A Population Model with Stochastic Processes and a Fitness Mechanic

Authors: *Parth Adhia*, Raj Joshi*, Mihir Rastogi*, Aditya Agashe*, Gil Shteyman*
Institution: *Middlesex County Academy for Science, Mathematics, and Engineering Technologies*
Department: *Electrical and Computer Engineering Track*
Location: *Edison, NJ, USA*

Abstract

We create a simulation modeling population growth, building upon the work of D J Aldous. Aldous creates an elementary model of city growth, akin to the "Chinese restaurant process". His two determinants of growth are α and β, which are two constants which influence the behavior of the simulation. However, the real world is the result of a myriad of factors, and we wish to better illustrate that. Thus, we add further parameters to Aldous's paper to represent a more accurate depiction of real world behaviors. Added features include "fitness", representing age and health, which influence tendencies for individuals to migrate based on preference and conditions. In addition, we have six inputs that dictate the behavior of the model's growth. In addition to Aldous's α and β, we utilize B, birthrate, C_0, a constant of influence, c, the rate of immigration, and m, an indication of when people migrate. Together, these inputs and factors create a more realistic sample of real world city growth phenomena. An analysis of the resulting data yields many patterns regarding total population, expected number of cities, the emergence of an infinite population, and critical points.

Keywords: Population Simulation, Spatial Growth, Stochastic Modeling, Asymptotic Behavior

1 INTRODUCTION

A popular topic in probability theory is the Chinese restaurant process (CRP). Based on the metaphor that Chinese restaurants have unlimited seating, it creates a scenario in which the probability of a new customer sitting at any given table is proportional to the amount of customers already at the table. At the fundamental level, it assigns new data based on the existing distribution of the partitioned data.

This process is the foundation of D J Aldous' paper, "A spatial model of city growth and formation." Aldous creates a mathematical model to reflect the aforementioned process in 2-dimensions. Aldous associates his model with the phenomenon of city population growth and formation, albeit it does not represent a realistic scenario. In his simulation, two main inputs α & β are defining factors of the end behavior. His results are summarized as follows: when $\beta > 2\alpha$, the model exhibits "Balanced Growth," where there is a steady increase in number of large cities with the same magnitude. When $\beta < 2\alpha$, the model undergoes "Unbalanced Growth," where a few cities contain majority of the population [1].

An interesting discussion can be created if certain complexities are added to the simulation to model population growth in a more realistic manner. Namely, an element of "fitness" can be associated with each person in the simulation. This makes way for the creation of more complex interactions in the simulation, such as the ability for individuals to migrate from one city to another if the population is more fit than the average fitness of its current city, and a life expectancy phenomena where members of a city die when they reach the life expectancy of the city. To further mimic real life, a birth mechanic is added in order to more closely simulate the growth of a city by adding the aspect of reproduction.

Our research aims to analyze the resulting patterns that are created by adding complexities to Aldous' population growth model. Specifically, we will research patterns revolving around the expected number of cities, total population, and the growth of cities.

An analysis of these results yields numerous patterns. One such is that of total population, a function comprised of many sub functions until the 77th year, reflecting different milestones over the course of our simulation. These milestones include the first births, and the first deaths. The function allows us to attain a highly accurate estimation of the population in any specific year in order to model the future. Some results of this function include: when $B < 1$, the population tends towards a specific value, while at all other values population diverges to infinity.

Another such pattern can be found with expected cities. Using a deduction of the average distance between two random points in the simulation, an analysis of the data yields a function representing a rough prediction of the expected number of cities at any point in time. This can be used to make decisions based on futuristic models.

This model can be a precursor to create more complex models where other patterns in the growth and development of cities can be analyzed. The primitive fitness function and moving mechanic that is used in this model can be replaced by more complex processes to model different social behaviors and demographic transformation.

2 MODEL

Our model requires six inputs that dictate the behavior of the model's growth:

1. Alpha, α: How much the population of cities affects their sphere of influence.
2. Beta, β: The rate at which sphere of influence decreases.
3. Birthrate, B: The chance that an adult will spawn a child(ren).
4. Constant of influence, C_0: Constant which determines the influence of cities.
5. Rate of Immigration, w: The amount of adults who immigrate every time interval.
6. Moving, m: The percentage above a person's fitness that the total average fitness must be in order to decide to move to a different city.

Our model has the potential to run on a infinite time interval starting at $t = 0, 1, 2, 3...$ years (however, in order to simulate and gather data, the model was run for a period of 500 years) where cities are created at positions p_j inside a square $[0, 1000]^2$. A city exists if its population $N_j(t) \geq 1$ where total population of the simulation $T = \sum_{j=1}^{n} N_j(t)$ and n is the current number of cities. There are 3 parameters of a city's influence:

$$0 < C_0 < \infty,\ \alpha > 0,\ \beta > 0$$

These parameters are used to define the influence of a city at a particular point inside the grid

$$I_j(C_0, N_j, d) = C_0 N_j^{\alpha} d^{-\beta} \qquad \text{Equation 2.1.}$$

where d is defined as the distance between p_j and a location inside of the grid and N is the population of city j.

