Analysis of Time Series: from SIMCoV to the Logistic Map

James Morris
Computer Science
University of New Mexico
jmorris7@unm.edu

John (Jack) Ringer
Computer Science
University of New Mexico
jrin42@unm.edu

***Abstract*—Within this report the dynamics of time series from SIMCoV [2], the logistic map, and real-world COVID-19 cases and deaths are explored in-depth. Parameter sets found within SIMCoV that give rise to stable, periodic, and chaotic dynamics are provided and related to real-world phenomenon. The logistic map is used as a toy model to investigate periodic and chaotic dynamics. For each set of data we calculate various metrics taken from information theory and relate these metrics to observed behavior. Finally, we discuss the relationship between SIMCoV and the logistic map along with the limitations of using the logistic map to model more complex systems. We note that the unique code and data used for this report may be found at the following GitHub repository: https://github.com/Jack-42/cs523_project-1**

I. Introduction

In this project we explore various potential dynamics of a viral infection using the SIMCoV model. Small changes in initial conditions and parameters can have dramatic effects on how the viral infection evolves within the model. This makes it a good problem to analyze from the perspective of complex adaptive systems.

Here we define a few terms of relevance:

The *logistic map* (equation 1) is a recurrence relation which displays interesting properties by varying the inputs. Particularly, by varying values of r it may display stable, periodic, or chaotic dynamics. It thus is a useful tool in the analysis of complex systems, and in this case we compare some results from the logistic map with that of SIMCoV. In this context, *chaos* is behavior which is *sensitive to initial conditions*. An example of such behavior is a double-rod pendulum dropped from some initial height. The behavior of the pendulum (modelled in a computer) is deterministic but is so sensitive to initial conditions that even slight perturbations in the initial height or position will result in vast divergence in motion after some period of time.

SIMCoV is a model for the progression of the SARS-CoV-2 infection within the lungs [2]. This problem is suitable for study as taking place within a complex adaptive system, because of the vast number of factors which can influence how SARS-CoV-2 infection plays out within the body. For example, every person has a unique immune system, and the SARS-CoV-2 infection can vary depending on variant, initial viral load, and other factors. SIMCoV therefore attempts to model some of these unique differences between cases to better understand how they can affect the course of disease.

II. Methods

A. Modifying SIMCoV

Within this project we modify the default configuration of SIMCoV [2] in order to explore parameters that lead to stable, periodic, and chaotic dynamics for viral load and infected cells. To discover these parameters we began with the default SIMCoV configuration, finding that this configuration led to a stable final state with all virions being eradicated from the body by the final timestep. Thus, to try and find interesting dynamics, parameters which increased viral load over time (e.g., infectivity) were initially increased and those which decreased viral load (e.g., virion clearance) were decreased. Through a trial-and-error process the parameters shown in Table I were found.

TABLE I
Overview of Parameters used for SIMCoV

	Default	Stable	Periodic	Chaotic
Incubation Period	480	480	600	480
Expressing Period	900	900	1100	900
Apoptosis Period	180	180	120	180
Infectivity	0.001	0.01	0.002	0.01
Virion Production	1.1	0.11	1.1	0.11
Virion Clearance	0.004	0.002	0.002	0.004
Virion Diffusion	0.15	0.15	0.2	0.15
T Cell Production	105000	105000	130000	105000
T Cell Initial Delay	10080	4320	9000	4320
T Cell Tissue Period	1440	1440	1600	1440
T Cell Binding Period	10	10	7	10

Please note that for the sake of space only parameters which were modified from the default configuration are included. To view full configuration files see the project repo.

To verify these findings, we ran each configuration (stable, periodic, dynamic) 5 times, using a different random seed for each run. In particular, we use a random seed of x for run x (i.e., a random seed of 1 for run 1, a random seed of 2 for run 2, etc). See Results section for visualizations of these runs.

B. Calculations

To plot and save values from the logistic map we use Python coupled with various packages including Pandas [1], Matplotlib [5], and Numpy [6]. The scripts used to generate all figures may be found at the project repo.

Values for the logistic map are generated for 100 timesteps ($t = 100$) according to the equation:

$$x_{t+1} = r x_t (1 - x_t) \tag{1}$$

Where x_t is the population at timestep t and r is the reproduction rate. We explore periodic and chaotic dynamics by using different values of r with slightly different initial conditions.

