

Markov Chain Models for COVID-19

Max Liu

Wyoming Seminary College Preparatory School, Pennsylvania, USA

Mentor: Dr. Fanhui Kong, Department of Mathematics and Computer Science, Wilkes University

INTRODUCTION

In late 2019, the first case of a deadly new virus was identified in Wuhan, China: the SARS-like coronavirus, also known as COVID-19. Despite immediate response, both in and outside of China, the coronavirus soon spread across China and the rest of the globe. After a strict quarantine period, the number of new cases slowly began to decrease in China, only for the number to rise in other parts of

the world. In the United States alone, thousands of new people were being infected daily, forcing almost all businesses and institutions to shut down or move to a virtual setting. Despite efforts to restrict contact and limit the spread of the coronavirus, the number of daily new cases remains consistently high in the United States. In my home state of Pennsylvania, hundreds of new cases are still being re-

ported daily, and many institutions and businesses remain closed or under strict regulations. This study attempts to use a model formed by a Markov chain to estimate the daily number of new coronavirus cases.

METHODS

A Markov chain is a sequence of the state of a group over a period of time, where the probability from one state to the next is determined by the current state only. We can think of the spread of coronavirus the same way: The amount of people infected and recovered in the next time period is dependent on the current state only.

Many works, such as a similar Markov Chain model for other infectious diseases^[1], categorize each individual in the population into one of three states: susceptible, infected, and removed. We define the distribution $S^{(n)}$, where

$$S^{(n)} = \begin{pmatrix} S_0^{(n)} & S_1^{(n)} & S_2^{(n)} \end{pmatrix}, \quad (1)$$

so that at time n , $S_0^{(n)}$, $S_1^{(n)}$, $S_2^{(n)}$ represent the number of susceptible, infected, and removed (including dead and recovered), respectively. We then define a transition matrix,

$$P_{ij} = \begin{pmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{pmatrix}, \quad (2)$$

where

$$\sum_{j=0}^2 P_{ij} = 1, i = 0, 1, 2, \quad (3)$$

and P_{ij} represents the probability for an individual to transition from state i to state j .

Our Markov chain consists of a set of S distributions, where

$$S^{(n+1)} = S^{(n)} \cdot P \quad (4)$$

To define the matrix P , we find the values for each row individually, with the knowledge that each row sums up to 1.

We assume P_{02} is equal to zero, as it is rare for an individual to enter a removed state without going through the infected state first within a one day period. Also, P_{10} is assumed to be zero, as it is rare for an individual to return to a susceptible state and possibly being reinfected after being infected once. Similarly, we assume P_{20} and P_{21} equal zero. P_{01} can be calculated,

$$P_{01} = \frac{(S_1^{(n+1)} - S_1^{(n)}) + (S_2^{(n+1)} - S_2^{(n)})}{S_0^{(n)}} \quad (5)$$

Similarly, we solve for P_{11} by taking the number of individuals that remained infected, also known as the people who were infected at time n and stayed infected at time $n+1$, otherwise known as the total number infected at time $n+1$ minus the number of new infected at time $n+1$, and dividing by the total number of infected at time n , so that

$$P_{11} = \frac{S_1^{(n+1)} - ((S_1^{(n+1)} - S_1^{(n)}) + (S_2^{(n+1)} - S_2^{(n)}))}{S_1^{(n)}} \quad (6)$$

which can then be simplified to

$$P_{11} = \frac{S_1^{(n)} - S_2^{(n+1)} + S_2^{(n)}}{S_1^{(n)}}. \quad (7)$$

By equation (3), $P_{22} = 1$, $P_{00} + P_{01} = 1$, and $P_{11} + P_{12} = 1$. With this, we have our completed P matrix, which we use in our Markov chain model.

To create our model, we employ two different types of Markov chains.

First, we use a stationary Markov chain. In this model, P remains constant throughout the entire simulation. To obtain P , we calculate the probability matrix for each day throughout our entire period, and average them to create a probability matrix. This idea is applied to a data analysis of China. A key assumption made for this simulation is that the people in China are properly following quarantine regulations, so that the number of possible susceptible people is significantly reduced.

Second, we create a non-stationary Markov chain model. Inspired by a study using an Erlang distribution^[6] and analysis of clinical observations, P_{01} and P_{12} are formulated using a Poisson distribution. With this model, we simulated the spread of coronavirus in China and Pennsylvania, both of which were able to capture the trend from observed data to a degree.