At the start of the simulation, a city with a population of 1 is spawned at a random location inside of the grid. In each subsequent time interval, w immigrants spawn at a random location inside of the grid and are assigned to city j with probability P_j:

$$P_j = \frac{I(C_0, N_j, d)}{(1 + \sum(I(C_0, N_j, d))} \quad \text{Equation 2.2.}$$

Immigrants can also spawn and start their own city at an unfounded location with probability R, where

$$R = \frac{1}{(1 + \sum(I(C_0, N_j, d))} \quad \text{Equation 2.3.}$$

Each immigrant enters the simulation with an age a of 18 years. Their age increases by 1 following each time interval. Furthermore, every person in the simulation has a fitness f defined by

$$f = 100e^{-\frac{(a-45)^2}{625}} \quad \text{Equation 2.4.}$$

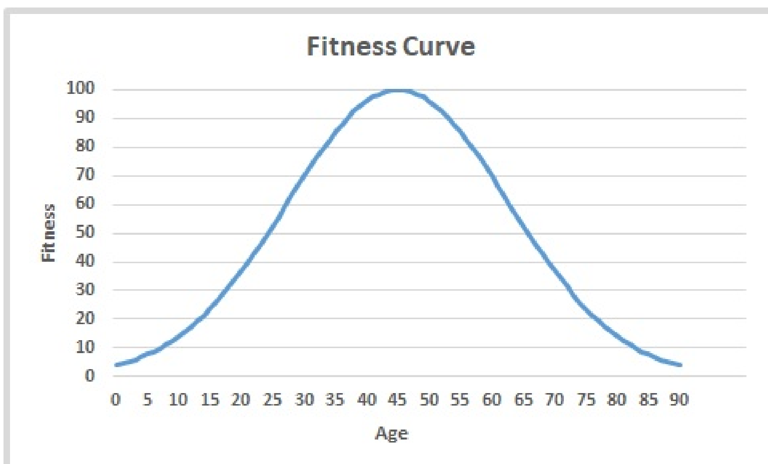

Figure 1. Fitness Curve

The population of each city is divided into two sections: the children and the adults. Each citizen with age $a < 18$ is considered to be a child, and all other citizens are adults. The number of adults in city j is denoted as A_j.

In each subsequent time interval, every person in the simulation at age $a = 30$ has a probability B of spawning children where B is inputted by the user. Let x be the largest integer smaller than B. The person will be guaranteed to spawn x children and will have a probability of $B - x$ to spawn 1 more.

The average fitness of city j, S_j, is given by

$$S_j = \frac{\sum_{k=1}^{A_j} f_k}{n} \quad \text{Equation 2.5.}$$

The life expectancy, for a given city j, L_j, is

$$L_j = \frac{114e^{\frac{n*S_j}{\sum S_j} - 0.3}}{e^{\frac{n*S_j}{\sum S_j} - 0.3} + 1} \quad \text{Equation 2.6.}$$

A person dies when their age surpasses the life expectancy of their city.

In each subsequent time interval, members of a city P_{s_j} can migrate to another when their fitness is m% higher than S_j. The person randomly moves to a new city j based off probability R_j:

$$R_j = \frac{S_j}{\sum S_j} \qquad \textbf{Equation 2.7.}$$

We define an "active city", j, as a city in which $N_j > 0$.

3 LONG-TERM ANALYSIS

3.1 *Total Population*

Lemma 3.1.1. *The average age of death for any given simulation is 77.*

Proof. The life expectancy, for a given city j, L_j (eq. 2.6), is

$$L_j = \frac{114e^{n*\frac{S_j}{\sum S_j}-0.3}}{e^{n*\frac{S_j}{\sum S_j}-0.3}+1}$$

Consider the function

$$F(x) = \frac{1.5e^{x-0.3}}{e^{x-0.3}+1}$$

Here x represents the quotient of the fitness of a city and the fitness of the total population. The quotient, x will have a mean value of 1. This is because on average, the fitness of a city will equal the fitness of the total population. This is because if one city has a fitness greater than that of the population, another city must have fitness less, thus creating an average of 1. We can calculate the average value of $F(x)$ by computing $F(1)$, which is 1.002. We can relate F(x) to life expectancy by the equation

$$F(x) * 76 = L_j$$

We see this holds because 114 is a factor of 76 larger than 1.5. Thus, the average value of L_j will be 76 times larger than that of $F(x)$, or 76.152. However, our simulation causes a person to die after exceeding the life expectancy. Thus, the average age of death is 77. $\square$

We now analyze the model for the total population of the entire simulation at any year t. To find the total population function, we must first calculate the immigrants entering the simulation at any year t. At year 0, w immigrants enter the simulation, and w immigrants enter the simulation every year until year 13. At year 13, w immigrants enter and wB kids enter due to the initial immigrants being 30 years of age, and therefore having children. People continue to enter at a rate of $w + wB$ until year 43, when the kids of the immigrants have wB^2 kids. This pattern is described below.

