Abstract

Severe acute respiratory syndrome (SARS), caused by a strain of coronavirus called SARS-CoV, is a viral respiratory infection that was first discovered in Guangdong Province of China in November 2002. It has subsequently led to an epidemic outbreak in Hong Kong in March 2013 before it became fully contained in July of the same year. Benchmarking with the daily reported cases published by the World Health Organization (WHO), our study simulates the outbreak using a basic compartmental susceptible-infected-removed-dead (SIRD) model and concludes that while our best-estimated model can simulate the total infected, recovered, dead cases, and the infectious peak relatively well, the model has a limited ability to simulate the spread and the containment of the virus on a more granular basis due to its simplistic model assumptions. Advanced models that address some of our model limitations are briefly mentioned at the end.

REPORT

1 Introduction

Severe acute respiratory syndrome, also known as SARS, is a viral respiratory infectious disease caused by severe acute respiratory coronavirus called SARS-CoV (CDC, 2017). The virus was first discovered in Guangdong Province of China in November 2002, giving rise to a world-wide epidemic in the following year before it was contained in July 2003 (CDC, 2017). According to the \textit{Overview on SARS in Asia and the World}, the emergence of SARS in Hong Kong SAR, China, was attributed to a medical doctor from Guangdong Province who travelled to Hong Kong in February 2003, unknowingly spreading the disease to his brother-in-law, Hong Kong health care workers, and a dozen of guests from the hotel where he stayed during his visit (WK et al., 2003). The infected guests further propagated the disease when they travelled back home, ultimately causing more than 8,000 infections with 774 death cases in 29 countries and regions globally (WK et al., 2003).

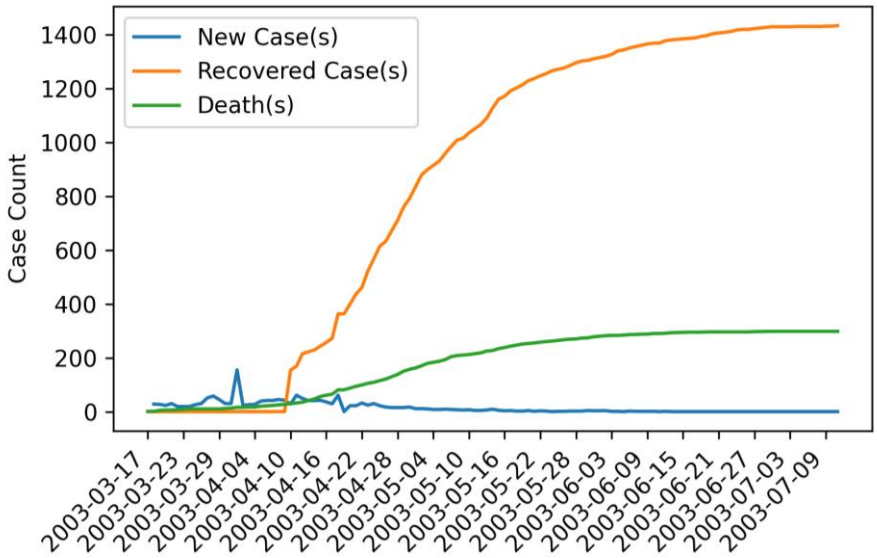
The SARS epidemic in Hong Kong can be described in three phases. The first phase took place in March 2003, where a large number of hospital staff and medical students were infected. The second phase started in late March in a residential estate called Amoy Gardens, reaching its peak in early April 2003, where the highest single-day new cases was reported. The third phase started in early May, where the virus was continuously spread among hospitals and residential estates (*reference 3)*, before accelerating into terminal decline by mid-May (*reference 4*).

During the outbreak, Hong Kong maintained a transparent attitude in its data collection and reporting, which was acknowledged by the World Health Organization (WHO) (*reference 5*). Data transparency is critical for our mathematical modelling purpose. In the rest of the report, we will introduce two of the most common basic mathematical models in infectious disease modelling - Susceptible-Infected-Removed (SIR) model and Susceptible-Infected-Removed-Dead (SIRD) model - and attempt to model the 2003 SARS outbreak in Hong Kong using the SIRD model. The purpose of our effort is to investigate whether a simplistic SIRD model is sufficient to capture the underlying mechanism of the spread and the containment of SARS in Hong Kong.

2 Methodology

2.1 Data Acquisition and Processing

Our analysis is based on a dataset compiled from the daily \textit{*Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS)*} report released on the WHO website *(reference 4)*, in which cumulative number of cases, number of deaths, and number recovered are made available from March 17, 2003 to July 11, 2003. Data was reported 6 times a week from Monday to Saturday in March, April and May, and 5 times a week from Monday to Friday in June and July. For the reporting gaps on weekends, linear interpolation was employed to generate data. Daily new cases were then derived by taking the difference between daily cumulative cases of the two adjacent days.



