Mathematical Modelling and Simulation of Logistic Growth: Theory Versus Experiment.

Stefano Faccio
DISI
University of Trento
Trento, Italy
stefano.faccio@studenti.unitn.it

Camilla Pelagalli
DISI
University of Trento
Trento, Italy
camilla.pelagalli@studenti.unitn.it

***Abstract*—We propose a reproducible pipeline of work, consisting in the time-driven simulation of discrete logistic growth based on the corresponding master equation, focusing on demographic variation under a carrying capacity limit. The mathematical modelling that leads to the stochastic implementation is presented in a step-by-step fashion to statistically ground the designed simulation. Tuning of the main parameters is performed to choose an accurate single simulation scenario, on which output analysis and benchmarking with the state-of-the-art Gillespie algorithm is conducted, to assess its statistical significance and discrepancies between the continuous and discrete logistic growth models.**

***Index Terms*—logistic growth, population, continuous time, discrete time, deterministic model, stochastic model, Gillespie algorithm, parameter tuning, analysis of residuals.**

I. Introduction

The term *population* stands for a group of individuals that belong in the same species and live in the same area, while *population ecology* is the study of how *biotic* (living) and *abiotic* (non-living) factors influence the *density*, *dispersion* and *size* of a population [1].
The population *size* is the number of individuals in the population, *density* is the number of individuals per unit area or volume, and *dispersion* is the pattern of spacing among individuals within the boundaries of a population [1].
The effects of *density* dependent factors, such as competition for resources (food, water, light, shelter), predation, disease, and waste accumulation, are biotic factors, depending on the behavior of other living beings, to be distinguished from *density independent* factors, such as natural disasters (forest fires, floods), which can be considered abiotic, depending on *Nature* [1].
Our project consisted in the discrete-time event simulation of limited, or *density* dependent, population growth, motivated by the relevance of this topic in the context of epidemics.
Our aim was to reproduce and propose a complete pipeline of work, from the mathematical modelling of the logistic growth master equation to its translation into a stochastic simulation, and subsequent parameter tuning and output analysis, trying to shed light on the discrepancies between continuous and discrete growth, and between deterministic and stochastic growth, which tend to be implicit and scattered around the available literature and experiments.
We tried to adopt an interdisciplinary approach, bridging mathematics, statistics, computer science and a glimpse of biology, combined with a step-by-step didactic imprint, to help a potential undergraduate student delve into the complexity and pitfalls of logistic growth modelling.

II. Related work

The logistic growth model is an essential analytical tool for ecologists and demographers, whose application can be extended to stock markets and forecasting models [2].
Recent papers examine the applicability of the logistic growth model for the study of the COVID-19 pandemic and other infectious diseases [3][4] and to evaluate mutation proportion in epidemics [5].
Online deterministic simulators (https://www.biointeractive.org) are available to explore logistic growth from a theoretical point of view and stochastic simulations are presented in the context of biological evolution (https://www.youtube.com/@PrimerBlobs).
The Gillespie algorithm [6] has become the standard for generating time-evolution trajectories of finite populations in continuous time [7] and it was employed as a reference to benchmark our simulation.

III. Problem definition

A. A deterministic model of unlimited growth

The science of ecology refers to the *Malthusian law* as the first founding principle of population dynamics [8]. The Malthus model is applicable to an isolated population of individuals, not interacting with other populations, with infinite food and space resources, hence growth is unlimited, and the rate at which the population changes is proportional to the population itself: $\frac{dN}{dt} = kN$. The parameter k is the *relative* growth rate (percent per unit time), to be distinguished from the *absolute* growth rate, $\frac{dN}{dt}$ (people per unit time) [2].
When the initial population is large enough and its evolution is observed for a long period of time, population growth can be mapped as a *continuous* variable and explained by the *exponential model*: (1) $N_t = N_0 e^{kt}$, whose *discrete* version is the *geometric progression*: (2) $N_t = N_0(1+r)^t$, where $r = e^k - 1$ [9][2].
The difference between these two cases resides in the value

of the base and the interpretation of the *continuous* and *discrete* growth rates, represented by the parameters k and r, respectively. As shown in Table I, if $r < 0.1$, *continuous* growth can be approximated with *discrete growth*.

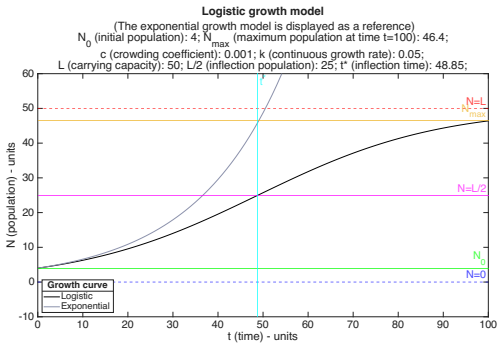
Discrete growth	Continuous growth
$N_t = N_0(1+r)^t$	$N_t = N_0 e^{kt}$
$r = 1$	$k = ln(1+1) = 0.69$
$r = 0.5$	$k = ln(1+0.5) = 0.40$
$r = 0.2$	$k = ln(1+0.2) = 0.18$
$r = 0.1$	$k = ln(1+0.1) = 0.095$
$r = 0.05$	$k = ln(1+0.05) = 0.0488$
$r = 0.02$	$k = ln(1+0.02) = 0.0198$
$r = 0.01$	$k = ln(1+0.01) = 0.0099$

TABLE I: Comparison of discrete (r) and continuous (k) growth rates. Ranges: $0.01 \leq r \leq 1$ and $0.0099 \leq k \leq 0.69$, respectively.