RESULTS

Figure 1

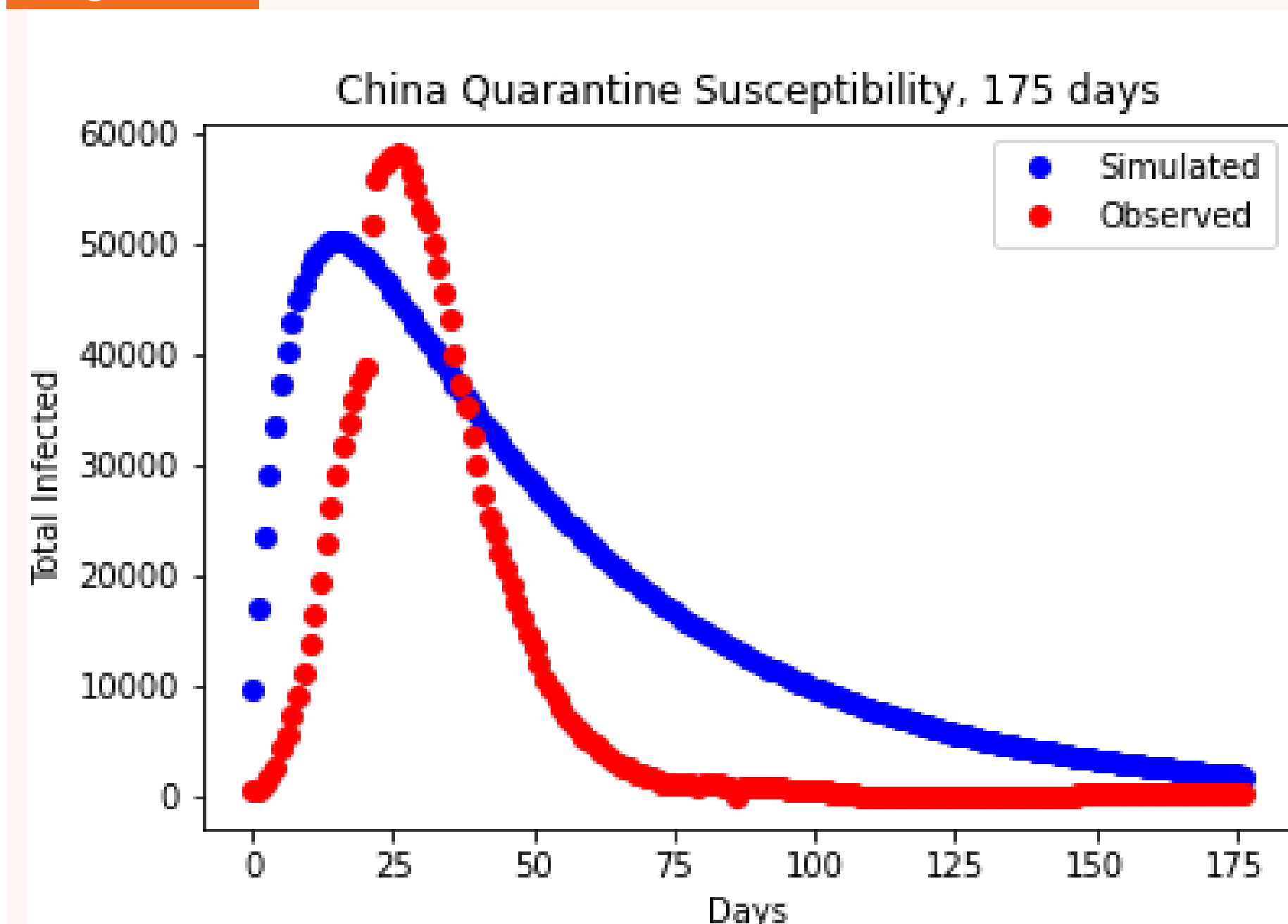


Figure 1 shows a comparison of the daily new infected from observed data and from simulated data, using our stationary Markov chain model with data from China, over a period of 175 days.