(Fig. 2.1.1, Hong Kong SARS Cases)

2.2 Susceptible-Infected-Removed (SIR) Model

Susceptible-Infected-Removed (SIR) model is a three compartmental model developed by Kermack and McKendrick, the work of which was first published in a 1927 paper *(reference 1)*. The model divides a fixed population of N individuals into three time-variant compartments:

- S(t), the number of susceptible individuals at time t;
- I(t), the number of infectious individuals at time t; and
- R(t), the number of recovered individuals at time t.

In its simplest form ,the model contains two parameters, β, the transmission rate per day, and γ, the recovery rate per day. At a constant transmission rate, susceptible individuals are unidirectionally transitioned into the infectious population over time. Similarly, at a constant recovery rate, transmission individuals are unidirectionally transitioned into the recovered population over time. The same transmission and recovery rate are assumed for

all individuals in the modelled population. The model is actualized by the following set of differential equations:

$$\frac{dS}{dt} = -\beta \cdot I(t) \cdot S(t)$$
$$\frac{dI}{dt} = \beta \cdot I(t) \cdot S(t) - \gamma \cdot I(t)$$
$$\frac{dR}{dt} = \gamma \cdot I(t)$$

As the sum of the three equations equates to zero, it reiterates the key model assumption that the population count is fixed at all time t - an individual is either susceptible to, infected by, or recovered from the infectious disease.

However, when modelling the SARS cases development in Hong Kong, fatality is a non-negligible driver that affects the rate at which individuals leave the infectious compartment, in tandem with the recovery rate. The case fatality ratio was estimated to be 15% in Hong Kong using the non-parametric competing risk analysis, as disclosed in the consensus document on the epidemiology of SARS by the WHO in November 2003 *(reference 2)* when the epidemic has fully concluded. Therefore, we are motivated to expand on the three compartmental SIR model to include death.

2.3 Susceptible-Infected-Removed-Dead (SIRD) Model

A four compartmental SIRD model is adopted as our choice of model, in which deaths resulting from the disease are counted towards the fourth compartment, Dead. The death rate is controlled by a new parameter, μ, and the model equations become:

$$\frac{dS}{dt} = -\beta \cdot I(t) \cdot S(t)$$
$$\frac{dI}{dt} = \beta \cdot I(t) \cdot S(t) - (\gamma + \mu) \cdot I(t)$$
$$\frac{dR}{dt} = \gamma \cdot I(t)$$
$$\frac{dD}{dt} = \mu \cdot I(t)$$

In the next section, we will use the SIRD model to simulate the 2003 SARS outbreak in Hong Kong. All computations are completed using Python (*reference 3*). In addition to the standard libraries, we also use Pandas (*reference 4*) to work with dataframes, Numpy *(reference 4*) to generate arrays, Scipy (*reference* 4) to integrate ordinary differential equations of the SIRD models, and Matplotlib *(reference 4*) to create visualizations.

3 Simulation Study

In this section, we will perform simulations of the SIRD model using various parameter settings to calibrate for a final model that best describes the collected data. We choose to tune the parameters with a trial-and-error approach to generate coarse parameter estimations after conducting a preliminary analysis about the collected data, in view for the observation that the reported cumulative recovered cases only starts to increase from zero 20 days in the study period when the reported daily infected cases has already moved past its peak, as shown in Fig. 2.1.1. The SIRD model, on the other hand, continuously transitions infected individuals into the recovered and dead compartments from the very beginning of the study period. Hence, while we align the simulated infection curve with the reported daily new cases trend, a gap between the simulated and the reported cumulative recovered cases is expected, especially early in the study period. Another observation we make is that the reported daily new cases display multiple sizable local extrema compared to the global extremum, making model fitting using the sum of squared errors (SSE) less desirable in this case, as it can cause misalignment of the infection peak. With those reasons, the trial-and-error approach is deemed appropriate for parameter estimations.

