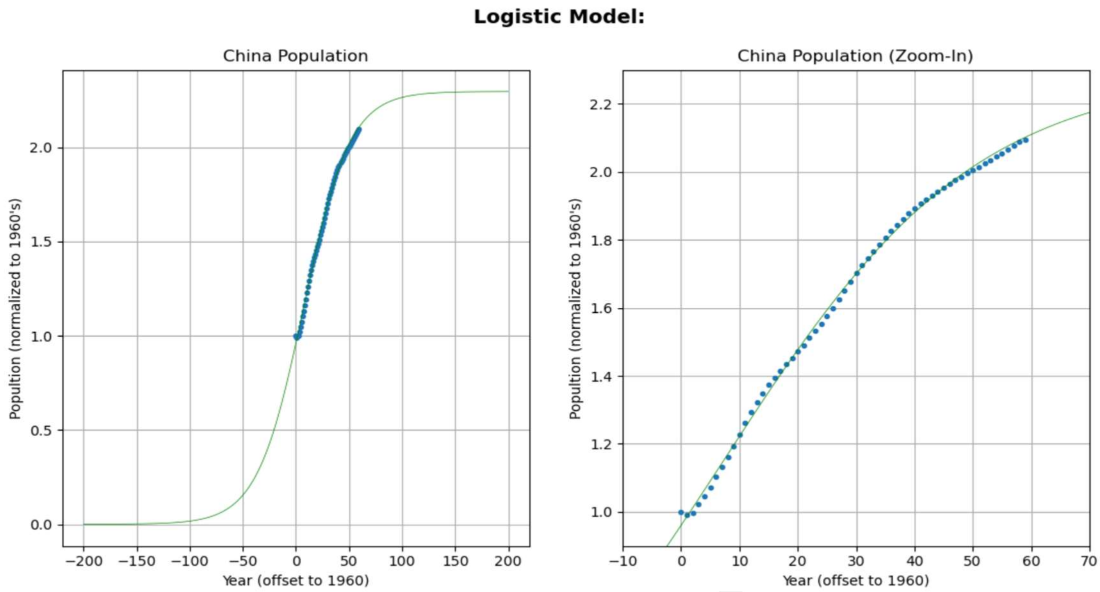
With the population data of China, here is the accuracy comparison between the various modellings.

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%



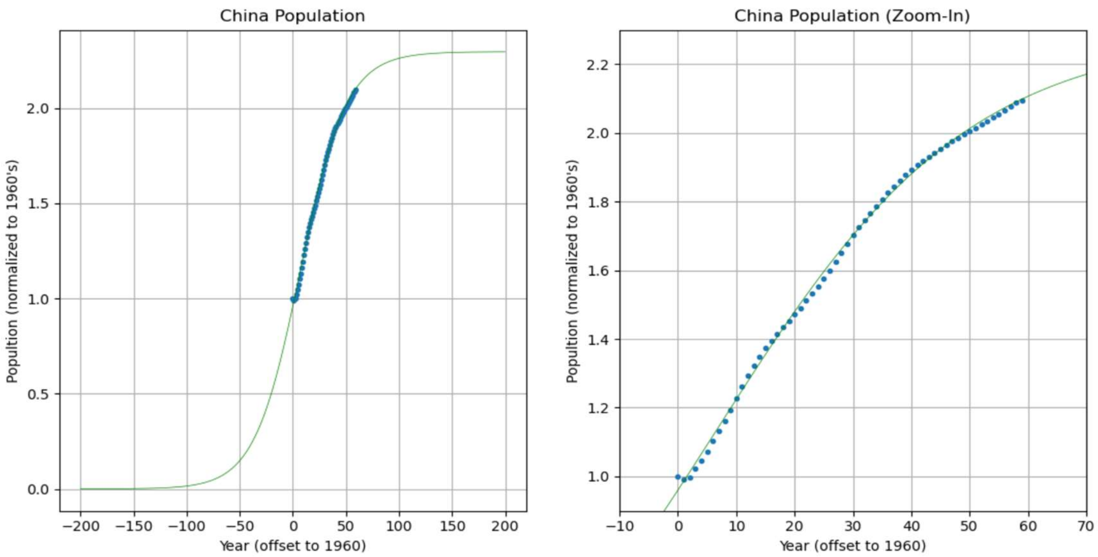
2. For the improved Logistics Growth Model from Yao Zheng,