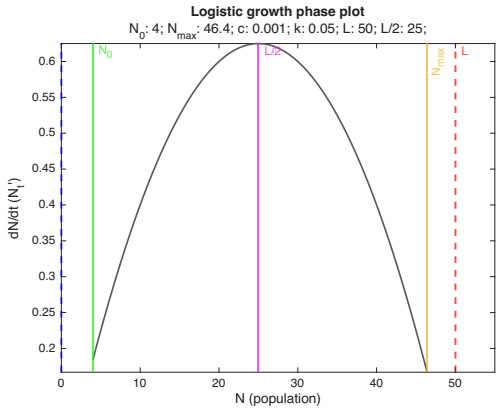
In general, if $r = 0$ and $k = 0$, the population remains constant; if $r < 0$ and $k < 0$, the population tends to become extinct, and if $r > 0$ and $k > 0$ the population "explodes" as a long-term behavior. Malthus himself highlighted that, if unchecked, the population would grow exponentially, whereas the food supply would grow linearly, leading to the population outstrip of the food supply [9].
The Malthusian law provided a good approximation of world population and that of many regions for decades and even centuries. Nevertheless, a growth without boundaries is unrealistic: population is indeed regulated by the availability of resources and damages caused by war and disease.

B. A deterministic model of limited growth

The Belgian mathematician Pierre F. Verhulst, under the guidance of Adolphe Quetelet, proposed a differential equation[1], known as the *logistic* equation, that takes into account the upper limit to population growth : $\frac{dN}{dt} = kN\left(\frac{L-N}{L}\right)$ [11]. L is called the *carrying capacity* of the environment, representing the largest population that can be supported, while k is the *continuous* growth rate.
The logistic growth equation can be obtained from the exponential growth equation, $\frac{1}{N}\frac{dN}{dt} = k$, assuming that the relative growth rate of the population is a linearly decreasing function of N: $\frac{1}{N}\frac{dN}{dt} = k - cN$.
The constant c act as a *crowding coefficient*: the greater it gets ($0 < c \leq \frac{k}{N}$) the smaller $\frac{dN}{dt}$ becomes, to account for the limiting resources of an over-crowded environment. When $c = \frac{k}{N}$, there is no total change in population size. Solving for N, we obtain: $N = \frac{k}{c} = L$, the *carrying capacity* [2].
The analytical solution to the logistic equation (Figure 1a) is the *sigmoid* function: $N_t = \frac{N_0 L}{(L-N_0)e^{-kt}+N_0}$.
The first derivative N_t' (Figure 1b) is positive, and increasing for $N_0 < N < L/2$, resulting in the sigmoid function being concave up. For $L/2 < N < N_{max}$, the slope $\frac{dN}{dt}$ is negative and N_t' is decreasing, making the sigmoid function become concave down [2][12]. The inflection time $t^* = \frac{\ln(\frac{L-N_0}{N_0})}{r}$ and $N^* = L/2$ are respectively the x and y coordinates of the *inflection point*, obtained by setting $N_t'' = 0$ [13].



(a) Population size N growth as a function of time t for $L = 50$ and $N_0 = 4$.
When $t = 48.85$ and $N = 25$, the logistic function grows the fastest and then flips its concavity, growing slower until L is reached.



(b) Growth rate dN/dt as a function of population size N for a population with $L = 50$.
The inflection population is where the *acceleration* changes from increasing to decreasing, i.e. the second derivative is zero at the top of the hump.

Fig. 1: Continuous model of logistic growth.

The discrete version of logistic growth:

$$N_{t+1} = N_t\left(1 + r\left(1 - \frac{N_t}{L}\right)\right)$$

also called *master* equation, has no general closed form solution and it was the reference method for our discrete-time event simulation. The ecologist Robert May studied this equation for populations and discovered that it could produce very complicated dynamics [14]. However, by keeping r small enough (Table I), population is prevented from changing in size rapidly over short periods of time and rounding errors are minimized (Figure 2) [15].

IV. System Setup

A. A stochastic model of limited growth.

The simulation system is defined as a set of synthetic creatures placed on a remote island with limited resources and

[1]The original equation is proposed in the *Recerches mathématiques sur la loi d'acroissement de la population* [10, p. 8].

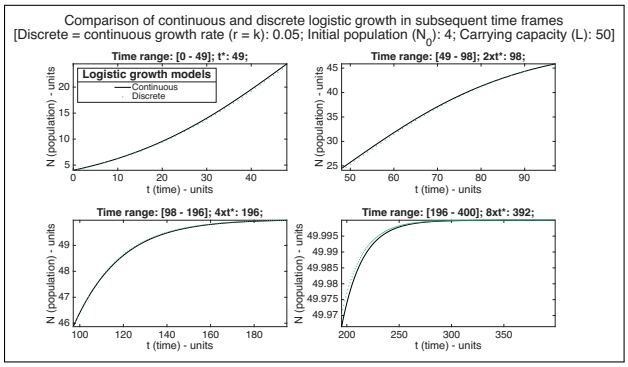


Fig. 2: Approximation of continuous logistic growth with a discrete model: subsequent time frames.
With r set to 0.05 and the example parameter settings, discrete population values respectively underestimate and overestimate continuous values in the time ranges $[t_0 - 2t^*]$ and $[2t^* - 6t^*]$, then both model realizations converge to L.

the population growth is monitored over time.
Two entities can be identified: the *environment*, which is *passive*, since it is represented by the constraint of *the carrying capacity*, and the *creature*, which is *active*, hence a *process*, being able to *replicate*, causing the "birth", or appearance, of another creature, and *die*, decreasing the overall population. Replication and death represent the simulation *events* and the *state variable* under observation is the current size N of the population as *density*, being the unit area an island, at each time step, or iteration.
We assumed an initial population N_0 and R and D to be the per-capita *replication* and *death rates* per unit time as a function of N, respectively.
A member of a population, or creature, A replicates when they instantiate a copy of themselves: $A \rightarrow A + A$, and die disappearing from the environment: $A \rightarrow \emptyset$. These are the simulation *activities*.
The change of the population size (ΔN) during a time interval (Δt) is then given as:

$$\Delta N = N_{t+\Delta t} - N_t = (R\ \Delta t\ N) - (D\ \Delta t\ N)$$

By letting $\Delta t \rightarrow 0$, we get the *absolute growth rate*, or the overall expected change of a population:

$$\frac{dN}{dt} = (R - D)\ N \qquad \left(\leftrightarrow \frac{dN}{dt} = kN\right) \qquad [2]$$

$R - D$ is the net per-capita *increase rate* per unit time [16].
However, population has to plateau in accordance with the criteria of limited growth: the expected change per time step has to go to zero when the maximum number of creatures that the environment can support is reached. An additional term c has to be included to adjust the *overall death rate* based on *crowding*:

$$\frac{dN}{dt} = (R - D - c\ N)\ N \qquad \left(\leftrightarrow \frac{1}{N}\frac{dN}{dt} = k - cN\right) \qquad [3]$$

[2]Deterministic continuous exponential growth model (III-A).
[3]Deterministic continuous logistic growth model (III-B).