The next step is to determine the total population in the model and the initial number of susceptible and infected individuals. The population of Hong Kong in 2003 was roughly 6.8 million (reference 1), which we attempt to use as our model total population. As the scale of infection is extremely small in comparison to the total population, the model has an infinitesimal transmission rate and a high recovery rate. We conclude that this setting is undesirable due to the reason that an infectious peak on April 1, 2003, with roughly 155 cases and an equilibrium for the simulated recoveries at roughly 1433 cases at the end of the study period cannot be reached simultaneously. To account for the fact that the peak of the daily infected case count being 155 and the ultimate recovered case count being 1433, we continue to simulate using a variety of total population N that are much smaller than the total population of Hong Kong, to observe whether they produce reasonably scaled simulated curves per the collected data. Among our attempts, we observe that a total population N between 2500 to 4500 produces the most reasonable simulated cases compared. We are led to propose an assumption that only a small subset of the Hong Kong population could be susceptible to SARS, while the rest of the population is shielded from the disease. A similar model assumption has been proposed by Ng, Turinici, and Danchin in \textit{A double epidemic model for the SARS propagation}, where a smaller modelled population size is adopted for their Hong Kong model to better fit the data (reference 2). They included possible reasons, such as vaccination and inherent immunity, but left the specific reason “yet to define” (reference 2). For our simulations and the rest of the report, we proceed with an assumption that a subset of 3300 individuals in Hong Kong can be susceptible to the disease, of which 1 individual is infected and 3299 individuals are susceptible initially.

There are three parameters in the SIRD model that we need to adjust: the transmission rate β, the recovery rate γ, and the death rate μ. As we aim to find a set of suitable parameters by estimation, we look at the following three metrics for each simulation: cumulative cases, cumulative deaths, and cumulative recoveries, all on the final day of the study period.

	transmission rate β	recovery rate γ	death rate μ	cumulative cases	cumulative recoveries	cumulative deaths
data	unknown	unknown	unknown	1755	1433	298

(Table 3.1.1)

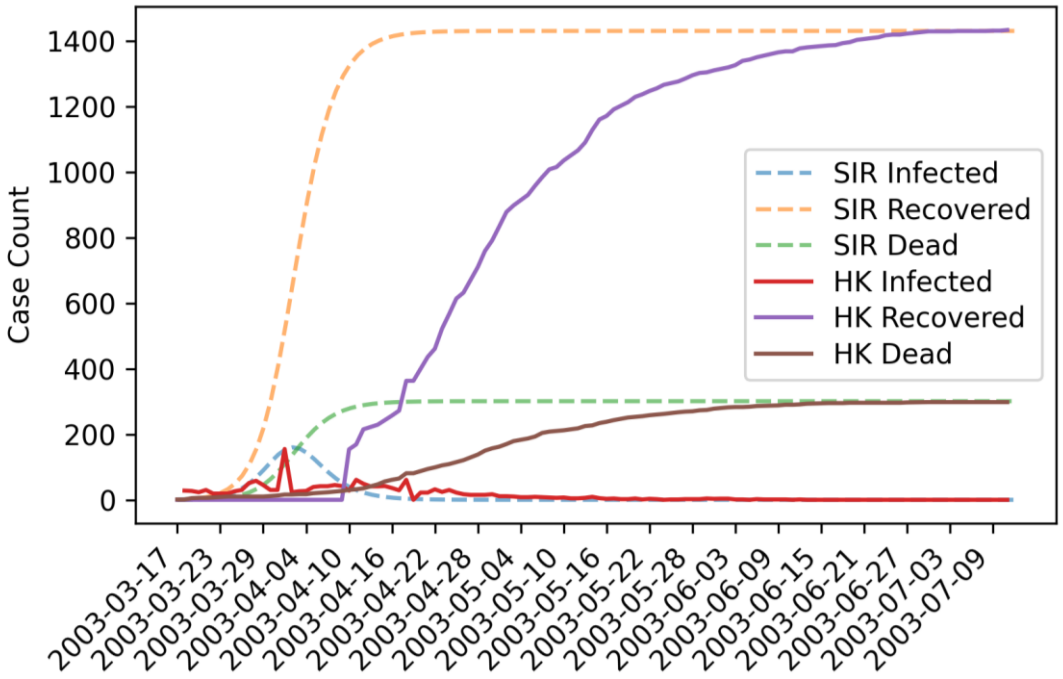
We initiate our model with a transmission rate of 4x10^ (-10), a recovery rate of 0.81, and a death rate of 0.17, which is the ultimate death rate on the last day of the study. For easy comparisons to the collected data, we compute the differences between the simulated counts and the actual counts for cumulative cases, cumulative deaths, and cumulative recoveries, as shown in the following table:

Simulation	transmission rate β	recovery rate γ	death rate μ	diff. cumulative cases	diff. cumulative recoveries	diff. cumulative deaths
1	0.00040	0.810	0.17	-194.85	158.77	30.60
2	0.00045	0.810	0.17	225.01	-184.26	-41.42
3	0.000425	0.810	0.17	32.69	-27.13	-8.44
4	0.0004125	0.810	0.17	-76.2	61.83	10.22
5	0.00042	0.810	0.17	-9.76	7.55	-1.17
6	0.00042	0.808	0.17	1.22	2.12	-3.05
7	0.00042	0.80823	0.17	-0.05	2.75	-2.83

(Table 3.1.2)

In the first four simulations, the recovery rate remains fixed at 0.810, while we adjust the transmission rate using a midpoint method in the hope to decrease the gap between the simulated and the collected case counts. The final three simulations are conducted with a trial-and-error principle, with the aim to close the gap between the cumulative cases while allowing an error margin of 3 to the differences of the cumulative recoveries and deaths.