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

,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%

YaoZheng's Model:



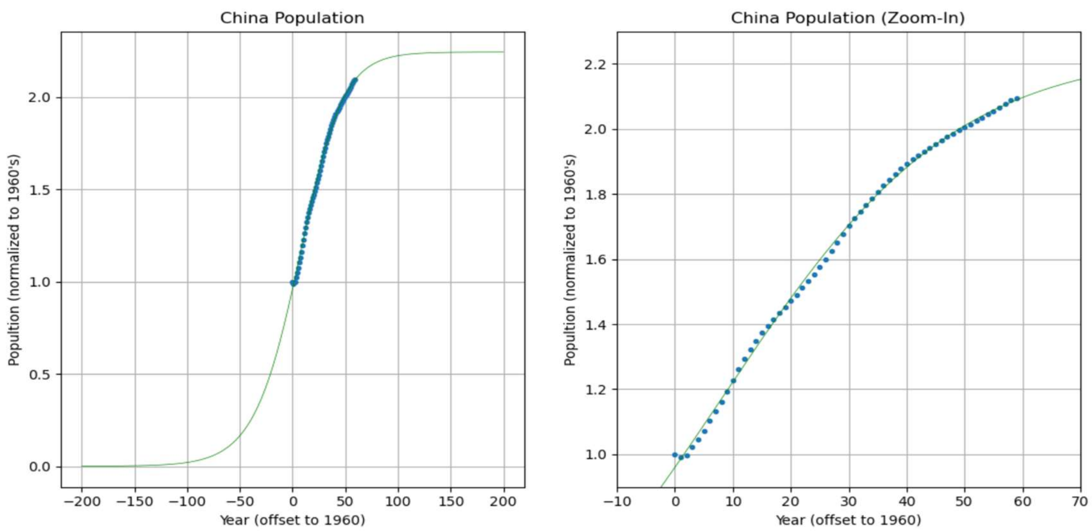
3. For the improved Logistics Growth Model that I proposed,

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

,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%

TroyWu's Model:



Model Source	Model Type	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%
Published by Yao-Zheng	$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$	2.295	0.047	$\alpha = 1.02$	1.244%
My proposal	$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$	2.795	0.043	$\alpha = 0.11$	1.216%

Table 1: Comparison between various models for China

(B) India:

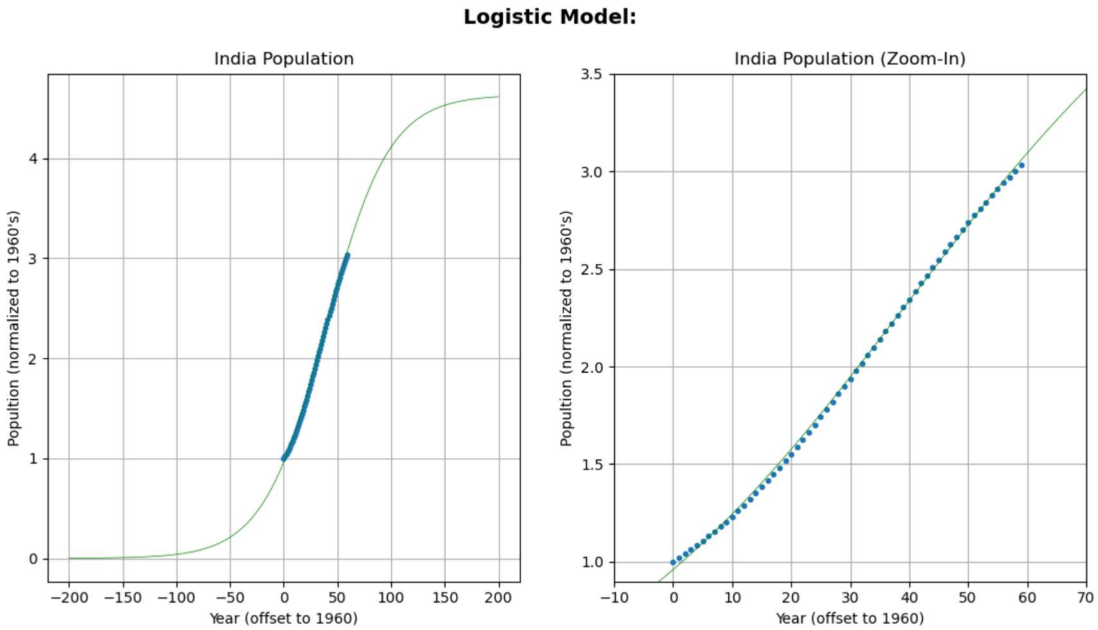
With the population data of India, here is the accuracy comparison between the various modellings.