Year	People Entering
1	w
13	$w + B$
43	$w + wB + wB^2$
73	$w + wB + wB^2 + wB^3$
103	$w + wB + wB^2 + wB^3 + wB^4$
30n-17	$w + wB + wB^2 + wB^3 + ... + wB^n$

In order to calculate total population, we must add the previous 77 years of people entering the simulation. Anyone entering the simulation before 77 years prior would be deceased because the average age of death is 77. The number of people entering each year is simply the sum of a finite geometric sequence. Moreover each sum must be multiplied by a constant because multiple years can have the same number of people entering the simulation. Thus, each sum would be: $C * w(\frac{1-B^n}{1-B})$, where C and n depend on the year t. However, immigrants enter at age 18, and are 19 years ahead because children start at age 0. Thus, we must subtract $19w$ from the final sum. Adding up all the sums and subtracting $19w$ yields:

$$\begin{aligned} Q_0(t) &= (((t-13) \bmod 30)(w(\frac{1-B^{\phi+1}}{1-B})) + 30(w(\frac{1-B^{\phi}}{1-B})) \\ &\quad +([47-((t-13) \bmod 30)])(w(\frac{1-B^{\phi-1}}{1-B})) \\ &\quad +(\{17-((t-13) \bmod 30)\})(w(\frac{1-B^{\phi-2}}{1-B})) - 19w \end{aligned}$$

where $[x]$ means that if $x > 30$ then $x = 30$, {x} means that if $x < 0$ then $x = 0$, and

$$\phi = \left\lfloor \frac{t+17}{30} \right\rfloor$$

Due to the fact that we are adding the previous 77 years, we can only use $Q_0(t)$ when $t \geq 77$. For $t < 77$, we can add up only the previous t years. Thus, we have:

Year	**Total Population**
$t \geq 77$	$Q_0(t)$
$73 \leq t \leq 76$	$Q_1(t) = (t-72)(w(\frac{1-B^4}{1-B})) + 30(w(\frac{1-B^3}{1-B})) + 30(w(\frac{1-B^2}{1-B})) + (70-t)w$
$59 \leq t \leq 72$	$Q_2(t) = (t-42)(w(\frac{1-B^3}{1-B})) + 30(w(\frac{1-B^2}{1-B})) + (70-t)w$
$43 \leq t \leq 58$	$Q_3(t) = (t-42)(w(\frac{1-B^3}{1-B})) + 30(w(\frac{1-B^2}{1-B})) + 12w$
$13 \leq t \leq 42$	$Q_4(t) = (t-12)(w(\frac{1-B^2}{1-B})) + 12w$
$t \leq 12$	$Q_5(t) = wt$

The final total population function is:

$$Q(t) = \begin{cases} Q_0(t) & t \geq 77 \\ Q_1(t) & 73 \leq t \leq 76 \\ Q_2(t) & 59 \leq t \leq 72 \\ Q_3(t) & 43 \leq t \leq 58 \\ Q_4(t) & 13 \leq t \leq 42 \\ Q_5(t) & t \leq 12 \end{cases}$$

where t is the current year

Solving future equations using $Q(t)$ would prove tedious and in some cases, impossible. Therefore, we introduce a piece-wise continuous function $Q'(t)$ which is an approximation of $Q(t)$.

$$Q'(t) = \begin{cases} Q'_0(t) = 77w(\frac{1-B^{t/30}}{1-B}) - 19w & t \geq 77 \\ Q'_1(t) = tw(\frac{1-B^{t/30}}{1-B}) + (70-t)w & 58 < t < 77 \\ Q'_2(t) = tw(\frac{1-B^{t/30}}{1-B}) & 0 \leq t \leq 58 \end{cases}$$

where t is the current year

We will examine the end behavior of the population of a city. This depends on the value of B, and so we will analyze this outcome for three cases: $B < 1$, $B = 1$, and $B > 1$.

Lemma 3.1.2. *When $B < 1$, the total population tends towards*

$$Q = w(\frac{77}{1-B} - 19) \qquad \textbf{Equation 3.1.1.}$$

Proof. As t tends towards infinity, we see that ϕ also approaches infinity. Thus, $\lim_{t\to\infty} Q(t) = \lim_{\phi\to\infty} Q(t)$. We also know that as ϕ approaches infinity, B^{ϕ} approaches 0, because B is less than 1. Hence: $\lim_{\phi\to\infty} w(\frac{1-B^{\phi+c}}{1-B}) = \frac{w}{1-B}$, where c is a constant. Because all of the coefficients of the sums of the geometric series in $Q(t)$ add up to 77, we have:

$$\lim_{\phi\to\infty} Q(t) = 77w(\frac{1}{1-B}) - 19w$$

Factoring out w, we get $w(\frac{77}{1-B} - 19)$. □

Lemma 3.1.3. *When $B > 1$ the total population goes to ∞.*

Proof. As t tends toward infinity, we see that ϕ also approaches infinity. Thus, $\lim_{t\to\infty} T(t) = \lim_{\phi\to\infty} T(t)$. This means that B^{ϕ} approaches infinity since $B > 1$. Thus

$$\lim_{\phi\to\infty} w(\frac{1-B^{\phi+c}}{1-B}) = \infty$$

Thus $Q_0(t) = \infty - 19w = \infty$ □

Lemma 3.1.4. *When $B = 1$ the total population goes to ∞.*

Proof. Here we see that our equation breaks down when B = 1. This means every person creates exactly 1 child. The person that made the child then proceeds to die. Thus the population is not affected. However, since w people enter every year, tw people will have entered by year t. Thus as t approaches infinity tw approaches infinity which means that the total population approaches infinity as the death rate is only enough to counteract the birth rate.

□

3.2 Expected Number of Cities

The expected number of cities at any given time interval is defined as C_t. We define C as $\lim_{t\to\infty} C_t$.

