

Pandemic Flu Simulation

Nathan Kurtz-Enko

Table of contents

Abstract	1
Background	2
Literature Review	2
Modeling	4
Method	4
Read & Describe Case Data	5
Discussion	5
Future Work	5
Conclusions	5
Old Work	6
References	6

Abstract

- Problem
- Previous approaches
- Novelty
- Key findings

Background

Understanding the behavior of the spread of infectious disease, like flu pandemics, is critical for advising policy making with respect to public health and wellness. For example, studying the behavior of virus transmission and recovery rates allows health professionals and policy-makers to define the quantity of antiviral agents that should be readily available, regions and demographics that should be prioritized for vaccination, and other strategies intended to prevent a pandemic. Furthermore, being able to communicate the results of an infectious disease outbreak model/simulation is very important as this will facilitate adoption of containment measures.

Many researchers studying the spread of infectious disease explicitly cite the 1918 influenza outbreak (Berkessel et al. 2021; Germann et al. 2006; Johnson and Mueller 2002) as a motivating factor for their work. In the case of Johnson and Mueller (2002) and Patterson and Pyle (1991), the objective was to look back through previous records and data to verify morbidity of the 1918 pandemic. The authors made the observation that each time the statistics are reviewed, the total deaths from the pandemic seem to increase. In the 1920s, the record was approximately 21.5 million deaths; in 1991, Patterson and Pyle (1991) updated the estimate to 24.7-39.3 million; most recently, Johnson and Mueller (2002) estimated the death toll is on the order of 50 million. However, Johnson and Mueller (2002) does not make this claim definitively. Instead, he allows for speculation on different records in future estimates because there are many places affected by the 1918 influenza for which we do not have any data.

Clearly, the capacity for large-scale damage from infectious disease outbreaks is a very real threat. Hence, understanding how these things spread and the impact that different mitigation strategies have on the rate of spread and severity of illness is critical.

Literature Review

Infectious disease epidemic models have been the subject of study since the 1980s. It was Russian scientists that first began to implement models for estimating spread patterns of influenza and other viruses (Longini, Fine, and Thacker 1986). Much of that work is relatively inaccessible due to the fact that it is written in Russian. Nonetheless, eventually their work made its way into academia across the globe.

There are deterministic and stochastic approaches to modeling pandemic flu. However, early models were deterministic, and they described the movement of viruses using differential equations and partitioned populations into a number of subgroups. These are referred to as SIR (susceptible-infected-recovered), SEIR (susceptible-exposed-infected-recovered), and other derivatives.

Initial reactions to these deterministic models were skeptical for a few reasons. First, geographic regions being studied were relatively isolated, hence the ability to apply these models for global study was uncertain (Longini, Fine, and Thacker 1986). Furthermore, for the time,

these approaches were computationally expensive. Longini, Fine, and Thacker (1986) noted that over 90% of time was dedicated to running the algorithms themselves. However, this is less of an issue now because high-performance computing is readily available.

While relatively old, these models, and similar derivatives, still have applications today. Valle, Mniszewski, and Hyman (2012) used a deterministic model to show that an individual’s behavior will likely change in response to an outbreak; some individuals will adopt new behavior to protect themselves, thereby influencing the probability of transmission per contact.

More recently, stochastic models have become prevalent because they are able to capture more complex spreading patterns. Colizza et al. (2007), Cooper et al. (2006), Ferguson et al. (2005), Germann et al. (2006), Samsuzzoha, Singh, and Lucy (2013), and Tan et al. (2021) all used some stochastic process to model the spread of influenza or other infectious disease, and analyze different mitigation strategies. Colizza et al. (2007) found that a large-scale application of prophylaxis antiviral agents could successfully contain a pandemic H5N1 influenza given a viral reproduction ratio of 1.9 or lower. Cooper et al. (2006) found that significant air travel restrictions could delay the spread of pandemic influenza, but not contain an epidemic altogether. The findings of Germann et al. (2006) determined that rapid and preferential vaccination of children could contain an outbreak when the reproduction ratio of influenza is at or below 1.9.