$c\ N$ is small when there are few N creatures and big when too many creatures are sharing the environment.
If we substitute k with r, to account for the approximation of continuous growth with discrete growth, a prerequisite for discrete-time event simulations, we can make the following equivalences: $R - D = r$, $\quad c = c$, $\quad \frac{R-D}{c} = L$.
By turning R and D rates into *replication* and *death probabilities* respectively, and setting $c\ N$ as *death probability adjusted for crowding*, we got the stochastic components of our simulation. We treated them as random variables and employed rejection sampling to draw them as samples from the uniform distribution, at each time step. The corresponding *threshold probabilities*, to compare the sampled ones with, are the *deterministic variables* of the system, that we set with predefined values, consistent with the deterministic model of logistic growth, and resulting from parameter tuning.
Considering a single creature, if the drawn probability for R is smaller than the threshold, then that creature replicates; D and cN are summed as a common death threshold probability, so if the second drawn probability is smaller than their sum, that creature dies. If both probability samples satisfy the inequalities, then "nothing happens". The event of "nothing happens", or absence of change per creature, can also manifest if both probability samples do not satisfy the inequalities. This event could be considered the third *(in)activity* of the simulation system.

B. Pseudo Random Number Generator testing

Since the simulation output is relative to multiple runs, we have chosen the PRNG that implements *Mersenne Twister* 1997 algorithm, from the *Math.NET* Project (https://numerics.mathdotnet.com/Random) and we have set sequential prime numbers as seeds to the corresponding simulation runs.
Due to limited time and computational resources, we have performed a *small scale* test of the PRNG, by analyzing 100 different streams of data points, matching 100 single simulation runs, containing each 500 iterations, multiplied by the two probabilities thrown, R and cN respectively, and the number of creatures alive per iteration, that we set to 100, for a total of 100000 data points per stream.
To check the absence of correlation between different streams, we have reviewed their pairwise Spearman rank correlation coefficient matrix and 8 out of 10000 possible couples, excluding the main diagonal, resulted in $rs \geq 0.01$ or $rs \leq -0.01$. Same results were also obtained with r Pearson coefficient.
One-sample Kolmogorov-Smirnov and *Chi-square goodness-of-fit* tests indicated that 96 and 95 streams of random numbers, out of 100, followed a uniform distribution.

V. Results and Discussion

A. Parameter space exploration

We have performed the designed simulation 100 times, with different parameter settings, to explore the resulting scenarios with an unbiased approach.
We have kept r, or $R - D$, and the number of iterations fixed to 0.05 and 500 respectively: lowering the discrete growth rate

implied increasing the simulation execution time, so we chose the first r that approximated k by at least four digits (Table I) and the corresponding number of iterations that showed all the phases of logistic growth, including the long-term behaviour.
We have then decided the ranges of initial population (N_0), crowding coefficient (c) and reproduction (R) - hence death (D) - probability to be: $1 \leq N_0 \leq 51$, $0.00001 \leq c \leq 0.0001$, and $0.05 \leq R \leq 1$ ($0 \leq D \leq 0.95$), with matching steps of 2, 0.00001 and 0.05. In this way, the theoretical L shifts from 5000 to 500 creatures (by one order of magnitude) and the parameter space dimension is: 26 (values of N_0) $\times$ 20 (values of R) $\times$ 10 (values of c) $\times$ 500 (number of iterations), for a total of 26×10^5 values of $\mu(N_i)$, or iteration means[4,5].
We have taken a subset of this space, made of the last 200 iterations, when growth is already stationary, and took the (grand) mean of the corresponding values of $\mu(N_i)$ for each parameter setting, to get the observed L values[6].
We analyzed the interplay between c, N_0, R and hence D through the impact they had on the ability of the population to reach theoretical L (Figure 3). We have used the scaled absolute difference between the observed and theoretical L, ΔL, as a measure of distance of the stochastic model from the mathematical one.
We noted that lower values of N_0 (*yellow dots*) led to a higher ΔL, independently from c and R. This could be explained by the exponential phase of growth: if not enough creatures are there to replicate and balance the dead ones out, the population could go extinct very quickly. Out of 100 simulations, the majority could reach 0 as the current N_0 before even getting to the inflection time.
Extreme values of R (*blue dots*), or at least symmetrically distanced from center values (e.g. values far from $R = 0.5$ and $D = 0.45$) seemed to imply the shortening of ΔL towards zero, while no significant effect is seen on the outcome because of parameter c variation.