Entropy, Mutual Information, and Transfer Entropy are all metrics used within this work to quantify the amount of information within a single series (Entropy) or the information between two series (Mutual Information, Transfer Entropy). All of these metrics have been covered in-class and so we do not describe them in this report.

The JIDT tool was used to compute each of these metrics [3]. Note that to compute these metrics the data must be discretized. Our logistic map data (where each value $x_t \in [0, 1]$) is discretized by assigning each of these values to 1 of 10 discrete bins, each bin of equal size. In particular we use the Pandas.cut() utility to bin data [1]. A similar procedure is used to discretize the data from SIMCoV, but we increase the bin size to 20 to reflect the larger set of values in SIMCoV data.

For the real-world data, we use the same binning technique with 15 bins. Data was retrieved from OurWorldInData.org [11] at their GitHub repository [12].

III. Results

A. SIMCoV Dynamics

As mentioned in the previous section, we run each of our found configurations 5 times using different initial states for each run. Fig. 1 shows the dynamics of both the epithelial cell state as well as T-cell and viral load states for the stable configuration over 5 runs. Note how changes to the random seed led to negligible changes in dynamics.

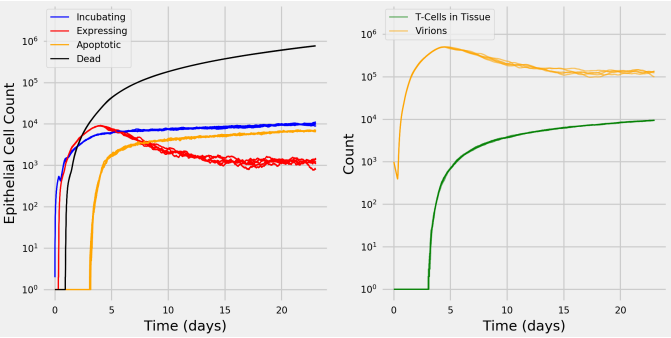

Fig. 1. Dynamics of epithelial cell state (left) and T-cell and viral load dynamics (right) for the stable configuration. The results of 5 separate runs are shown

Fig. 2 shows the time series of both infected cells (incubating + expressing) as well as viral load for each configuration over 5 separate runs. It is interesting to note that while changes to the random seed led to almost no changes in dynamics for the stable configuration, for both the periodic and chaotic configurations the random seed had a noticeable impact. These deviations in behavior are not seen until the

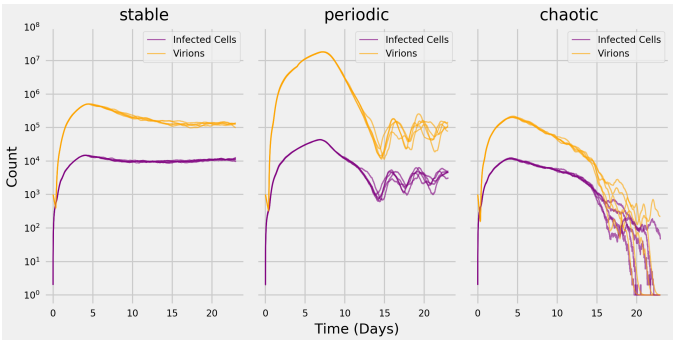

Fig. 2. Plot of infected cells (expressing + incubating) and number of virions for each configuration. The stable configuration is shown left, periodic middle, and chaotic on the right. For each configuration 5 separate runs are shown.

interesting dynamics (periodic and chaotic) start to take place around days 13 and 15. Prior the onset of periodic and chaotic dynamics, the deviation between each run is negligible for both configurations.

For all configurations we made some modification to the behavior of the virus (e.g., altering the infectivity rate). These modifications could reflect variants of the virus, as mutations can alter viral dynamics in alarming ways [10].

As can be seen from Table I the difference in parameters between the stable and chaotic regimes is rather small, with the only difference being the viral clearance rate (which is twice as high in the chaotic regime). Given that the viral clearance rate is an abstract value representing the general strength of the immune response [2], it's not surprising that decreasing this value can cause the virus to remain for longer and at higher levels. A decreased viral clearance could represent an individual with a relatively weaker immune system, such as an elderly person [8] or someone who is immunocompromised [9]. These populations are often at-risk for experiencing longer and/or more severe COVID-19 symptoms, which is reflected in the simulation by the increased viral load and number of infected cells shown in Fig 2.