Others have explored the relationship between geographic variables (e.g., altitude and solar radiation) and incidence rates for COVID-19, as well as severity of illness (Arias-Reyes et al. 2021; Stephens, Chernyavskiy, and Bruns 2021). Arias-Reyes et al. (2021) concluded that above 1000 meters above sea level, incidence rates and severity for COVID-19 decline. Through analysis of public geographic and COVID-19 data, Stephens, Chernyavskiy, and Bruns (2021) determined that for increases of 495 meters in elevation above LA county, and for areas of equivalent population density, infection rates were 12.82%, 12.01%, and 11.72%. However, mortality rates were not statistically significantly different between high and low elevation areas. Stephens, Chernyavskiy, and Bruns (2021) notes that other environmental variables such as increased solar radiation, larger swings of high/low temperatures throughout the day, and long-term exposure to altitude hypoxia could explain these observed differences.

Since influenza and similar viruses spread via aerosols, Sagripanti and Lytle (2007) proposed solar radiation as a contributing factor to the seasonality of influenza. Specifically, the authors calculated the expected inactivation of influenza by UV radiation in several cities and different times of the year. They estimated that a full day of sunlight exposure will reduce influenza by 99% in regions at a similar latitude to Mexico City, and by 90% in regions at a similar latitude to Miami. However, Weber and Stilianakis (2008) note that the impact on survivability of influenza by humidity and temperature are as significant as solar radiation. Furthermore, it is unlikely that outdoor transmission is a significant mechanism for the spread of influenza because it is so sensitive to wide range of environmental variables.

Regardless, it is clear that higher altitude, higher humidity/temperature, and higher solar radiation could theoretically influence the spread of pandemic influenza. The purpose of

this paper will be to simulate pandemic flu spread and identify critical thresholds for these covariates in order to minimize the risk of a severe outbreak.

Modeling

As previously mentioned, there are two very general approaches to simulating pandemic virus spread. We will be discussing the SIR model (Débarre, n.d.; “An Introduction to Deterministic Infectious Disease Models,” n.d.) because it is relatively simple, and establishes a foundation for considering other approaches.

The SIR model categorizes a population into three groups: susceptible, infected, and recovered. The rate of change of each of these groups is defined using differential equations, and we can account for mortality and birth rates in their definitions as well.

$$\begin{aligned}\frac{dS}{dt} &= \delta - \frac{\beta \times S \times I}{N} - \mu \times S \\ \frac{dI}{dt} &= \frac{\beta \times S \times I}{N} - (\gamma + \mu) \times I \\ \frac{dR}{dt} &= \gamma \times I - \mu \times R\end{aligned}$$

Here, S , I , R stand for the susceptible, infected, and recovered groups respectively. The population is denoted by N , and the rates of transmission, recovery, mortality, and birth are denoted by β , γ , μ , and δ .

The outbreak status is described by the reproduction ratio, which is the ratio between the transmission rate and recovery rate. When $\beta/\gamma > 1$, more people are becoming infected than are recovering, hence the system is in a state of outbreak. And, supposing that there is some percentage of the population that is already vaccinated (i.e., already in the recovered group), this ratio becomes $(\beta \times S(0))/(\gamma \times N)$, where $S(0)$ is the number of people in the susceptible group at time $t = 0$.

These facts make it easy to define different transmission, recovery, and vaccination target rates to avoid a significant outbreak. Hence, the SIR is a simple model, but very effective at clearly communicating actionable results to policy makers and authority figures.

Method

Using Epiverse-TRACE, RECONhub, and Epiforecasts suites of tools, and the R programming language (Wickham et al. 2025; Wickham and Henry 2025, 2025; Wickham, Vaughan, and Girlich 2025; Wickham, Hester, and Bryan 2024; Wickham 2016; Lambert, Kucharski, and Tamayo 2025; Gupte, Eggo, and Van Leeuwen 2025; Gupte, Van Leeuwen, and Kucharski 2025; Rudis 2022; Funk, Willem, and Gruson 2025).

Read & Describe Case Data

USA contact data (Breen, Mahmud, and Feehan 2022).

Build contact matrix.

Define initial conditions.

Estimate population data/demographics.

Define intervention measures.

Estimate parameters.

Simulate epidemic

Analyze results.