B. Parameter tuning

We then focused on parameter setting(s) that minimized ΔL in percentage, by looking at their pair-wise relationships (Figure 4). The scatter plot of ΔL against N_0 (4.A) confirmed that at least 30 creatures are necessary to reduce the gap between theoretical and observed L to $< 3\%$. Choosing $N_0 = 51$, the upper limit of the range, is a clear tuning outcome.
As for c and R, two box plots (4.B-C) were produced to evaluate the distribution of $\Delta L\%$ within the respective ranges. Regarding c, the 75^{th} percentile is approximately $\leq 3\%$ and the upper adjacent is $< 8\%$ for all the ten settings, with a median very close to 0%, ranging between 0.13% of $c = 0.00001$, that corresponds to a theoretical $L = 5000$, and 0.95% of $c = 0.0001$.
A greater L probably allows for a smoother growth and a more stable steady-state: the addition or subtraction of a single creature affects $L = 5000$ by a $\pm 0.02\%$, compared with $\pm 0.2\%$ when $L = 500$, the lower limit of the range.
Instead, R variation shows a roughly symmetric pattern. The values of 0.5 and 0.55 record upper adjacent limits close to 20% and 75^{th} percentiles $> 8\%$. Probabilities farther from the center yield shorter inter-quartile ranges. The extremes of $R = 0.05$ and $R = 1$ should be excluded because they respectively represent creatures not dying except for overcrowding and always replicating. The R value recording the lowest median (of 0.15%) is 0.95, followed by $R = 0.1$, with a median of 0.22%.
We cannot choose R from an absolute biological prospective: we simulated the growth of an *in silico* population, so any R and D values are feasible. However, in a real world-scenario, $R > 50\%$ is not sustainable.
Remembering that $R - D$ is kept constant and the growth rate never changes, this pattern may be due to the stochasticity of the model, so their setting cannot be tuned unambiguously or may not be tunable at all.
A speculative guess may be that values of R closer to the lower and upper extremes result in a lower probability of simulation failure (P_f), i.e. when population gets stuck into 0. For example, in a single creature and single iteration scenario,

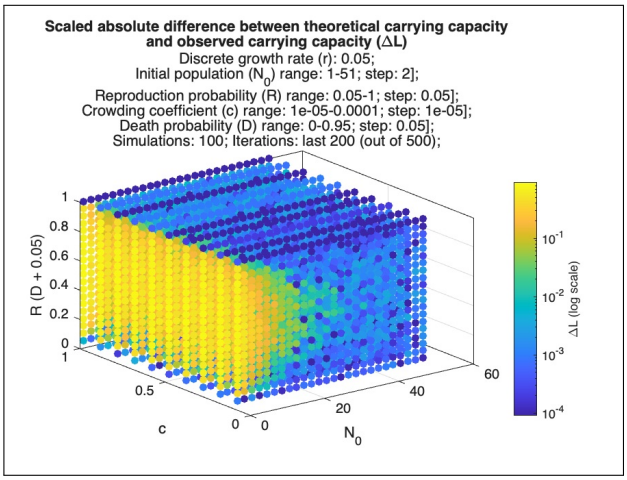


Fig. 3: Parameter space hypercube.
Each 4-dimensional point represents the absolute value of the difference between the theoretical and observed carrying capacity L, scaled by theoretical L (ΔL), for each setting of N_0, c and $R(D + 0.05)$. In this way, we obtain weighted differences, comparable with each other, given that theoretical L ranges from 5000 to 500. The 4^{th} dimension, the value of ΔL, depicted by color, is in logarithmic scale to accentuate the "yellow-ish" half-circle pattern that can be detected on each vertical section made by c, hinting at the relevance of the parameter N_0.

[4]The iteration mean $\mu(N_i)$, computed on iteration i over all the j simulations can be defined as:

$$\mu(N_i) = \frac{\sum_{j=1}^{max} N_{ij}}{|S|}$$

where $|S|$ is the total number of simulations. Taken together, the iteration means are representative of the whole trend of population evolution, each averaging the current size N of a population, at a given time unit.

[5]95% confidence intervals for each $\mu(N_i)$ were not calculated due to computational and time costs that could not be met.

[6]95% confidence intervals (computed with the Bootstrap method) of 5200 observed L (iteration grand mean) values, for each setting, are given in Supplementary Table 1.

under a value of L so great that death by overcrowding becomes negligible (by approximation), $P_f = (1-R)(D) = (1-R)(R-0.05)$. If $R = 0.95$ or $R = 0.1$ then P_f is 0.045 for both, else if $R = 0.5$ or $R = 0.55$, P_f peaks at 0.225. For $R = 0.6$ or $R = 0.45$, $P_f = 0.22$, following the same symmetric pattern as that found for $\Delta L(\%)$ in Figure 4.C).
We have thus decided to choose the setting of $R = 0.1$ to proceed with the pipeline of the output analysis, being a realistic option that also minimizes $\Delta N\%$[7], while the choice of $N_0 = 51$ and $c = 0.00001$ is motivated by the tuning results. The frequency distribution of $\Delta N\%$ across all simulations (4.D) validated the overall consistency of the stochastic model, confirming its statistical soundness.

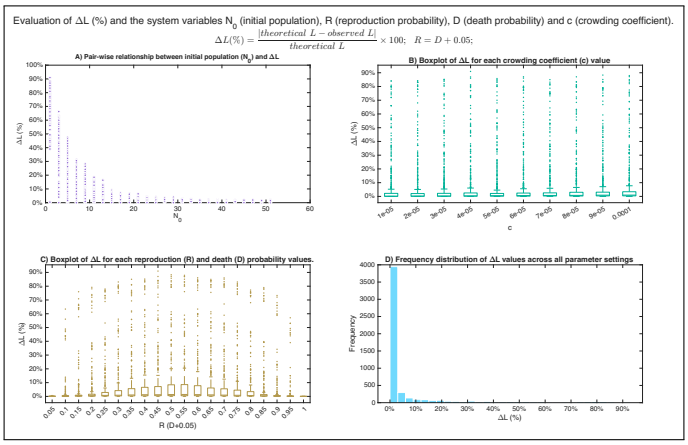
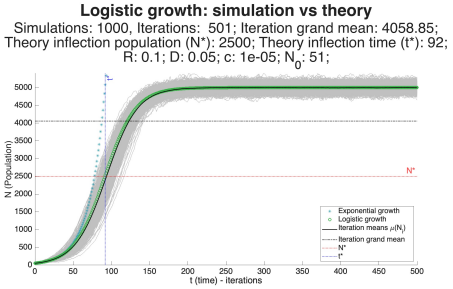


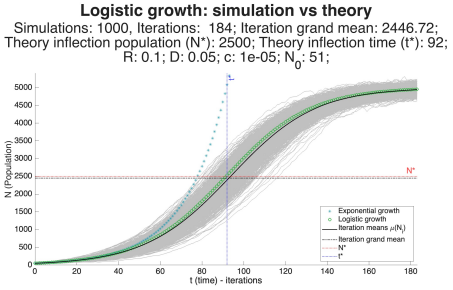
Fig. 4: Parameter tuning of $\Delta N\%$ against N_0, c and R (D). $\Delta L\%$ is used to facilitate visual comparison across plots. The $\Delta L\%$ gaps of C) box plots derive from the inclusion of all N_0 values in the range $[0-51]$.