Simulation 7 seems to produce a reasonable model. We plot the simulated I(t), R(t), and D(t) curves against the collected data:



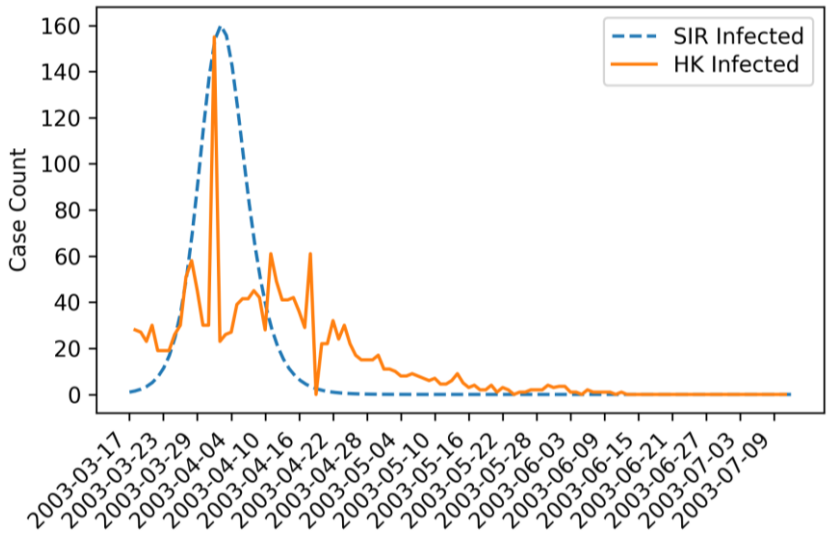
(Figure 3.1.1, simulation 7 v.s. collected data)

As expected, to ensure the proper modelling of the infection curve, as a trade-off, the recovery and death curve do not tailor the collected data well in the early study period, but the limiting behaviours seem to be captured. The data reports the highest single-day new cases (155 cases) on April 1, 2003; in comparison, our simulation has the highest single-day new cases (160 cases) on April 2, 2003, which is reasonable.

4 Application & Analysis

In this section, we continue to compare and contrast our best-estimated SIRD model against the collected data to examine whether the model captures the underlying mechanism of the spread and the containment of SARS in Hong Kong, and identify a few model limitations that contribute to the gaps between the modelled and the collected data.

Our best-estimated model (simulation 7) has a transmission rate (β) of 4.2 x 10^(-4), a recovery rate (γ) of 0.80823, and a death rate (μ) of 0.17. The model produces a relatively close approximation to the limiting behaviours of the collected recovered and dead case counts, but deviates from the collected data noticeably in the early study period. This is because the model transitions the infected individuals continuously into the recovered and dead compartments, while in reality, recoveries and deaths from SARS can occur after several weeks of illness (reference 1). The infection curve, from a first glance, looks acceptable - it produces an infectious peak around the same time as the collected data. However, when we separate the infection curve and data from the rest of the plot, we can see that the simulated infection curve decreases monotonically while the actual infected cases continue to fluctuate, and overall, takes longer to diminish.



(Figure 4.1.1)

Recall from the introduction that the epidemic in Hong Kong can be described in three phases: outbreak in hospitals, outbreak in Amoy Gardens, and the continuing spread among the hospitals and the housing estates. Our simulated curve is roughly aligned to the peak of the second phase when the outbreak took place in the Amoy Gardens; the highest single-day new cases was reported in this phase. In phase three, sizable daily new cases continue to be reported, which our model fails to include.

The basic reproduction number R0, an epidemiological metric that measures the average number of secondary infections an infected individual can cause, can be computed with

$R_0 = \frac{\beta}{\gamma + \mu}$ (reference 3) for an SIRD model. It is evaluated to be less than 0.01 for our model, suggesting the decline and eradication of the disease over time. This, by no means, solely suggests low infectivity in comparison with recovery and death. It is not to overlook the roles of mask wearing or selective quarantine (reference 4), the effect of which our model does not consider but plays an integral part in disease control.