The parameters which gave rise to periodic dynamics are further from the default configuration than those of the stable and chaotic regimes. Periodic dynamics are found in some autoimmune diseases, with large incubation periods being found to be partly-responsible for this behavior in computational models [7]. Hence why we increased the incubation period to try and discover periodic dynamics. Changes to the immune system behavior (from default) could be explained by differences in populations touched-on above.

B. Logistic Map Experiments

In addition to our investigation of SIMCoV dynamics, we also use the logistic map as a toy model to explore periodic and chaotic behavior. Fig. 3 shows the dynamics of the logistic map over 100 timesteps for different values of r and x_0. For the top plot, a value of $r = 3.2$ is used whereas a value of $r = 3.75$ is used in the bottom plot. For each value of r two series using slightly different initial conditions ($x_0 = 0.1$ in red and $x_0 = 0.100001$ in blue) are shown.

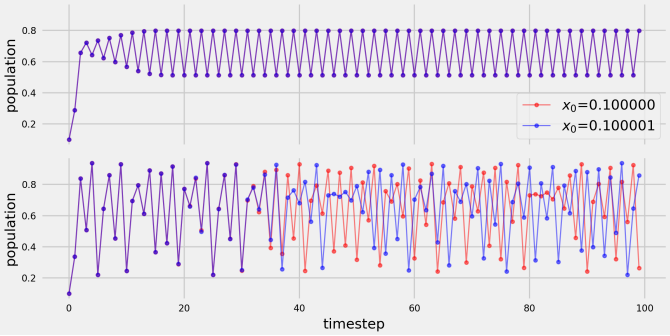

Fig. 3. Plot of logistic map for $r = 3.2$ (top) and $r = 3.75$ (bottom) over 100 timesteps for slightly different initial conditions (x_0)

As can be seen by Fig. 3, a logistic map with $r = 3.2$ converges to a periodic time series for both initial conditions. The series for both initial conditions overlap and converge to the same behavior. On the other hand, a logistic map with $r = 3.75$ gives chaotic dynamics. Although the series with different initial conditions overlap initially, around timestep $t = 37$ they diverge. After 100 timesteps neither series converges towards predictable behavior. These results are in-line with the fact that there is an onset of chaotic dynamics at $r \approx 3.57$ [4]. This means the logistic map will converge to either a stable point or exhibit periodic behavior for $r \leq 3.57$, whereas most values that fall within $3.57 \leq r \leq 4$ give chaotic dynamics (there are some islands of stability) [4].

C. Information Theory Metrics

In this section we present the calculated Entropy, Mutual Information, and Transfer Entropy for/between many of the series explored within this work. These metrics are shown in Table II. The following paragraphs discuss the calculated metrics for each of the major series (Logistic Map, SIMCoV, and Real-World COVID-19 Cases) explored within this work.

In Table II data for the logistic map is shown in the first four columns. As would be expected, the Shannon Entropy was higher for the chaotic series (right two columns) than the periodic series (left two columns). After $n = 37$ timesteps the periodic series has converged to bouncing between two values, which explains why the Shannon Entropy goes to 1 (only need 1 bit to store 2 values). For the chaotic series, the mutual information is initially high (1.8974) but after the two series diverge at about $n = 37$ timesteps the mutual information drops, as would be expected. Transfer entropy was 0 for the periodic data (both before and after timestep $n = 37$), which follows from the fact that the two periodic series overlap and so there is no information to be gained from looking at the other series. For the chaotic data transfer entropy is 0 when the two series are coupled, but when they diverge there is the potential for some information to be gained when looking at the other series. This is why the transfer entropy increases from the top to the bottom row for the chaotic logistic map data.

The results for the SIMCoV data are not as clean as the logistic map data, in particular for Shannon Entropy. One would expect that the Shannon Entropy for both viral load and infected cells would be highest for the chaotic series (followed by periodic), but this is not what was found. The reason for this discrepancy between expectation and calculation is likely due to the inconsistency of plotting data on a log-scale while using fixed-size (linear) binning methods. As is shown in Fig. 4, our configurations do indeed demonstrate the dynamics we propose they do when plotted on a log-scale. However, the smaller magnitude of the chaotic data caused much of it to be placed in similar bins. On the other hand, although the stable data appears flat on a log-scale, there are some large fluctuations which get captured by using linear binning. Hence the discrepancy between theory and calculation for Shannon Entropy. Unfortunately, the authors did not have enough time to investigate logarithmic binning methods.