C. Single simulation output analysis

We have performed the simulation 1000 times with the same settings resulting from parameter tuning and we focused on both the long-term (Figure 5a) and short-term (Figure 5b) behaviours of the system. The $\mu(N_i)$ values capture the trend of the simulated growth, the iteration grand mean of the long-term plot ($\sim$ 4059) gets closer to L (the carrying capacity), while it is almost at the inflection population N^* ($\sim$ 2447 with respect to $N^* = 2500$) in the short-term plot.
A closer look at the distribution of $\mu(N_i)$ and $\sigma^2(N_i)$[8] values in the two scenarios, trough box plots 6.A and 6.B respectively, reveals the impact of the *tail* on the iteration means: when the *tail* is cut, their median shifts from 4994 to 2366, matching the iteration grand mean decrease towards



(a) Long-term behaviour of logistic growth.
The steady-state phase starts approximately after $2t*$ and it is terminated by the simulation within 501 iterations. We also refer to it as *tail*.



(b) Short-term behaviour of logistic growth.
The lag, acceleration and deceleration phases, included in the time range $[0-2t*]$ are displayed, giving rise to the familiar symmetrical S-shaped curve, which we also refer to as *head*.

Fig. 5: Single simulation results: population evolution over time.

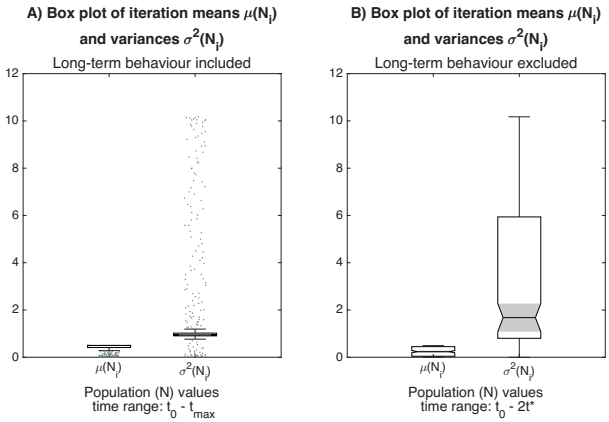


Fig. 6: Box plot of the iteration means and variances distribution. A) Long-term behaviour included vs. B) excluded. Values on the Y axis need to be multiplied by 10^4.

the inflection population. The number of outliers goes from 99 out of 501 data points (19.76%) to zero in the *head*. This is due to the iteration means belonging to the exponential phase of the logistic growth, which appear as outliers with respect to the means calculated when population approaches L.
A similar pattern can be seen also for the variance, which counts 151 outliers out of 501 data points (30.14%) in the long-term behaviour, and zero in the short-term. However, the median of $\sigma^2(N_i)$ values increases from 9648 to 16770 and the 75^{th} percentile stretches out from 10220 to 59380 (almost a sixfold). This variability seems to be flattened by values

[7]The value of R could be replaced by any other setting to carry out the output analysis, probably yielding different results. Also, the same output analysis could be applied to all the values in the R range in parallel for further comparison.

[8]The iteration variances $\sigma^2(N_i)$, computed on iteration i over all the j simulations can be defined as:

$$\sigma^2(N_i) = \frac{\sum_{j=1}^{max} \mu(N_i) - N_{ij}}{|S-1|}$$

where S is the total number of simulations.
95% confidence intervals for each $\mu(N_i)$ and $\sigma^2(N_i)$ are given in Supplementary Table 2 and 3, and they were computed with the Bootstrap method.

approaching L from above or below, at steady-state.

This observation led us to look for heterosckedasticity in the raw residuals deriving from the comparison of $\mu(N_i)$ values with the two theoretical models used as mathematical references, continuous (*CM*, Figure 7.B) and discrete (*DM*, Figure 7.C). The residuals exhibited a similar pattern, a roughly symmetric *V*-shape in the short-term, that converges around $\Delta N = 0$ in the long-term, with one main difference: the minima are ~ -138 in the *CM* and almost -48 in the *DM*, in line with the underlying the simulation implementation.

From the analysis of the 95% Confidence Intervals of the

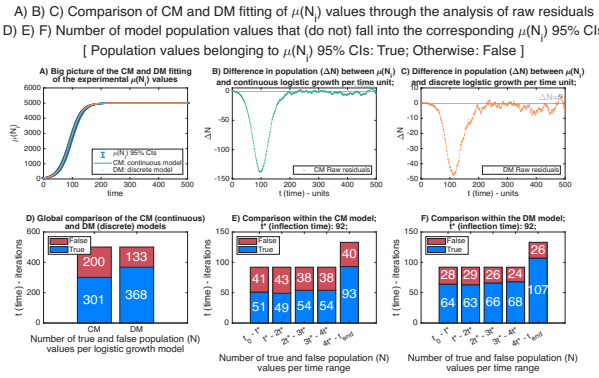


Fig. 7: Comparison between the continuous and discrete model fittings of the iteration means: raw residuals and belonging to $\mu(N_i)$ 95% CIs.
CM residuals (B) look more compact than the *DM* ones (C) in the *tail*, but the Y axis scale is different. Actually, the *DM* (F) fits the $\mu(N_i)$ CI values better than the *CM* (E). It is interesting to note that a similar number of true/false outcomes is present from t_0 to $4t^*$, also after the sigmoid shape, probably reflecting fluctuations above and below L.

$\mu(N_i)$ values, we can globally state that a higher number of population values from *DM* falls into the $\mu(N_i)$ CIs than from *CM* (Figure 7.D), and locally, considering subsequent time frames (Figure 7.E-F), the *DM* population values fits better than *CM* on both short-term and long-term ranges, especially in the latter.