5 Conclusion

In this report, we introduced two basic mathematical models, SIR and SIER model, commonly used in infectious disease modelling. We simulated the 2003 SARS outbreak in Hong Kong by estimating the transmission rate, recovery rate, and death rate based on the data reported by the WHO, in an attempt to answer whether a simplistic SIRD model can sufficiently capture the underlying mechanism of the spread and the containment of SARS in Hong Kong.

We conclude that although the model is able to produce similar total infected, recovered, dead cases, and a similar infectious peak in the beginning of April 2003, the model has a limited ability to simulate a few other aspects accurately due to its assumptions. Those aspects include but not limited to: time differences between the time of infection and the time of recovery or death; the incubation period and the time from onset to hospital arrival; outbreaks taking places in different phases in different regions (i.e. hospitals, residential estates); the effect of mask wearing and quarantine; differences between typical close-contact transmission and faecal-droplet transmission due to the inadequate plumbing system (*reference 1*); and identifications of the super spreaders (*reference 2*). To address these limitations, possible model directions include superimposing two compartmental models to simulate the spread of the disease in different regions (*ref 2 from simulation study*), or using a Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) model (*reference 3*) to account for the period from incubation to hospital arrival.

References

CDC. (2019). SARS. Retrieved from Centers for Disease Control and Prevention website: https://www.cdc.gov/sars/about/fs-sars.html

Census and Statistics Department. (2003). Hong Kong 2003 - Hong Kong: The Facts. Retrieved from www.yearbook.gov.hk website: https://www.yearbook.gov.hk/2003/english/hkfact/hkfact.html

Fonseca i Casas, P., García i Carrasco, V., & Garcia i Subirana, J. (2020). SEIRD COVID-19 Formal Characterization and Model Comparison Validation. *Applied Sciences*, *10*(15), 5162. https://doi.org/10.3390/app10155162

Guido Van Rossum. (1995). *Python reference manual*. Amsterdam, The Netherlands: Centrum Voor Wiskunde En Informatica.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hethcote, H. W. (1989). Three Basic Epidemiological Models. *Applied Mathematical Ecology*, 119–144. https://doi.org/10.1007/978-3-642-61317-3_5

Hung, L. S. (2003). The SARS epidemic in Hong Kong: what lessons have we learned? *Journal of the Royal Society of Medicine*, *96*(8), 374–378. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC539564/

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *115*(772), 700–721. https://doi.org/10.1098/rspa.1927.0118

Kong, H. (2003). *SARS in Hong Kong : from experience to action : a summary report of the SARS Expert Committee.* Hong Kong: Government Logistics Dept.

LAM, W., ZHONG, N., & TAN, W. (2003). Overview on SARS in Asia and the World. *Respirology*, *8*(s1), S2–S5. https://doi.org/10.1046/j.1440-1843.2003.00516.x

Mckinney, W. (2010). Data Structures for Statistical Computing in Python. In *PROC. OF THE 9th PYTHON IN SCIENCE CONF* (p. 51). Retrieved from https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

Ng, T. W., Turinici, G., & Danchin, A. (2003). A double epidemic model for the SARS propagation. *BMC Infectious Diseases*, *3*(1). https://doi.org/10.1186/1471-2334-3-19

Novel Coronavirus (2019-nCoV) situation reports. (2020). Retrieved from www.who.int website: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

Siu, A., & Wong, Y. C. R. (2004). Economic Impact of SARS: The Case of Hong Kong. *Asian Economic Papers*, *3*(1), 62–83. https://doi.org/10.1162/1535351041747996

Small, M., Tse, C. K., & Walker, D. M. (2006). Super-spreaders and the rate of transmission of the SARS virus. *Physica D. Nonlinear Phenomena*, *215*(2), 146–158. https://doi.org/10.1016/j.physd.2006.01.021

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

World Health Organization. (2003a). *Consensus document on the epidemiology of severe acute respiratory syndrome (SARS) DEPARTMENT OF COMMUNICABLE DISEASE SURVEILLANCE AND RESPONSE*. Retrieved from https://www.who.int/csr/sars/en/WHOconsensus.pdf

World Health Organization. (2003b, May 7). WHO | Update 49 - SARS case fatality ratio, incubation period. Retrieved from WHO website: https://www.who.int/csr/sarsarchive/2003_05_07a/en/

World Health Organization. (2003c, September 26). Inadequate plumbing systems likely contributed to SARS transmission. Retrieved from www.who.int website: https://www.who.int/news/item/26-09-2003-inadequate-plumbing-systems-likely-contributed-to-sars-transmission