

Plan for modeling in the Bay Area

Lucy M. Li

April 3, 2020

Executive summary

Ever since the SARS epidemic in 2003, mathematical modeling has been used to infer epidemiological quantities, uncover the natural history of disease, and evaluate effectiveness of interventions. In more recent years, modeling has been using in real time to help policymakers to make decisions on interventions and resource allocation by forecasting the number of morbidities and mortalities as well as resource requirements.

During the current pandemic, models used for past outbreaks have been adapted for capturing COVID-19 transmission. This document provides:

- 1) an overview of the process of modeling for infectious diseases;
- 2) broad categories of models that are generally used for infectious diseases and their application to COVID-19;
- 3) sources of bias and uncertainty, and approaches to address these issues;
- 4) computational setup and requirements for modeling;

A proposed strategy for modeling the transmission of COVID-19 in the San Francisco Bay Area is laid out in the last section.

Process of modeling for infectious diseases

1. Decide on question(s) to address

- What are key epidemiological parameter values?
- How many hospitalizations and deaths will occur in the next week?
- Who are more likely to infect others?
- Where are hotspots of transmission?
- What is the impact of social distancing?
- What is the fatality rate per infection?

2. Pick model(s) based on what we find out about the disease and pathogen.

- Routes of transmission
- Natural history of infection (CDC resource)
- Risk factors for infection and transmission?
- Animal and environmental reservoirs
- Immunity after infection (is re-infection possible?)
- Strains or serotypes of the pathogen
- Availability of treatments or vaccines
- Seasonality

3. Constraining or estimating parameters and initial model state values.

- Demography data to estimate population sizes
- Contact tracing data from the current outbreak can provide direct measurements of model parameters

- Serological surveys can provide information about who has already been infected
- Statistical distributions of generation times and incubation periods from similar infectious diseases could be used as priors to constrain parameter search space
- The ability to estimate parameters from data decreases with the number of parameters relative to the number of data points (paper on the identifiability of parameters in infectious disease epi models)
- Sometimes it's better to fix more parameters, and perform sensitivity analyses to assess dependence of results on assumptions

4. Model fitting

- ODE solvers for deterministic models
- General purpose optimization algorithms for simple models
- Approximate Bayesian computation
- MCMC is the main workhorse for more complex models and for Bayesian analyses. There are many flavors that are used: Metropolis-Hastings, Gibbs, Hamiltonian, particle marginal Metropolis Hastings)

5. Model selection / ensemble modeling

- Reversible jump MCMC
- AIC/BIC/DIC/Bayes factor
- Bayesian model averaging
- Ensemble modelig

6. Summarizing results

- Forecast of epidemic trajectory with credible intervals
- Parameter estimates with credible or confidence intervals
- Sensitivity analysis results
- Probabilities of events or outcomes
- Spatial visualizations

Categories of models

This paper in CID from 2015 provides a fairly comprehensive categorization of models, the epidemiological questions they address, and the types of data they need. Although the paper was written to provide guidance on modeling for the next flu pandemic, the principles are the same for modeling coronavirus.

Instead of regurgitating what was written in that paper, I will highlight a few papers that capture a range of options for modeling COVID-19.

SEIR models with spatial spread

Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study (Feb 29):

- Fixed incubation and infectious periods
- Fixed zoonotic infections
- Used Tencent mobility data to fix the number of people moving between provinces in China, and between China and neighboring countries
- Estimated R_0 from Wuhan data using a deterministic SEIR model
- Gibbs sampling

- Predicted infections in China and surrounding countries using forward simulation of a deterministic metapopulation SEIR model
- Showed how quickly the epidemic would die out assuming different reduction in transmission

Early dynamics of transmission and control of COVID-19: a mathematical modelling study

- Fixed the number of travellers to different countries
- Fixed the incubation period and delay in from onset to isolation
- Estimated time-varying R using stochastic SEIR model
- Estimated the relative rate of detection among exported cases compared to in Wuhan
- Used Sequential Monte Carlo to fit model to data

The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak

-
- Used the GLEAM model to estimate R_0
- Approximate Bayesian Computation
- Sensitivity analysis assuming different combinations of average latent and infectious periods, detection rates, initial conditions, and generation times

Branching processes

Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts

Individual-based models

<https://spiral.imperial.ac.uk/bitstream/10044/1/77482/5/Imperial%20College%20COVID19%20NPI%20modelling%2016-03-2020.pdf>

Statistical models

Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months

Biases and uncertainty

Common types of sampling issues

- Fraction of infections that are asymptomatic/mild symptomatic infections \rightarrow time-invariant
- Time-varying in type of tests being used
- Time-varying in testing criteria
- Uneven distribution geographically and according to human demographics
- False negatives \rightarrow people who were mis-diagnosed
- Potentially conflicting information from different data sources

Sensitivity to assumptions

Models tend to be sensitive to these assumptions

- Generation time distribution

- Mixing patterns if there is significant heterogeneity in transmission across different demographic groups in the population
- Rate of ascertainment

Computational

R/Python

Usually data cleaning, data transformation, and visualizations are done in R and Python, with the model and MCMC algorithm written in C or C++ for performance.

Most frequently used R packages: <https://www.repidemicsconsortium.org/projects/>

RStan

PyStan

C/C++

Stan

BUGS

JAGS

Containers

Not used that much for infectious disease outbreak modeling, even though Docker containers are widely used for pathogen genomics research. This would be an area of great technological contribution.

Plan for predictive modeling of COVID-19 in the Bay Area

Data

County case counts Hospitalizations Number of tests Mobility data → Google/Unacast Number of healthcare facilities, beds, and ventilators → Get data from SFDPH or individual hospitals?

ESRI's data

Sampling

Computational setup

Questions to address

- Predictions about COVID and hospitalizations
- Predictions about how long shelter-in-place needs to be in place
- Estimate real-time reproductive number before and after shelter-in-place
- Cross-jurisdictional transmissions based on genomic data

- Estimate ratio of introduced vs local transmission over time
- Side project on impact for healthcare by reducing load not just by COVID:
 - Impact on influenza
 - Impact on RSV infections
 - Food-borne outbreaks