We also noted that both *CM* and *DM* population points ranging from t_0 to $2t^*$ have a similar number of true/false outcomes with respect to the iteration means, validating the symmetry arising from the raw residuals.

These considerations led us to compare the two theoretical models, so we plotted ΔN values (Figure 8) between the *CM* and *DM* N values across time, making the symmetrical *V*-shape appear, with a minimum at around -100.

As a consequence, heterosckedasticity in the raw residuals of the iteration means does not imply a simulation implementation-related error, but a constitutive characteristic of our discrete-event simulation trying to keep up with continuous growth.

D. Benchmarking with the Gillespie algorithm

In order to assess the statistical significance of the $\mu(N_i)$ values obtained by our simulation, we evaluated how the same information, produced by the application of the Gillespie algorithm (exact method) [6], would compare to *CM* and the

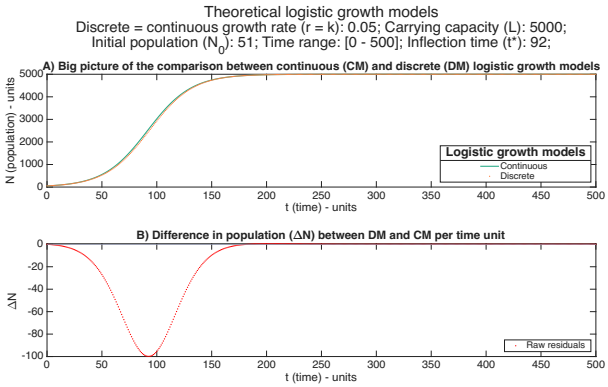


Fig. 8: Comparison of continuous and discrete deterministic models of logistic growth. The v-shaped ΔN values across time is also a characteristic of the two models. Up to t^*, *DM* increasingly underestimates *CM*, in harmony with an exponential phase of growth, which causes *DM* to round N down to the nearest integer, and afterwords, the difference is gradually re-absorbed during the deceleration phase of logistic growth. At steady-state, $\Delta N = 0$ for both models.

corresponding raw residuals.

One of the main differences between the proposed simulation and the state-of-the-art implementation of the Gillespie algorithm for logistic growth [17] resides in the interpretation of time: by taking finite unit steps in time (mirroring the discrete logistic master equation), we dictated that both replication and death may occur in the same step, updating the probability. However, these steps are not always synchronous. Birth (by replication) and death can be considered Poisson processes, meaning that the waiting time (τ) until either process occurs is exponentially distributed [18][19].

The underlying continuous-time Markov process has the following events and transition rates:

- Birth: $N \to N+1$, with rate $\gamma = rN$[9];
- Death: $N \to N-1$, with rate $\omega N = \frac{rN^2}{L}$. This rate is dynamical, depending on the current value of N[10].

The corresponding transition probabilities are given as:

- $P(N \to N+1) = \frac{\gamma}{\gamma+\omega N}$;
- $P(N \to N-1) = \frac{\omega N}{\gamma+\omega N}$,

and they are used for rejection sampling. If the sample drawn from the uniform distribution is smaller than $P(N \to N+1)$, then the creature replicates, otherwise it dies, making these two events mutually exclusive, another major difference with respect to the proposed simulation, where they can happen at the same time, resulting in no net change [20].

To determine how long (τ) it would take for a birth (or death) events to occur, a random number from an exponential distribution with a mean of $\frac{1}{\gamma+\omega N}$ is drawn[11].

As a consequence, time points are not evenly spaced across simulations. We had to extract the mean of the current N at

[9] $\gamma = rN$ accounts for $\frac{dN}{dt} = rN\left(1 - \frac{N}{L}\right)$, the first part of the logistic equation.

[10] $\omega N = \frac{rN^2}{L}$ accounts for $\frac{dN}{dt} = rN\left(1 - \frac{N}{L}\right)$, the second part of the logistic equation.

[11] cfr. Theorem on Poisson arrivals.

each time point over the individual simulations, by weighting each count by the length of time that it existed [18].
In this way, we could compute the $\mu(N_i)$ values[12] and analyze the distribution of raw residuals with respect to the continuous model for both the Gillespie simulation (Figure 9) and the discrete master equation simulation (ours, Figure 10).

The maximum raw residual of Gillespie is around -27

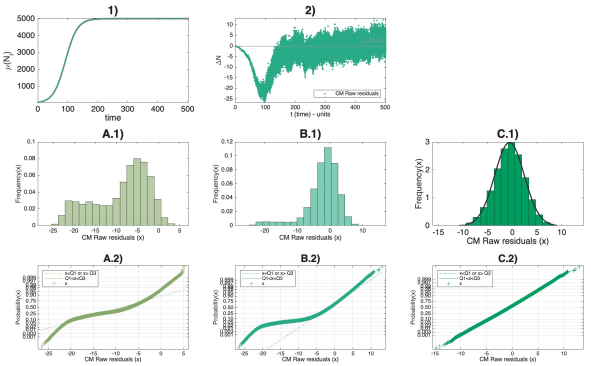


Fig. 9: Gillespie simulation. **1)** Big picture of the CM fitting of the experimental Gillespie $\mu(N_i)$ values; **2)** Difference in population (ΔN) between Gillespie $\mu(N_i)$ values and *CM* growth per time unit; **A.1)** *CM* histogram plot of the head (short-term) raw residuals, time range: $[t_0 - 2t * (\sim 184)]$; **B.1)** *CM* histogram plot of the head+tail (long-term) raw residuals, time range: $[t_0 - t_{max}(\sim 501)]$; **C.1)** *CM* histogram plot of the tail (steady-state only) raw residuals, time range: $[2t * - t_{max}]$; **A.2)-B.2)-C.2)** Corresponding *CM* normal probability plots. *NRSME* indicators, for each time window, are respectively 0.01, 0.009 and 0.38.