Lemma 3.2.1. *The average distance between two random points in the simulation is*

$$\frac{200}{3}(2+\sqrt{2}+5ln(1+\sqrt{2})) \qquad \textbf{Equation 3.2.1.}$$

Proof. We will first demonstrate the average distance in a unit square. This can be represented from the distance formula, a derivation of the Pythagorean theorem.

$$\int_0^1\int_0^1\int_0^1\int_0^1 \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}\, dx_1dx_2dy_1dy_2$$

We can simplify this by considering the distances in the x and y directions, delta x and delta y. To do this, we have to consider the change in the probability distributions of $x_1 - x_2$ and $y_1 - y_2$. This is known to be a triangular distribution, where x, the difference in x coordinates, will have distribution of 2*(1-x). A similar distribution is seen for y, the difference in y coordinates. This results in the following integral

$$\int_0^1\int_0^1 \sqrt{x^2+y^2} * 2(1-x) * 2(1-y)\, dxdy$$

We now solve this integral through a change to polar coordinates. We will let $x = rcos\theta$ and $y = rsin\theta$. Instead of integrating over the whole unit square, we will only integrate over the lower half triangle, and then double the value of this integral. θ ranges from 0 to 45, while the radius ranges from 1 to $1/cos\theta$, or $sec\theta$. We also need to include an extra factor of r when changing our coordinates. This results in the following integral

$$8\int_0^{\frac{\pi}{4}}\int_0^{\sec\theta} \sqrt{(r\cos\theta)^2+(r\sin\theta)^2} * (1-r\cos\theta) * (1-r\sin\theta)\, rdrd\theta$$

Evaluating this gives the value $\frac{2+\sqrt{2}+5*\ln(\sqrt{2}+1)}{15}$. However, we do not use a unit square, and we must thus multiply this result by the length of our square, 1000. This results in $\frac{200}{3}(2+\sqrt{2}+5ln(1+\sqrt{2})) \approx 521.4$, which is what we intended to prove.

□

Lemma 3.2.2.

$$C = \frac{1}{1 + C_0 * (\frac{T(t)^\alpha}{{C_i}^{\alpha-2}}) * r_a}$$

Proof. We define C_i to be the current number of cities in any year t. Then, on average, city population = $\frac{Q(t)}{C_i}$. Additionally, the probability that one person begins a new city is given by the following:

$$\frac{1}{1+\sum(C_0 * r^{-\beta} * (\frac{Q(t)}{C_i})^\alpha)}$$

The summation sums the term for every current city, thus we have: $\sum(C_0 * (\frac{Q(t)}{C_i})^{\alpha}) = C_0 * \alpha * (\frac{Q(t)}{C_i})^{\alpha}) = C_0 * (\frac{Q(t)^{\alpha}}{C_i{}^{\alpha-1}})$ This simplifies the previous equation to:

$$\frac{1}{1 + C_0 * (\frac{Q(t)^{\alpha}}{C_i{}^{\alpha-1}}) * \sum r^{-\beta}}$$

Thus, the only remaining unknown in the above equation is $\sum r^{-\beta}$. Let r_S denote the average closest distance to a city from the spawning point of a new immigrant. Let r_L denote the average largest distance to a city from the spawning point of a new immigrant. The following average value integral can be used to find the average of $r^{-\beta}$, which we define as r_a:

$$r_a = \frac{\int\limits_{r_S}^{r_L} r^{-\beta}}{r_L - r_S}$$

An estimation of r_S can be derived as follows:

On average, the C_i cities will be uniformly distributed across the simulation (a square) with the center of the distribution the center of the square. If we create congruent circles around each center, we see that r_S is equal to the radius of the circles when the sum of the areas inside the circles is equal to the sum of the areas outside the circle. Thus, r_S can be determined by writing the equation: $C_i*\pi*r^2 = 10^6 - C_i*\pi*r_S^2$. This results in $2C_i*\pi*r_S^2 = 10^6$. This equation is just an estimate as we are dividing the grid into 2 regions but we are not accounting for the circles going outside the region. When there are many cities, the circles will be small, so the area going outside will be negligible. Solving for r_S, we get $r_S^2 = \frac{10^6}{2C_i*\pi}$. Taking the square root of both sides, this yields $r_S = \frac{10^3}{\sqrt{2C_i*\pi}}$

An estimation of r_L can be derived as follows:

The largest r_L can be determined by considering going a small distance h along the diagonal away from a corner. The probability of finding a point in the corner triangle with height h is $C_i * h * 2$, and such a point is at least $x = \sqrt{2} - 2h$ away from a point in the same triangle at the opposite corner. If D is the maximum distance, we have for small enough h

$$Pr[D > \sqrt{2} - 2h] \approx C_i h^2$$

We now rewrite $h = (\sqrt{2} - x)/2$, to get the approximate distribution function

$$F(x) = Pr[D \leq x] \approx 1 - \frac{C_i}{4}(\sqrt{2} - x)^2$$

with the lowest acceptable x_0 where $F(x_0) = 0$ or $x_0 = \sqrt{2} - 2\sqrt{n}$. Integrating gives the expectation

$$E[D] \approx \int\limits_{\sqrt{2}}^{x_0} xF'(x)dx = \sqrt{2} - \frac{4}{3\sqrt{C_i}}$$

However, our unit square is of side length 1000. Thus, the answer must be scaled by a factor of $1,000$. This results in a final answer of

$$E[D] \approx 1000\sqrt{2} - \frac{4000}{3\sqrt{C_i}}$$

After plugging in the determined values for r_S and r_L into the average value integral, the average value of $r^{-\beta}$, r_a, is yielded. Clearly, $\sum_{C_i} r^{-\beta} = C_i * r_a$ Thus, the estimation for the expected number of new cities given a year and the current number of cities is modeled by the following equation:

$$\frac{1}{1 + C_0 * (\frac{T(t)^\alpha}{{C_i}^{\alpha-2}}) * r_a}$$ **Equation 3.2.2.**

□

3.3 *Unbalanced Growth*

Let a city be called dominant if the rate of growth of that city is greater than the rate of growth of every other city combined at a given year. A simulation experiences unbalanced growth if and only if there exists a year t_0 such that there exists a city which is dominant for every year $t > t_0$. In this section, we aim to find under which conditions a simulation experiences unbalanced growth. To avoid tedious calculations, we will use $Q'(t)$ rather than $Q(t)$.