Discussion

Future Work

- Underexplored covariates or regions.
- Need for interdisciplinary models combining epidemiology, climatology, and public health.
- Potential of machine learning to enhance traditional SIR frameworks.

Conclusions

- Summary of key findings.
- Implications for pandemic preparedness and policy.
- Call for more nuanced, data-rich modeling approaches.

Old Work

The metapopulation stochastic model relies on both partitioning the population into sub-groups, and also randomness to model the spread of infectious diseases. (apolloni2014metapopulation?) describes this approach in great detail.

Effectively, there are two layers to this mode. First, there is a social layer, which follows the SIR paradigm to a degree. Second, there is a spatial layer, which models the distribution of individuals in space as well as their mobility.

For the social layer, we partition the population into two groups, $N_1 = N \cdot \alpha$ and $N_2 = N \cdot (1 - \alpha)$, where N is the population and α is the proportion of individuals in the first group.

These groups interact, and we use a two-by-two contact matrix to define these interaction rates.

$$C = \begin{pmatrix} \frac{p_1 q_1}{\alpha} & \frac{(1-p_1)q_2}{\alpha} \\ \frac{(1-p_1)q_1}{\alpha} & \frac{p_2 q_2}{1-\alpha} \end{pmatrix}$$

Where q_* is the contact rate of the group, and p_* is the contact rate within a group.

We use another two-by-two matrix to describe the secondary infections. Effectively, this next-generation matrix is a function of the contact matrix because each new case is produced by exposures from preceding contacts.

$$R = \frac{\beta}{\mu} \begin{pmatrix} C_{11} \cdot \alpha & C_{12} \cdot \alpha \\ C_{21} \cdot (1 - \alpha) & C_{22} \cdot (1 - \alpha) \end{pmatrix}$$

- Strengths and limitations in modeling flu pandemics.
- Case studies using SIR models for H1N1, SARS, or seasonal influenza.
- Use of stochastic vs. deterministic approaches.

References

- “An Introduction to Deterministic Infectious Disease Models.” n.d. <https://documents1.worldbank.org/curated/en/888341625223820901/pdf/An-Introduction-to-Deterministic-Infectious-Disease-Models.pdf>.
- Arias-Reyes, C., F. Carvajal-Rodriguez, L. Poma-Machicao, F. Aliaga-Raduán, D. A. Marques, N. Zubieta-DeUrioste, R. A. Accinelli, et al. 2021. “Decreased Incidence, Virus Transmission Capacity, and Severity of COVID-19 at Altitude on the American Continent.” *PLOS ONE* 16 (3): e0237294. <https://doi.org/10.1371/journal.pone.0237294>.