(Figure 9.2), a value lower than our simulation against the underlying discrete model, and of the theoretical continuous model against its discrete version. This may be explained by Gillespie correctly accounting for the inherent fluctuations and correlations that are necessarily ignored in the deterministic formulation. In addition, this algorithm never approximates infinitesimal time increments dt by finite time steps Δt, in harmony with population levels that change suddenly and sharply with time [6].
However, a similar *V*-shape of the short-term residuals confirms the presence of an unavoidable bias, probably depending on the intrinsic symmetry of the sigmoid. To further investigate this discrepancy in the residuals, we divided logistic growth into three time windows, relative to the short-term or *head* $[t_0 - 2t*]$, long-term (head and tail), and only steady-state or *tail* $[2t* - t_{max}]$ respectively, knowing how our iteration means and variances behaved in the first and second window (Figure 6).
The distribution of raw residuals was observed for both simulation approaches through histogram plots ($A.1$, $B.1$, $C.1$ rows) and normal probability plots ($A.2$, $B.2$, $C.2$ rows). Figure 10.A.1 shows greater multimodality than 10.B.1, reflecting the deletion of tail $\mu(N_i)$ values. The long-term histogram is extremely left skewed and the smaller bars match the sigmoid growth outliers. The corresponding 10.A.2 and 10.B.2 normal probability plots resemble a symmetric distribution of residuals, as the *V*-shape of Figure 7.B, and a significant horizontal deviation from the reference line, in agreement with the long histogram tail, respectively. The short-term and long-term residuals of our simulation do not exhibit a normal distribution. Only the last window (10.C.1 and 10.C.2) seem to have a close to normal distribution.

Regarding the Gillespie $\mu(N_i)$ raw residuals, noting that

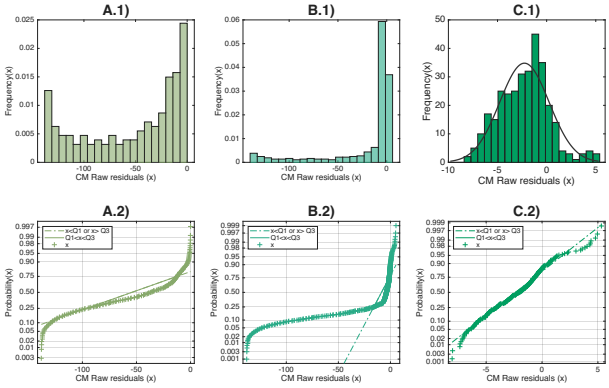


Fig. 10: Discrete master equation simulation. **A.1)** *CM* histogram plot of the head (short-term) raw residuals, time range: $[t_0 - 2t * (184)]$; **B.1)** *CM* histogram plot of the head+tail (long-term) raw residuals, time range: $[t_0 - t_{max}(501)$; **C.1)** *CM* histogram plot of the tail (steady-state only) raw residuals, time range: $[2t * - t_{max}]$; **A.2)-B.2)-C.2)** Corresponding *CM* normal probability plots. *NRSME* indicators, for each time window, are respectively 0.04, 0.03 and 0.39.

the number of points under scrutiny is much greater than our simulation, their distribution in Figure 9.C.1 and 9.C.2 is consistent with normality and also the long-term histogram shape, aside from the bottom-left tail clearly visible from the normal probability plot, tends to normality. Instead, the histogram in 9.A.1 is bimodal, probably in concordance with the acceleration/deceleration phases of logistic growth[13].
Without plotting the distribution of residuals, we would have been profoundly misled by the NRMSE indicator, which compared the iteration means with the background continuous model population values. Indeed, a low NRMSE does not exclude the possibility that a model is biased [21].
We realized that the continuous time approach of the Gillespie algorithm yields more statistically significant $\mu(N_i)$ values, but does not adjust for the inherent bias of the logistic growth continuous model: being time-series data, the error terms are correlated in the sigmoid phase of growth.

VI. Conclusions

We have successfully designed a generalizable and reproducible pipeline of work for logistic growth modeling, simulation, tuning and performance evaluation, and highlighted

[12]95% CIs are not available for the iteration means due to time and computational costs. While the Gillespie algorithm (exact method) is relatively simple to implement, it requires long periods of simulation for convergence to the steady state value. Compared with our single simulation, where we set 501 iterations, to include the long-term behaviour, Gillespie required 200000 iterations to match up.

[13]A mixture of Gaussians could explain away the bimodality of the frequency distribution.

the differences between the implementation of a *time-driven* simulation, as ours, which was mapped from the discrete master equation, and the implementation of an *event-driven* simulation, as Gillespie's, which yielded more statistically significant results, when compared to the continuous model of logistic growth. We were able to distinguish pitfalls of our simulation from intrinsic characteristics of the mathematical models thanks to an in-depth analysis that only an interdisciplinary approach, combined with a clear didactic purpose, could foster.

VII. Further improvements

To test the PRNG generator on a *global scale*, we should have produced as many data streams as the maximum number of simulation runs we carried out throughout the research, hence 1000, and to account for the number of data points within a single simulation or stream, we should have multiplied the total number of iterations (500) by the two probabilities thrown, R and cN respectively, and the number of creatures alive, at maximum $L \pm \epsilon$, per iteration, resulting in at least 5×10^9 data points to process.

Linear model fitting of the iteration means, exploiting the rewriting of the logistic equation in the form: $\frac{dN}{dt}\frac{1}{N} = k\left(1 - \frac{N}{L}\right)$, where the first term is a linear function of N, could be applied to check if the assumptions of homoskedasticity and normality of the raw residuals are fulfilled.
Parameters estimated for r and L could then be compared with the theoretical reference line $y = mx + b$, where $r = -m$ and $L = \frac{b}{r}$ [12][2].

Field observations, coming from epidemics or census data, biological results, involving bacteria competence models or sheep and harbor seals growth, could enhance the exploratory power of our simulation. In order to exploit real-world data, however, we would have to incorporate, for example, individual heterogeneity (each individual has a different probability of birth and death), spatial variation (birth and death probabilities differed by areas), temporal variation (birth and death probabilities differed in time). So far, our simulation accounts only for demographic variation.