If there exists a year t_0 that satisfies the condition for a simulation to experience unbalanced growth, then there exists a year t_1 that satisfies the same condition such that $t_1 > 77$. Thus, for ease of calculation, we will analyze the simulation after year 77 and we have that $Q'(t) = Q_2'(t) = tw(\frac{1-B^{t/30}}{1-B})$.

Let $S(t)$ represent the population of a single city and $A(t)$ represent the population of all the cities. $A(t)$ simply equals $Q'(t)$. Then, a city is dominant if $\frac{dQ_2}{dt} - \frac{dS}{dt} < \frac{dS}{dt}$. By definition, a city experiences unbalanced growth if there exists a year t_0 such that $\frac{dQ_2}{dt} - \frac{dS}{dt} < \frac{dS}{dt}$ for all years $t > t_0$.

$\frac{dS}{dt}$ can be split up into three components: the number of people being born within the dominant city, the number of immigrants entering the dominant city, and the number of people leaving the dominant city.

We choose t_0 as the year where the populations of all other cities can be considered dominant relative to the dominant city.

3.3.1 Number of immigrants entering the dominant city

We know that each immigrant is assigned to city j with probability P_j (eq 2.2). We see that I_j (eq 2.1) increases as N_j increases when $\alpha \geq 1$. Thus, as the population grows out of control, I_j will similarly tend to infinity. Thus, P_j tends to $\frac{\infty}{1+\infty}$, evaluating to 1. Thus, each person will have probability of essentially 1 of being assigned to the dominant city, making the immigration rate to the dominant city equal w.

3.3.2 Number of people being born in the dominant city

The number of people being born in the dominant city $w(\frac{1-B^{t/30}}{1-B})$.

We see this holds because the above equation represents the total number of babies being born out of all the cities. However, due to the dominant city's unbalanced growth, it has essentially all the population, and thus essentially all the births. Thus, the above equation also holds for the number of people being born in the dominant city.

3.3.3 Number of people leaving the dominant city

Using the equation behind Figure 1, we can calculate that under an assumption of uniform population distribution from ages 0 to 90, the average fitness of the city is 48.16.

In our model we have that m = The percentage above a person's fitness that the total average fitness must be in order to decide to move to a different city. For simplicity, redefine m as $1 + \frac{m}{100}$ for this section. By definition, a person will leave the dominant city if their fitness is greater than $48.16 * m$. Given that the fitness function is $f = 100e^{-\frac{(a-45)^2}{625}}$, we set the equation to $48.16 * m$ and find that a person leaves the city if their age is between $-25 * \sqrt{\ln \frac{100}{48.16*m}} + 45$ and $25 * \sqrt{\ln \frac{100}{48.16*m}} + 45$. The number of people leaving the dominant city will be the number of people who's age is between these two values. Finding this number is similar to finding total population (see section 3.1). Using the fact that the number of people entering the simulation is about $w(\frac{1-B^{t/30}}{1-B})$, we can estimate the number of people entering the city between those two ages to be $50 * \sqrt{\ln \frac{100}{48.16*m}} * w(\frac{1-B^{(t-45)/30}}{1-B})$. Thus, the number of people leaving the dominant city is $50 * \sqrt{\ln \frac{100}{48.16*m}} * w(\frac{1-B^{(t-45)/30}}{1-B})$.

Thus, we have:

$$\frac{dQ_2}{dt} = \frac{w*(30-30B^{\frac{t}{30}} - tB^{\frac{t}{30}} \ln B)}{30(1-B)}$$

$$2\frac{dS}{dt} = 2w * (\frac{1-B^{\frac{t}{30}}}{1-B} + 1) - 100 * \sqrt{\ln \frac{100}{48.16*m}} * w(\frac{1-B^{(t-45)/30}}{1-B})$$

If $\frac{dQ_2}{dt} < 2\frac{dS}{dt}$, then a city will remain dominant and the simulation will experience unbalanced growth. However, a dominant city must first emerge, which can be highly random due to the probabilistic nature of the simulation. However, a higher α, a lower β, and a higher m will clearly result in a higher chance of a dominant city emerging.

4 SIMULATION

In order to determine the behavior of the model, a computer simulation was created. Using C++ and the Object Oriented Paradigm, the simulation employed classes such as *Person*, *City*, and *World*. Each of the classes is designed to monitor a certain aspect of the model, such as the *Person* class controlling age, *City* class collecting average fitness and life expectancy, and the *World* class intertwining all of the other classes in order to move people and keep track of the growth of the cities.