The mutual information and transfer entropy findings for the SIMCoV data are reasonable. Given that viral load is highly correlated with the number of infected cells (see Figs 2 and 4), it is no surprise that the mutual information between these two series is high while the transfer entropy is close to 0 (the two series being highly correlated entails that little information is to be gained about one by looking at the other).

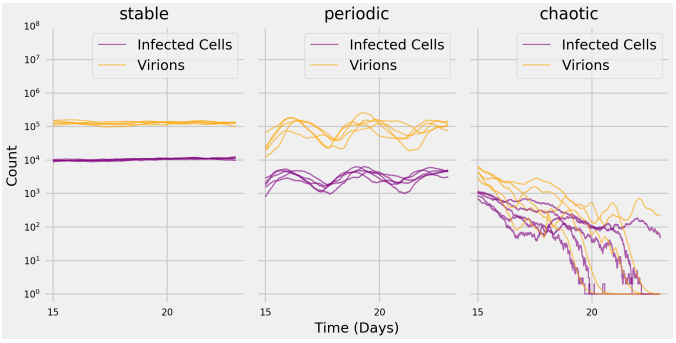

Fig. 4. Close-up of infected cells and number of virions for each configuration between days 15 and 23. The x-axis is restricted to days ≥ 15 to highlight the regions where interesting dynamics occur.

We now discuss the data from the real world setting. We chose to compare the number of new daily worldwide COVID-19 cases with the number of new deaths.

Figure 5 is a $log(y)$ graph of the daily new cases (in red) and the daily new deaths (in black). We see that at log scale the cases and deaths track quite closely with each other. For the first ~50 days of the pandemic, we observe somewhat chaotic dynamics, but this is probably attributable to poor data collection in many countries - i.e. different countries started collecting this data at different points in the first days of the pandemic. It then enters a somewhat stable growth regime from roughly day 50 until day ~350, after which point we start to see more periodic dynamics - which coincides with the emergence of variants as well as the first vaccines [14]. We start to see a large increase in cases around day 700 (the Omicron variant) with a somewhat smaller increase in deaths

TABLE II
OVERVIEW OF INFORMATION METRICS

	Logistic Map Data [1]				**SimCov Data** [2]			**Real-World Data** [3]	
Case	*r=3.2, x_0=0.1*	*r=3.2, x_0=0.100001*	*r=3.75, x_0=0.1*	*r=3.75, x_0=0.100001*	*Stable*	*Periodic*	*Chaotic*	*New Cases*	*Deaths*
Shannon Entropy [a]	2.1065	2.1065	2.8844	2.8613	3.7098	2.6868	3.3639		
	1	1	2.9532	2.9514	3.6367	3.5044	3.7886	2.4041	3.6007
Mutual Information [b]	1.3576		1.8974						
	1		0.8443		1.9408	2.0177	2.5688		0.6387
Transfer Entropy [c]	0		0		0.0064	0.0031	0.006		
	0		0.642		0.0021	0.0025	0.0024	0.3494	0.2608

Presented here is the Shannon Entropy, Mutual Information, and Transfer Entropy related to the series explored in this work. Given the large amount of information in the table, we provide the following guide for rows/columns that are not self-explanatory:

- Logistic Map Data (1)
  - General notes: The top row is for the first $n = 37$ timesteps (where chaotic data still overlap), the bottom row for the remaining 63 timesteps.
  - 1b: Mutual information between series with $x_0 = 0.1$ and the series with $x_0 = 0.10001$.
  - 1c: Transfer entropy from the series with $x_0 = 0.1$ to the series with $x_0 = 0.10001$.
- SIMCoV Data (2)
  - 2a: Viral load shown on top row, infected cells on bottom row.
  - 2b: Mutual information calculated between viral load and infected cells.
  - 2c: Transfer entropy going from viral load to infected cells shown on top, vice-versa on bottom.
- World-Wide COVID-19 Cases and Deaths (3)
  - 3c: Left column (New Cases) shows transfer entropy going from new cases to new deaths, right column is vice-versa.