- Berkessel, J. B., T. Ebert, J. E. Gebauer, T. Jonsson, and S. Oishi. 2021. “Pandemics Initially Spread Among People of Higher (Not Lower) Social Status: Evidence from COVID-19 and the Spanish Flu.” *Social Psychological and Personality Science*. <https://doi.org/10.1177/19485506211039990>.
- Breen, C. F., A. S. Mahmud, and D. M. Feehan. 2022. “Novel Estimates Reveal Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States.” *PLOS Computational Biology* 18 (12): e1010742. <https://doi.org/10.1371/journal.pcbi.1010742>.
- Colizza, V., A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani. 2007. “Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions.” *PLoS Medicine* 4 (1): e13. <https://doi.org/10.1371/journal.pmed.0040013>.
- Cooper, B. S., R. J. Pitman, W. J. Edmunds, and N. J. Gay. 2006. “Delaying the International Spread of Pandemic Influenza.” *PLoS Medicine* 3 (6): e212. <https://doi.org/10.1371/journal.pmed.0030212>.
- Débarre, F. n.d. “SIR Models of Epidemics Level 1 Module in ‘Modelling Course in Population and Evolutionary Biology’” <https://ethz.ch/content/dam/ethz/special-interest/usys/ibz/theoreticalbiology/education/learningmaterials/701-1424-00L/sir.pdf>.
- Ferguson, N. M., D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. 2005. “Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia.” *Nature* 437 (7056): 209–14. <https://doi.org/10.1038/nature04017>.
- Funk, Sebastian, Lander Willem, and Hugo Gruson. 2025. *Socialmixr: Social Mixing Matrices for Infectious Disease Modelling*. <https://github.com/epiforecasts/socialmixr>.
- Germann, T. C., K. Kadau, I. M. Longini, and C. A. Macken. 2006. “Mitigation Strategies for Pandemic Influenza in the United States.” *Proceedings of the National Academy of Sciences* 103 (15): 5935–40. <https://doi.org/10.1073/pnas.0601266103>.
- Gupte, Pratik, Rosalind Eggo, and Edwin Van Leeuwen. 2025. *Epidemics: Composable Epidemic Scenario Modelling*. <https://github.com/epiverse-trace/epidemics>.
- Gupte, Pratik, Edwin Van Leeuwen, and Adam Kucharski. 2025. *Finalsize: Calculate the Final Size of an Epidemic*. <https://github.com/epiverse-trace/finalsize>.
- Johnson, N. P. A. S., and J. Mueller. 2002. “Updating the Accounts: Global Mortality of the 1918-1920 ‘Spanish’ Influenza Pandemic.” *Bulletin of the History of Medicine* 76 (1): 105–15. <https://doi.org/10.1353/bhm.2002.0022>.
- Lambert, Joshua W., Adam Kucharski, and Carmen Tamayo. 2025. *Epiparameter: Classes and Helper Functions for Working with Epidemiological Parameters*. <https://doi.org/10.5281/zenodo.11110881>.
- Longini, I. M., P. E. M. Fine, and S. B. Thacker. 1986. “Predicting the Global Spread of New Infectious Agents.” *American Journal of Epidemiology* 123 (3): 383–91. <https://doi.org/10.1093/oxfordjournals.aje.a114253>.
- Patterson, K. D., and G. F. Pyle. 1991. “The Geography and Mortality of the 1918 Influenza Pandemic.” *Bulletin of the History of Medicine* 65 (1): 4–21. <https://www.jstor.org/stable/44447656>.
- Rudis, Bob. 2022. “Cdcfluview: Retrieve Flu Season Data from the United States CDC FluView Portal.” <https://github.com/hrbrmstr/cdcfluview>; GitHub.

- Sagripanti, J.-L., and C. D. Lytle. 2007. “Inactivation of Influenza Virus by Solar Radiation.” *Photochemistry and Photobiology* 83: 1278–82. <https://doi.org/10.1111/j.1751-1097.2007.00177.x>.
- Samsuzzoha, Md., M. Singh, and D. Lucy. 2013. “Parameter Estimation of Influenza Epidemic Model.” *Applied Mathematics and Computation* 220: 616–29. <https://doi.org/10.1016/j.amc.2013.07.040>.
- Stephens, K. E., P. Chernyavskiy, and D. R. Bruns. 2021. “Impact of Altitude on COVID-19 Infection and Death in the United States: A Modeling and Observational Study.” *PLOS ONE* 16 (1): e0245055. <https://doi.org/10.1371/journal.pone.0245055>.
- Tan, Y., D. Cator III, M. Ndeffo-Mbah, and U. Braga-Neto. 2021. “A Stochastic Metapopulation State-Space Approach to Modeling and Estimating COVID-19 Spread.” *Mathematical Biosciences and Engineering* 18 (6): 7685–7710. <https://doi.org/10.3934/mbe.2021381>.
- Valle, S. Y. D., S. M. Mniszewski, and J. M. Hyman. 2012. “Modeling the Impact of Behavior Changes on the Spread of Pandemic Influenza.” In *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*, 59–77. https://doi.org/10.1007/978-1-4614-5474-8_4.
- Weber, T. P., and N. I. Stilianakis. 2008. “A Note on the Inactivation of Influenza a Viruses by Solar Radiation, Relative Humidity and Temperature.” *Photochemistry and Photobiology* 84: 1601–2. <https://doi.org/10.1111/j.1751-1097.2008.00416.x>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2025. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Wickham, Hadley, and Lionel Henry. 2025. *Purrr: Functional Programming Tools*. <https://purrr.tidyverse.org/>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://readr.tidyverse.org>.
- Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2025. *Tidyr: Tidy Messy Data*. <https://tidyr.tidyverse.org>.