4.1 *Random Number Generator*

A limitation that modern computers have is the ability to generate true random numbers quickly. For the simulation, a pseudo-random number generator introduced in the C++11 standard was employed. It is based on the Mersenne Twister Algorithm and our implementation has a period of $2^{19936} - 1$, making it, for our purposes, a true random number engine.

4.2 *Data Collection*

Our simulation saves certain statistics regarding the growth of the model, including *Total Cities*, *Active Cities*, *Total Population*, *Average Fitness*, *Average Age of Death*, and *Total Age*. Each active city's *Total Population*, *Total Kids*, *Average Fitness*, *Life Expectancy*, *Average Adult Age*, and *Average Kid Age*. This data is saved in the form of a ".csv" file that can easily be viewed and analyzed in a tabular format in Microsoft Excel or Google Sheets.

4.3 *Multi-threading*

Simulating large processes like our model requires a lot of computational time and power. Gaining enough data to verify results becomes virtually impossible as the amount of entities in the simulation reaches large numbers. To verify out results, we needed at least 500 years of data across over 100 different input parameter values, which would be computationally impossible to complete in a reasonable amount of time. Because the simulation model is serialized, a custom threadpool was employed. The computer that was used to simulate the model had 4 cores, so the 4 threads were created in the threadpool. Each simulation was run on a different thread, which effectively ran 4 simulations at once.

4.4 *Input Variation*

Certain inputs were kept constant, while one was varied. The inputs that were kept constant were assigned their value based off of the base case which is as follows:

Input	α	β	B	C_0	c	m
Base Value	1	1	1.05	1000	1000	50

Due to time and computing restraints, the simulation was run for 500 time intervals.

4.5 *Simulation Process*

Input Processing In order to create a streamlined and efficient simulation process, the inputs were loaded into the simulation via a text file rather than user input. The first line of the text file contains a number, n, that signifies the number of test cases that will be simulated. The following n lines contain unique test cases that will be simulated. The inputs have been described above, so there is no need to provide redundant information. The inputs are used to initialize the World object for each specific test case so that the simulations can begin. Following the input assignments, the World object is fully initialized.

Threading Once a World is initialized, it is given to a method which runs the simulation. The method is then inserted into a queue for a worker thread to run. Once a thread is available, it will run the simulation, meaning that multiple simulations can run concurrently.

Simulation process The process of updating each year in the simulation is as follows: New immigrants are entered into the simulation with the random number generator engine. Fitness for each person in the world is updated followed by the moving mechanic. Data for that year is stored and the process is repeated for 500 years.

Output All of the essential data of the simulation is tracked and stored in a specific data structure. Following the conclusion of the simulation, the data structure and its contents are passed to an output handler which writes the

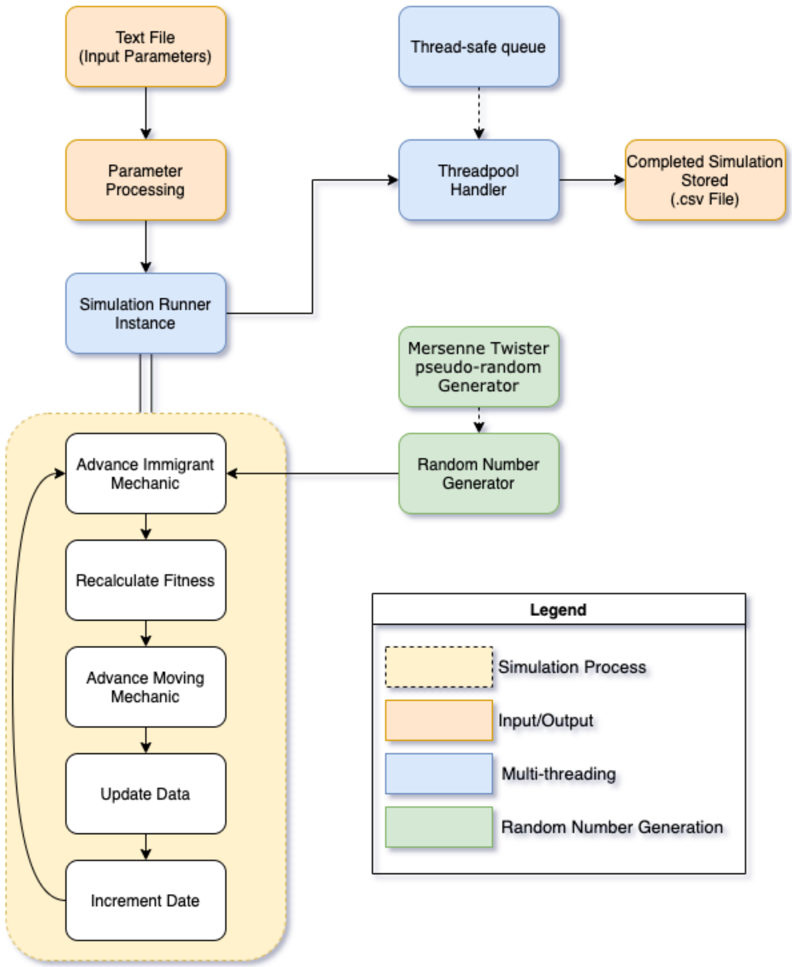

Figure 2. High-level Block Diagram

data into a .csv (comma separated list) file. In order to view the results of the simulation, the .csv file can be opened as either a Google Sheet or a Microsoft Excel project.