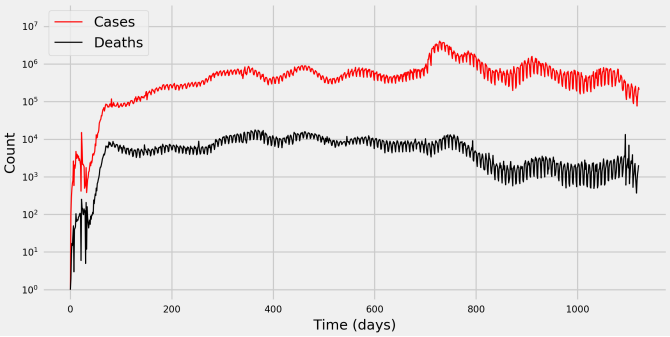

Fig. 5. COVID-19 worldwide daily new cases vs new deaths, from January 2020 until present [11]

at this point, and from this time until the present we see roughly periodic dynamics in both the case and death data.

Note that the Shannon Entropy obtained from the new cases (Table II) time series is lower than that obtained from the new deaths time series. This may be explained by the fact that the number of new cases from one day to the next is "predictable" in the sense that exposure to the virus is correlated with a somewhat predictable incubation period followed by testing positive for the virus some days later - and this phenomenon can be analyzed on the level of population dynamics. On the other hand, the death rate may be less predictable because the chance of dying from COVID-19 is influenced by a host of factors including vaccination status, the presence of comorbidities, access to health care, and many others. The death data also contains some strong outliers and hence has more variability from day to day - including one day in January 2023 on which China reported almost 60,000 deaths, far, far higher than any other country on that day, although this data point was deleted from the dataset to facilitate better results in the entropy analysis above (it appears that this number was for a period of about one month between early December and early January). Indeed, the variation at the tail end of the death curve coincides with China's rollback of their most stringent COVID-19 policies [15].

The mutual information quantity is not trivial to interpret on the population scale, but given the immense number of factors which influence both the case data and the death data the value obtained seems to suggest a reasonable reduction in uncertainty between the two time series. This makes sense, as the number of daily deaths would be reasonably expected to track with the number of daily cases, given some lagging period which we would expect to be the average time from diagnosis to death (roughly 18 days in an analysis of Australian data [13]).

The final point to make is the disparity between the two values of transfer entropy obtained. It seems somewhat obvious that the transfer entropy from cases to deaths would be higher than that of deaths to cases: some percentage of people who are infected with COVID-19 will die of the disease, while the reverse doesn't hold, hence we gain more information about the death rate by looking at case rate, rather than vice versa.

IV. Discussion

SIMCoV - or more generally, the progression of SARS-CoV-2 infection - can be thought of as analogous to the logistic map. In the logistic map equation, the growth variable is reduced to one value, r, whereas the growth of a variable in a SIMCoV run is dependent on several parameters. The "infectivity" parameter is one such example, which modulates the probability of a cell being infected at a given time step. Simultaneously, at each time step infected cells are being attacked by the immune system parameters which prevents the infected cells from completely overrunning the lungs. For example, the clearance parameter is an abstraction of the innate immune response and is the fraction by which the virion count drops per minute, so increasing this term would result in a lower r. Presumably if we set all the immune system parameters to 0 and maximized the growth parameters, while sufficiently increasing the time before an infected cell enters apoptosis and dies, we would simply see the infected cells arrive at a maximum and stay there until the end of the simulation. Infected cells are also restricted to infecting nearby cells. In this way, r can be thought of as some function of the growth parameters and the immune parameters, along with the number of cells nearby an infection site, at a given time step. The carrying capacity k for a variable is the number of available cells in the lung; e.g. the carrying capacity for the number of infected cells is the number of infectable (not dead, expressing, or apoptotic) cells.

As such, by varying the parameters in the SIMCoV model, we can produce results which display stable, periodic, and chaotic dynamics, in keeping with the behavior of the logistic map.