Finally, statistical error indicators different from (N)RMSE should be investigated, being population evolution time-series data.

VIII. Additional information

All the presented results can be fully reproduced on both simulation and output analysis side, since the code is available under GPL-2.0 license on the Github repository for the Simulation and Performance Evaluation final project[14], by respectively executing the *C# project* and *Matlab* files, belonging to the folders matching the paper paragraphs: *Mersenne-Twister Test* for section IV-B, *Parameter Tuning* for V-A and V-B, and *Single Simulation* for V-C and V-D. A guide on how to set the parameters for each folder can be found on the repository wiki, along with the corresponding cited *Supplementary Tables*.

[14]Github repository web URL: https://github.com/spam-thunder-with-sun/SimulationPerformanceEvaluationFinalProject.git

References

[1] G. Miller and S. Spoolman, *Essentials of Ecology*. Brooks/Cole, 2009, ISBN: 9780538734271. [Online]. Available: https : / / books . google . it / books ? id = elkxmgEACAAJ.

[2] D. Hughes-Hallett, A. Gleason, and W. McCallum, *Calculus: Single and Multivariable, 6th Edition*. Wiley, 2012, ISBN: 9781118475713. [Online]. Available: https://books.google.it/books?id=FT4cAAAAQBAJ.

[3] C. Y. Shen, "Logistic growth modelling of covid-19 proliferation in china and its international implications," *International Journal of Infectious Diseases*, vol. 96, pp. 582–589, 2020, ISSN: 1201-9712. DOI: https://doi.org/10.1016/j.ijid.2020.04.085. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1201971220303039.

[4] E. Pelinovsky, A. Kurkin, O. Kurkina, M. Kokoulina, and A. Epifanova, "Logistic equation and covid-19," *Chaos, Solitons Fractals*, vol. 140, p. 110 241, 2020, ISSN: 0960-0779. DOI: https://doi.org/10.1016/j.chaos.2020.110241. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077920306378.

[5] S. Zhao, I. Hu, J. Lou, *et al.*, "The mechanism shaping the logistic growth of mutation proportion in epidemics at population scale," *Infectious Disease Modelling*, vol. 8, no. 1, pp. 107–121, 2023, ISSN: 2468-0427. DOI: https://doi.org/10.1016/j.idm.2022.12.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468042722001129.

[6] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977. DOI: 10.1021/j100540a008. eprint: https://doi.org/10.1021/j100540a008. [Online]. Available: https://doi.org/10.1021/j100540a008.

[7] M. Pineda-Krch, "Gillespiessa: Implementing the gillespie stochastic simulation algorithm in r," *Journal of Statistical Software*, vol. 25, no. 12, pp. 1–18, 2008. DOI: 10.18637/jss.v025.i12. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v025i12.

[8] A. A. Berryman, "On principles, laws and theory in population ecology," *Oikos*, vol. 103, pp. 695–701, 2003. [Online]. Available: https://api.semanticscholar.org/CorpusID:18151192.

[9] T. R. Malthus, *An Essay on the Principle of Population* (History of Economic Thought Books malthus1798). McMaster University Archive for the History of Economic Thought, 1798.

[10] *Nouveaux mémoires de l'Académie royale des sciences et belles-lettres de Bruxelles*, bel. L'Académie Royale de Bruxelles et de l'Université Louvain, 1845. [Online]. Available: https://www.digizeitschriften.de/id/129323640_0018%7Clog1.

[11] P.-F. Verhulst, "A note on the law of population growth," in *Mathematical Demography: Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1977, pp. 333–339, ISBN: 978-3-642-81046-6. DOI: 10.1007/978-3-642-81046-6_37. [Online]. Available: https://doi.org/10.1007/978-3-642-81046-6_37.

[12] L. Lipkin and D. Smith, "Logistic growth model," *Convergence*, 2004. [Online]. Available: https://maa.org/press/periodicals/loci/joma/logistic-growth-model (visited on 10/31/2023).

[13] F. Cavallini, "Fitting a logistic curve to data," *The College Mathematics Journal*, vol. 24, no. 3, pp. 247–253, 1993, ISSN: 07468342, 19311346. [Online]. Available: http://www.jstor.org/stable/2686488 (visited on 10/31/2023).

[14] R. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 26, p. 457, Jul. 1976. DOI: 10.1038/261459a0.

[15] S. P. Gordon, "Comparing the discrete and continuous logistic models," *PRIMUS*, vol. 18, no. 5, pp. 449–455, 2008. DOI: 10.1080/10511970701884566. eprint: https://doi.org/10.1080/10511970701884566. [Online]. Available: https://doi.org/10.1080/10511970701884566.

[16] F. Takasu, *Lecture 10: Logistic growth models*, https://gi.ics.nara-wu.ac.jp/~takasu/lecture/old/archives/L10_logistic-growth.pdf, Accessed: 2024-06-13, 28 June 2006.

[17] M. Pineda-Krch and R. Cannoodt, *Gillespiessa: Gillespie's stochastic simulation algorithm (ssa)*, R package version 0.6.2, 2022. [Online]. Available: https://CRAN.R-project.org/package=GillespieSSA.

[18] G. Chure, *Tutorial 6: Stochastic simulation of constitutive expression*, https://www.rpgroup.caltech.edu/ncbs_pboc/code/t06_gillespie_algorithm.html, Accessed: 2024-07-02, 2018.

[19] D. Sànchez-Taltavull, *Stochastic modelling in mathematical biology*, https://www.ub.edu/simba/slides/simba130304-DSanchez.pdf, Accessed: 2024-07-02, 2013.

[20] A. Policriti, *Stochastic simulation (and Gillespie's algorithm)*, https://users.dimi.uniud.it/~alberto.policriti/home/sites/default/files/bioinformatica-supe/Stochastic-Simulation-advanced.pdf, Accessed: 2024-07-02.

[21] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013. [Online]. Available: https://faculty.marshall.usc.edu/gareth-james/ISL/.