5 MATHEMATICAL MODEL COMPARISON

Every comparison between the simulation and the mathematical model is based upon whether or not the mathematical conclusions match the data produced by the simulation for the 7 base cases. The table below outlines the 7 base cases that are run on the simulation:

Test Number	α	β	B	C_0	c	m
1	1	1	1.05	1000	1000	50
2	**0.1**	1	1.05	1000	1000	50
3	1	**0.1**	1.05	1000	1000	50
4	1	1	**1.15**	1000	1000	50
5	1	1	1.05	**100**	1000	50
6	1	1	1.05	1000	**100**	50
7	1	1	1.05	1000	1000	**5**

5.1 *Total Population*

Figure 3 demonstrates that Q(t) is an accurate representation of the population growth in the model. The seven base cases were simulated and the relationship between the total population and the time (in years) was graphed. Q(t) was used to make predictions of the population for each given base case, and the resulting data of Q(t) was graphed on the respective base cases' simulation data. As shown in Figure 3, the red line, which represents the Q(t) prediction, follows the trend of the data modeled by the blue dots (total population data from the simulation) for each of the given base cases. Thus, Q(t) can be considered accurate mathematical model for the population growth exhibited in the simulation.

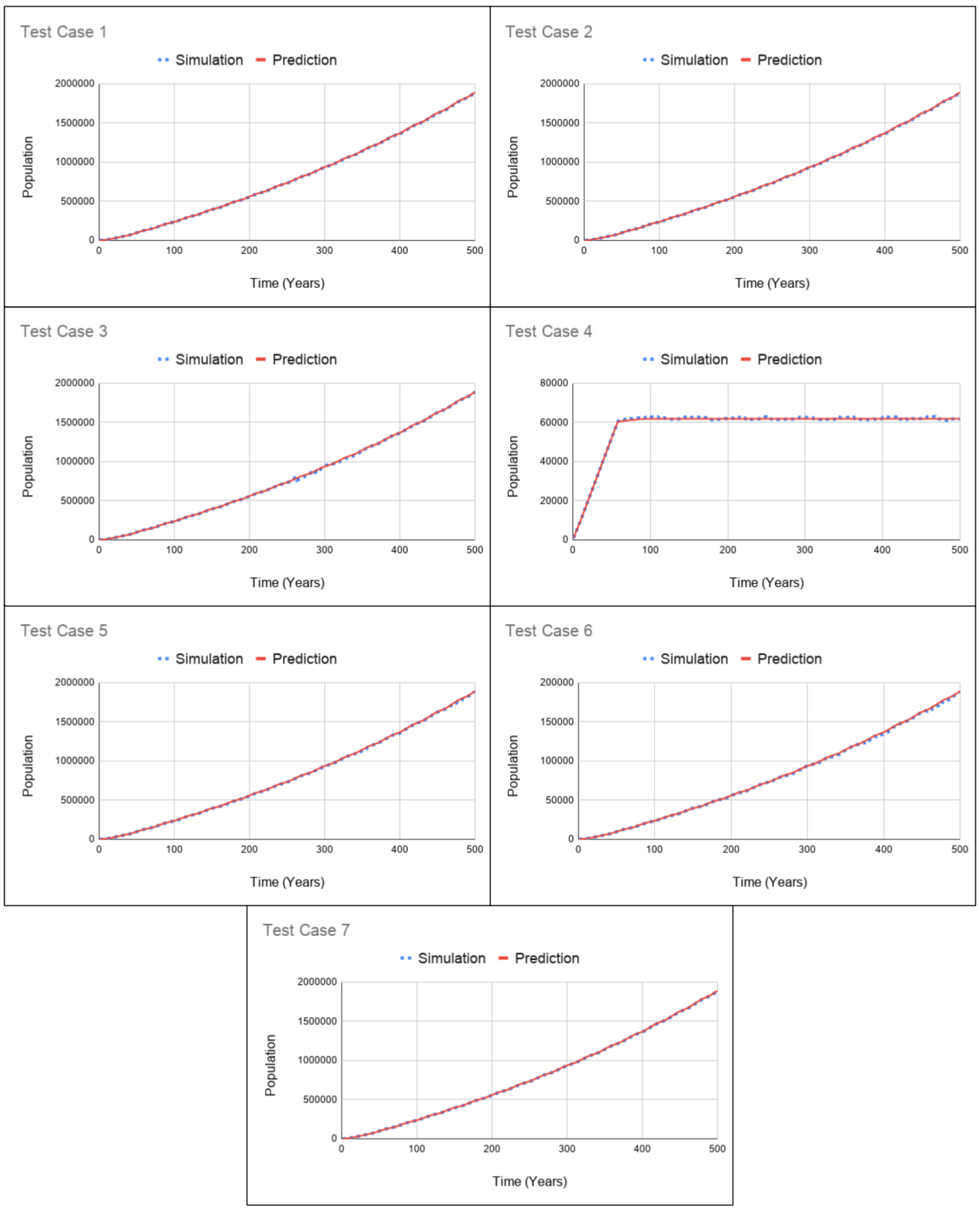

Figure 3. Total Population Comparison - Simulation vs. Formula-Based Prediction

5.2 *Life Expectancy*

The prediction matches with the simulation results. Test 2, as seen in the graph, is clearly not at the 77 years asymptote, but it has an upward trend through the entire graph. If the simulation was ran for longer than 500 years, the Age of death will asymptote at 77 year as well.