Figure 2

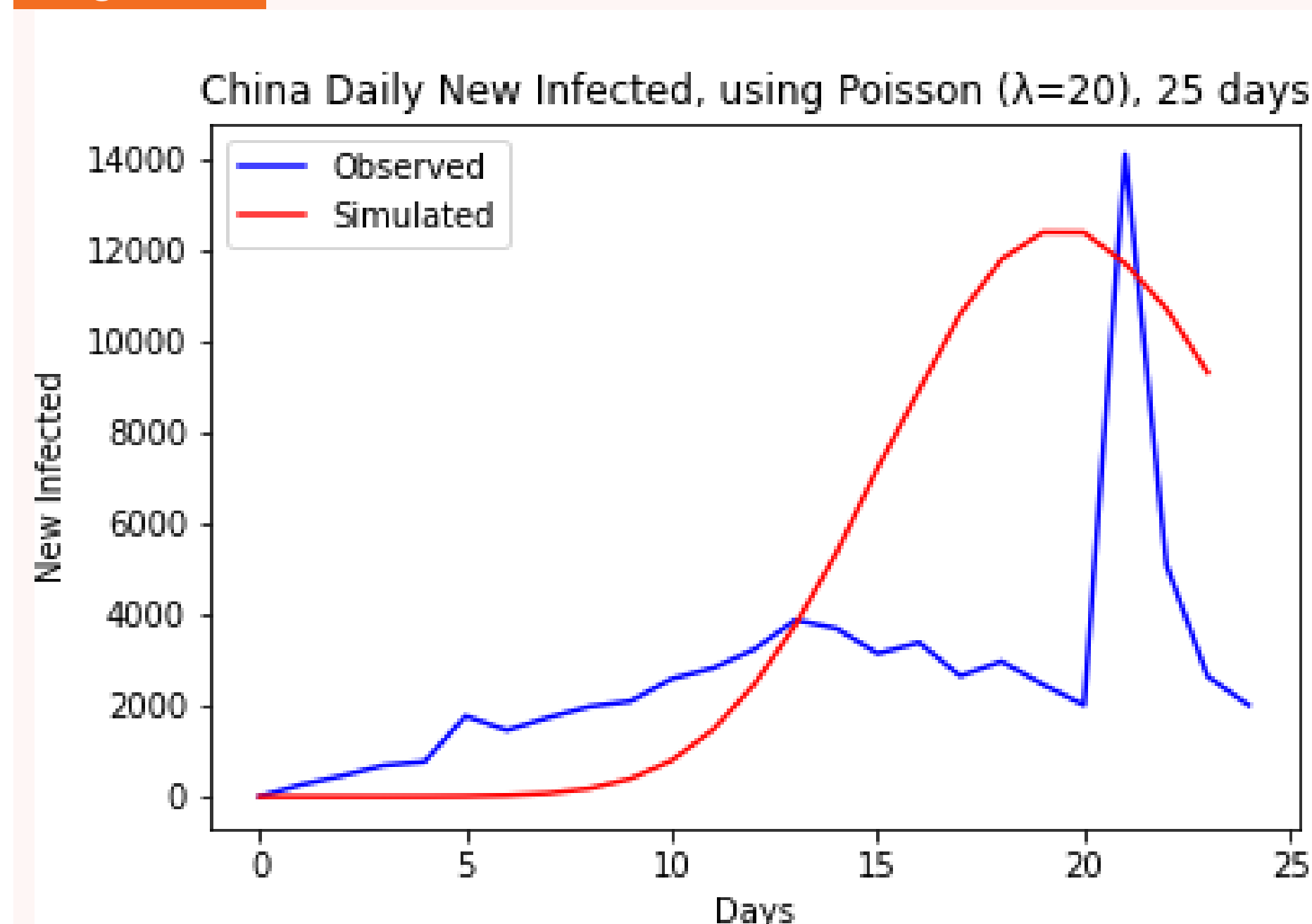


Figure 2 shows a comparison of the daily new infected from observed data and from simulated data, using a non-stationary Markov chain model with data from China, over a period of 25 days. This model uses a Poisson distribution with $\lambda = 20$ to model the value of P_{01} .