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 1.632%

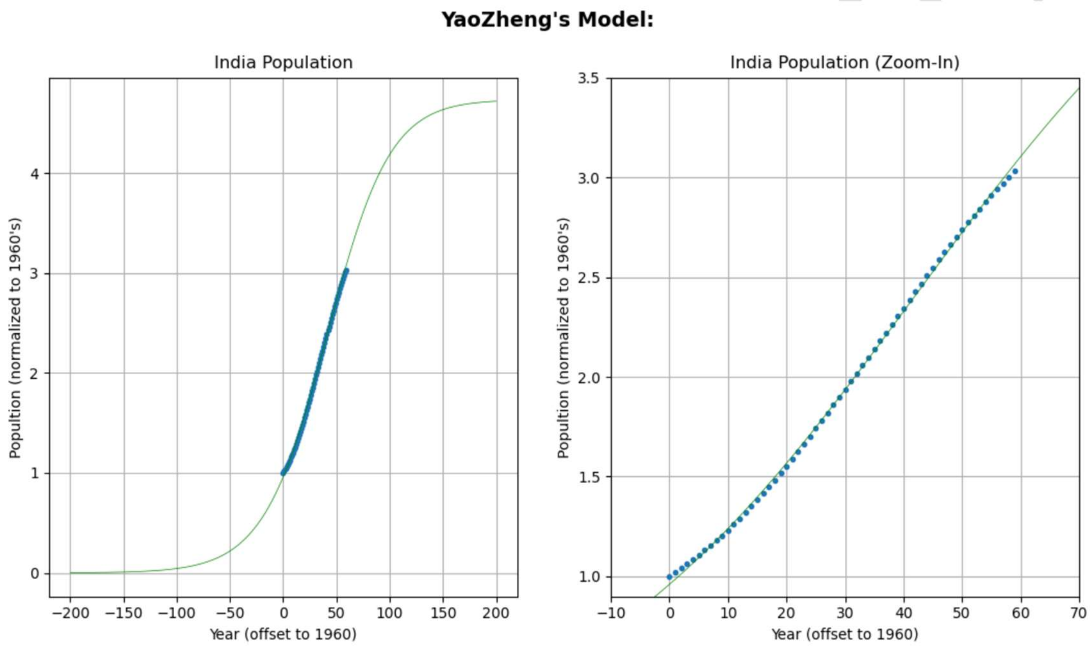


2. For the improved Logistics Growth Model from Yao Zheng,

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

,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%



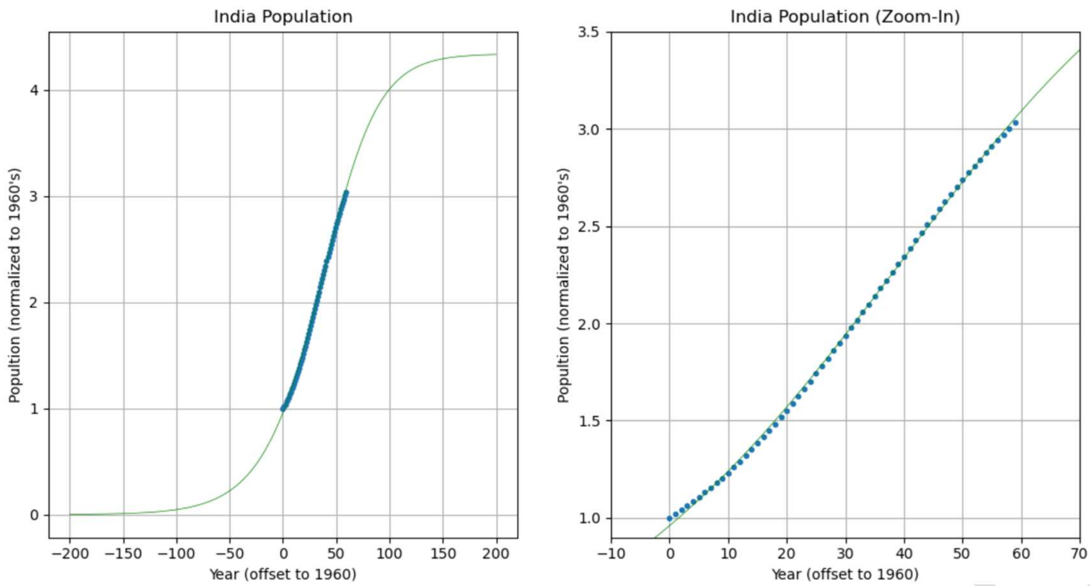
3. For the improved Logistics Growth Model that I proposed,

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

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

the minimum RMSE is 1.430%

TroyWu's Model:



Model Source	Model Type	Carrying Capacity K	Growth Rate r	Parameter of New Model	Minimum RMSE
Typical	$\frac{dP}{dt} = rP(1 - \frac{P}{K})$	4.633	0.034	N/A	1.632%
Published by Yao-Zheng	$\frac{dP}{dt} = rP(1 - \frac{P}{K})^{\alpha}$	4.733	0.033	$\alpha = 0.97$	1.625%
My proposal	$\frac{dP}{dt} = rP(1 - \frac{P+\alpha P^2}{K})$	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 equation works for all models. Besides, both Table 1 and Table 2 show my proposed model has the least estimation error.

Conclusion

My model has the least root mean square error to reflect that better accuracy is acquired with my proposed model. However, my proposing model to beat others existing is not the purpose of this exploration. Instead, I'd like to highlight the merit of my algorithm that makes the model-proposing and the model-verification easier for scientists. Also, it is demonstrated that ,by adding more terms, we can easily improve a model in the differential equation when the handling math is made simple.

Besides, I'd like to highlight the merit that my method tells where the the sample data stand along the fitting curve. For instance, we could observe China population is way closed than India to the carrying capacity which is set by nature or by government. It reflect China has truly conducted more effective population control, compared with India.

I hope my exploration to bring any positive influence to the world.

Bibliography

1. Cramer, J. S. (2002) *The origins of logistic regression* (Technical report). 119. Tinbergen Institute. pp. 167–178 https://papers.tinbergen.nl/02119.pdf
2. 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/
3. 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
4. Wu, Troy (2020). *Finding actual solutions for nonlinear Equations* https://troywuofficial.com/?p=266
5. Enciso, G, A. [UCI Open]. (2014). *Mathematical Biology. 14: Predatory Prey Model*. YouTube. Retrieved from: http://ocw.uci.edu/lectures/math_113b_lec_14_introduction_to_mathematical_modeling_in_biology_predator_prey_model.html
6. Hirsch, Morris, W. (1982). *Systems of Differential Equations that are Competitive or Cooperative I: Limit Sets*. UC Berkeley. eScholarship. https://escholarship.org/content/qt01n7r6wm/qt01n7r6wm.pdf?t=obc1qg