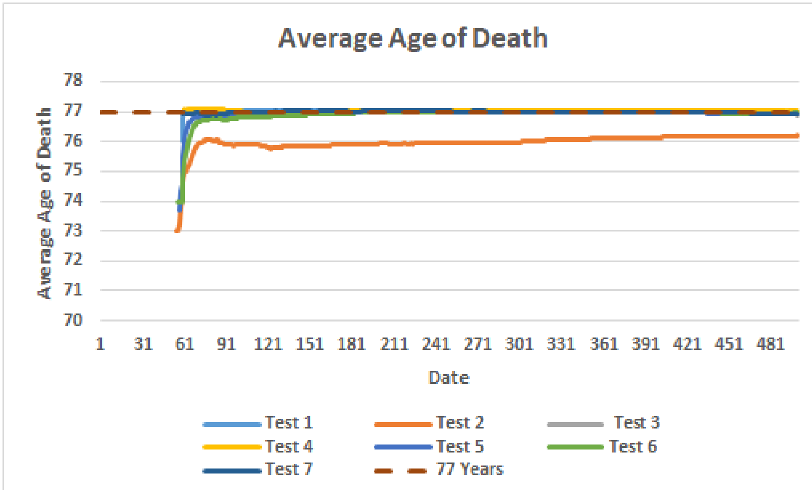

Figure 4. Average Age of Death

5.3 *Average Distance Between Two Points*

In expected cities, we estimated the average distance between two randomly generated points to be $\frac{200}{3}(2 + \sqrt{2} + 5ln(1 + \sqrt{2})) \approx 521.4$. This graph shows that our calculation matches simulation results, with the simulation approaching about 523. This result will hold true for every test case as the inputs of the simulation do not affect the generation of random points.

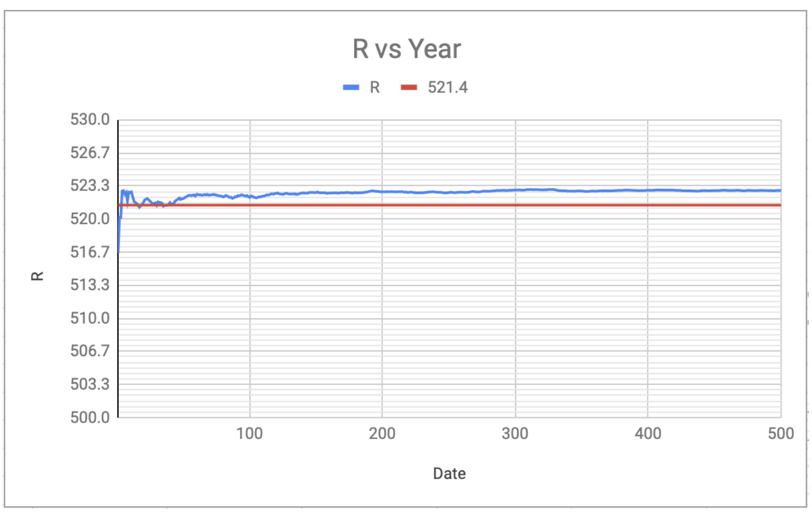

Figure 5. Average Distance Between Two Points

6 OTHER MODELS

The model in this paper focuses on the interactions that occur when the population moves within the bounds of the model itself, along with the behavior that new members of the simulation have when immigrating to a city.

Models of populations have become a topic of a lot of research in the years, and here are a two models which have other interesting interactions.

6.1 *Population Model with Catastrophe*

There is potential for many complex interactions to be added onto the framework of the simulation in order to vary the behavior. "Variants of such models are introduced allowing logarithmic, exp-algebraic or even doubly exponential growth" [3]. In Goncalves, Huillet, and Locherbachs' paper, a "catastrophe" is added which shrinks population by random amounts, even causing extinctions in parts of the simulation [3]. Models such as that can help understand natural phenomenon, such as the growth of bacteria and the effect that cleaner have on them.

6.2 *Population Extinction on a Random Fitness Seascape*

In the model of our paper, fitness of the population was a organized process with a bell curve which was similar to behavior in nature. In the paper by Ottino-Loffler and Kardarm, Random Fitness Seascape, a model is created that uses a random noise to determine people fitness at any given location and time. [4] Models such as that can help extend our simulation results to other environmental population models. The ability to characterize population distribution has far reaching effects in ecology, evolutionary biology, epidemiology, among many other fields. This can allow a prediction and protection of various ecosystems.

REFERENCES

[1] Aldous D and Huang B. "A spatial model of growth and formation" September 2012. https://arxiv.org/abs/1209.5120

[2] Silvia (https://mathoverflow.net/users/132350/silvia), Mean maximum distance for N random points on a unit square, URL (version: 2018-12-11): https://mathoverflow.net/q/317410

[3] Branda Goncalves, Thierry Huillet, and Eva Locherbach. "On Population Growth with Catastrophes" July 2020. https://arxiv.org/abs/2007.03277

[4] Bertrand Ottino-L¨offler and Mehran Kardar. "Population Extinction on a Random Fitness Seascape" *Department of Physics, Massachusetts Institute of Technology* https://arxiv.org/abs/2007.11749

[5] J. Antonio Perez (https://math.stackexchange.com/users/242788/j-antonio-perez), Average distance between two randomly chosen points in unit square (without calculus), URL (version: 2015-05-22): https://math.stackexchange.com/q/1294800