Figure 3

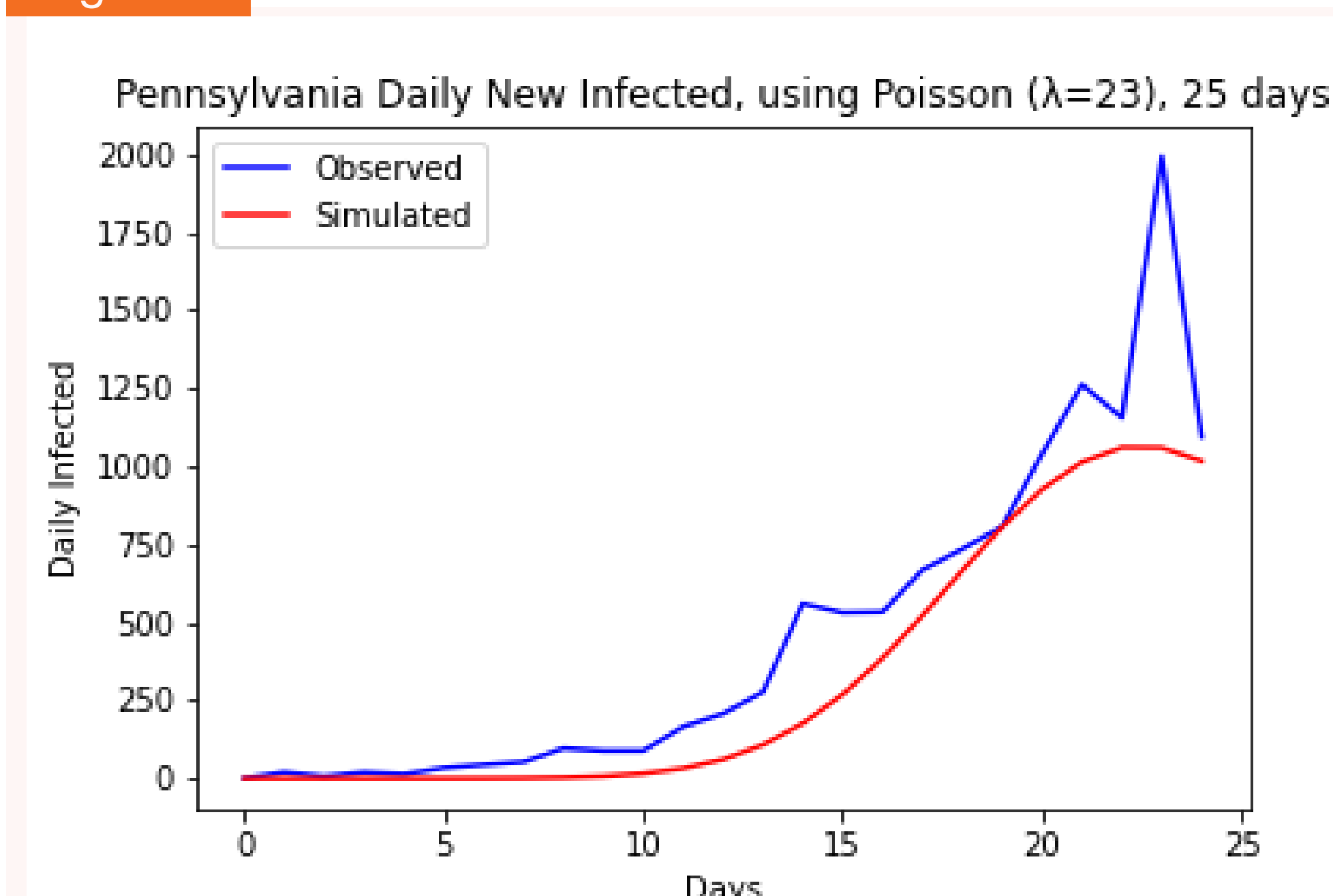


Figure 3 shows a comparison of the daily new infected from observed data and from simulated data, using a non-stationary Markov chain model with data from Pennsylvania, over a period of 25 days. This model uses a Poisson distribution with $\lambda = 23$ to model the value of P_{01} .