We find that Shannon Entropy, mutual information, and transfer entropy all can offer insight into the dynamics of and relationships between time series. High values for Shannon Entropy generally indicate chaotic behavior, whereas lower values correspond with more predictable behavior. This is in-line with a common interpretation of Shannon Entropy, which is that it is the amount of "surprise" in a message. The mutual information between two series is relatively high (i.e., close to the Shannon Entropy's of the two series being compared) when the series are strongly correlated, such as the two periodic series generated from the logistic map (see Table II). Transfer entropy is high when going from series X to series Y when knowing the past values of X provides additional information for predicting future values of Y (i.e., more information than we would have gotten from just looking at the past values of Y). This is why series which overlap have a transfer entropy close to 0, whereas those which have more complex and intermingled relationships (such as the COVID-19 new cases and deaths data) have higher transfer entropy values.

For SIMCoV data there was some inconsistency between how data was visualized (on a log-scale) versus how data was binned, which caused calculations for Shannon-Entropy to give counter-intuitive results. However, the calculated mutual information still provided insight into the degree of shared information between viral load and infected cells. A low transfer entropy from viral load to infected cells (and vice-versa) reflects the high degree of correlation between these two series.

The data generated with SIMCoV was usually not obviously periodic, chaotic, or stable. A shortcoming of thinking about SIMCoV results in terms of the logistic map is that the logistic map has predictable behavior for certain values of r, whereas in SIMCoV, our hypothetical function which generates r at each time step varies, which can lead to behavior that is sometimes periodic, sometimes chaotic, and sometimes stable, or behavior that doesn't match any of those descriptions completely.

References

[1] McKinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics.

[2] Moses ME, Hofmeyr S, Cannon JL, Andrews A, Gridley R, Hinga M, et al. (2021) Spatially distributed infection increases viral load in a computational model of SARS-CoV-2 lung infection. PLoS Comput Biol 17(12): e1009735. https://doi.org/10.1371/journal.pcbi.1009735

[3] Joseph T. Lizier, "JIDT: An information-theoretic toolkit for studying the dynamics of complex systems", Frontiers in Robotics and AI 1:11, 2014; doi:10.3389/frobt.2014.00011 (pre-print: arXiv:1408.3270)

[4] May, R. Simple mathematical models with very complicated dynamics. Nature 261, 459–467 (1976). https://doi.org/10.1038/261459a0

[5] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science and Engineering, vol. 9, no. 3, pp. 90-95, 2007.

[6] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2. (Publisher link)

[7] Canabarro, A., Gleria, I., and Lyra, M. L. (2004). Periodic solutions and chaos in a non-linear model for the delayed cellular immune response. Physica D: Nonlinear Phenomena, 342(1–2), 234–241. https://doi.org/10.1016/j.physa.2004.04.083

[8] Bajaj V, Gadi N, Spihlman AP, Wu SC, Choi CH, Moulton VR. Aging, Immunity, and COVID-19-19: How Age Influences the Host Immune Response to Coronavirus Infections? Front Physiol. 2021 Jan 12;11:571416. doi: 10.3389/fphys.2020.571416. PMID: 33510644; PMCID: PMC7835928.

[9] COVID-19-19 and Your Health. (2020, February 11). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-who-are-immunocompromised.html

[10] Yao, H., Lu, X., Chen, Q. et al. Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. Cell Discov 6, 76 (2020). https://doi.org/10.1038/s41421-020-00226-1

[11] Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

[12] https://github.com/owid/COVID-19-data

[13] Marschner, I.C. Estimating age-specific COVID-19 fatality risk and time to death by comparing population diagnosis and death patterns: Australian data. BMC Med Res Methodol 21, 126 (2021). https://doi.org/10.1186/s12874-021-01314-w

[14] Vaccine timeline. Historic Dates and Events Related to Vaccines and Immunization. (n.d.). Retrieved February 19, 2023, from https://www.immunize.org/timeline/

[15] Press, T. A. (2022, December 7). China announces a roll-back of its strict anti-covid-19 measures. NPR. Retrieved February 19, 2023, from https://www.npr.org/2022/12/07/1141172723/china-announces-a-roll-back-of-its-strict-anti-covid-19-measures

Contribution Statement

Both authors worked on finding SIMCoV configuration files that gave interesting dynamics. Jack worked on generating

data for the logistic map and calculating metrics for logistic map data and SIMCoV data. James gathered real-world data and generated relevant figures and metrics. Both authors contributed equally to the writing of this report.