Figure 4

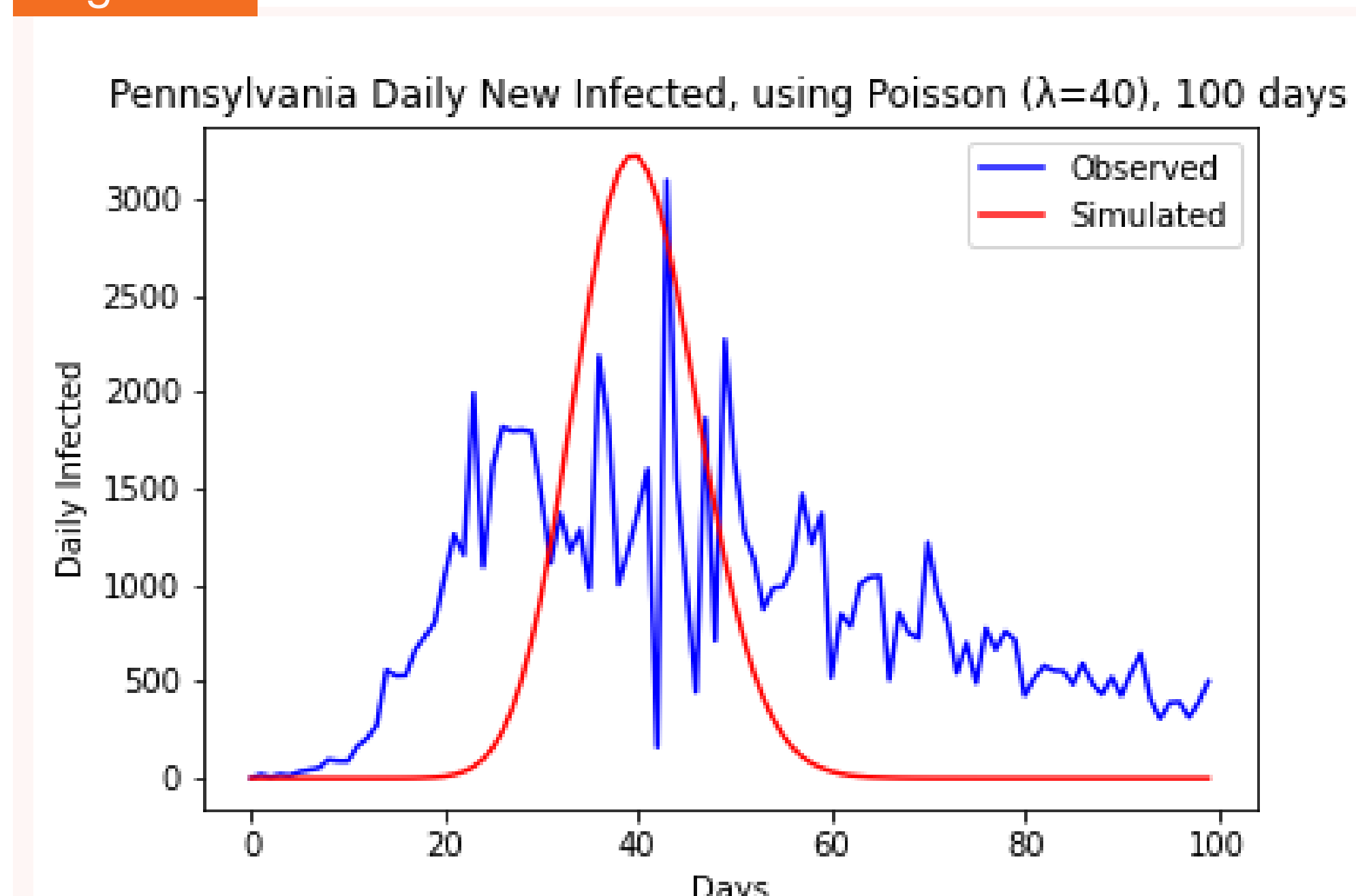


Figure 4 shows a comparison of the daily new infected from observed data and from simulated data, using a non-stationary Markov chain model with data from Pennsylvania, over a period of 100 days. This model uses a Poisson distribution with $\lambda = 100$ to model the value of P_{01} .

DISCUSSION

Our results suggest that it is possible to model the overall spread of the coronavirus using Markov chains to some degree. The trend shown by the spread of the virus seems to generally match the trend simulated. Obviously, the models are not perfect. One of our future goals is to use ODEs and other methods to improve the results. As a preliminary model, we were unable to take into account other factors, such as the popular behavior during the pandemic, including the impact of hospitalization and asymptomatic

individuals, as well as changes within the virus itself. We were unable to take into account the resurgence in the spread of the virus caused by such behaviors. Even now, many areas, including Pennsylvania, are experiencing a renewed growth in the number of cases, especially influenced by the reopening of schools, which is not considered in any of our models. We plan to keep investigating and improving our model to fit these behaviors.

References

- [1] Twumasi 2019 <https://doi.org/10.1155/2019/9362492>
- [2] Catak 2020 <https://doi.org/10.1101/2020.04.21.20073668>
- [3] Bendtsen Cano 2020 <https://doi.org/10.1101/2020.03.29.20046870>
- [4] Vivanco-Lira 2020 <https://arxiv.org/ftp/arxiv/papers/2003/2003.06758.pdf>
- [5] Zimmer 2020 the-scientist.com/features/why-r0-is-problematic-for-predicting-covid-19-spread-67690
- [6] Abrams 2020 <https://doi.org/10.1101/2020.06.29.20142851>