Mathematical Modeling

Parameter Estimation and Forecasting Using ODE Models

Clement Ampong

October 2024

Introduction

The objective of this study is to estimate model parameters and forecast the spread of the 1918 influenza pandemic in San Francisco using a system of ordinary differential equations (ODEs). Specifically, this work implements the QuantDiffForecast MATLAB toolbox, utilizing both Nonlinear Least Squares (NLS) and Maximum Likelihood Estimation (MLE) methods to parameterize a Susceptible-Exposed-Infectious-Recovered (SEIR) model. The study aims to generate forecasts with quantified uncertainty, assessing the performance and accuracy of both estimation techniques.

Model Specification

To analyze the 1918 influenza pandemic in San Francisco, we utilize the SEIR model, which categorizes the population into four groups: susceptible (S), exposed (E), infectious (I), and recovered (R). This model tracks how individuals move through these stages over time, offering a structured method to study the progression and spread of the epidemic.

The compartmental diagram of the SEIR model, clearly showing the rates of flow between compartments is given below.

$$\boxed{\mathbf{S}} \xrightarrow{\beta \frac{I}{N}} \boxed{\mathbf{E}} \xrightarrow{k} \boxed{\mathbf{I}} \xrightarrow{\gamma} \boxed{\mathbf{R}}$$

From the model above, the variables are defined as:

- $S(t)$: number of susceptible individuals at time t
- $E(t)$: number of exposed individuals at time t
- $I(t)$: number of infected individuals at time t

- $R(t)$: number of recovered individuals at time t

The SEIR Differential equations for the model is given below:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} \\
\frac{dE}{dt} &= \frac{\beta SI}{N} - kE \\
\frac{dI}{dt} &= kE - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

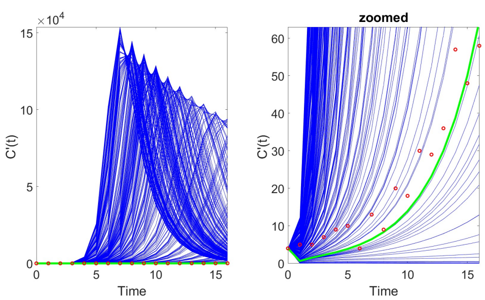
where:

β is the transmission rate,
κ is the latent period (κ = 1/1.9),
γ is the infectious period (γ = 1/4.1).

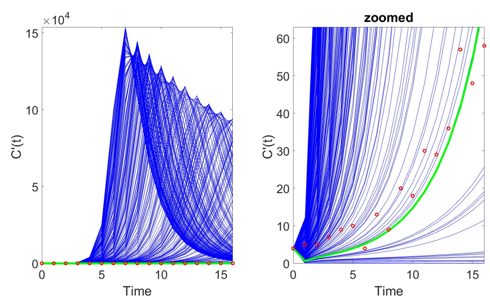
Parameter Estimation

We used two methods, Nonlinear Least Squares (NLS) and Maximum Likelihood Estimation (MLE), to estimate the transmission rate β, while keeping the latent period fixed at $\kappa = 1/1.9$ and the infectious period at $\gamma = 1/4.1$. Both techniques were implemented in MATLAB. We also performed parametric bootstrapping with 300 bootstrap realizations to quantify the uncertainty of the estimated parameters.

The diagrams below shows the parameter estimation for the two methods.

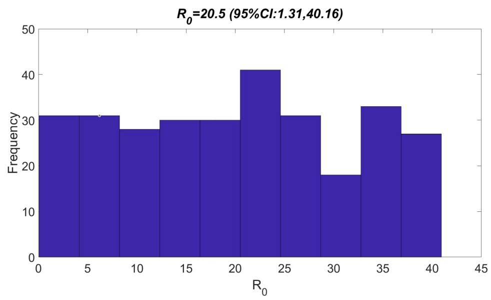


Model fit by NLS

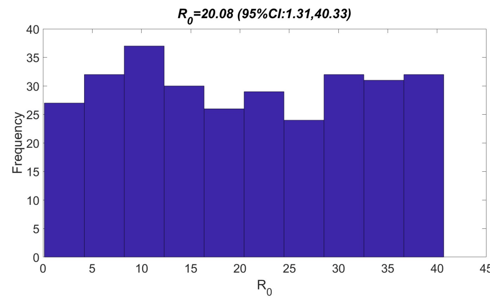


Model fit by MLE

The green line in both methods represents the best-fit solution, while the blue lines indicate the variability in predictions from bootstrapped estimates. In the NLS plots, the blue lines show a wider spread indicating greater uncertainty in the long-term predictions. The green line in the zoomed NLS plot fits the observed data (red circles) reasonably well but slightly underestimates the early growth. Also, the MLE plots show a tighter grouping of blue lines suggesting less uncertainty and a more consistent fit. The green line in the MLE zoomed plot follows the data points more closely than in NLS, capturing the early outbreak dynamics more accurately. Overall, MLE provides a more precise fit with less spread in predictions compared to NLS.

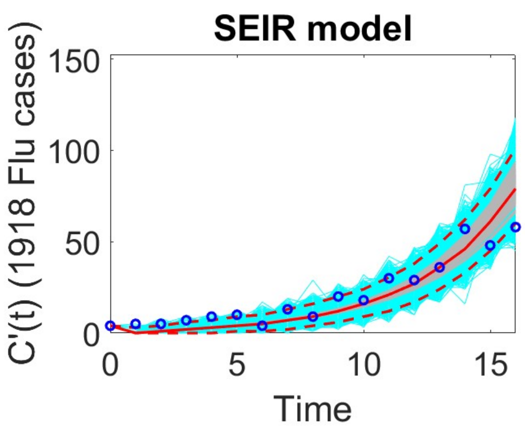


NLS histogram

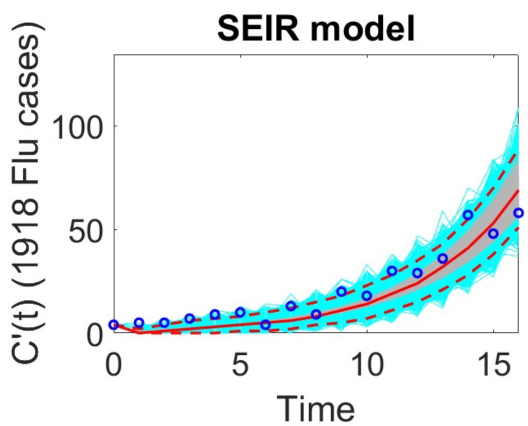


MLE histogram

The histograms for NLS and MLE shows that the NLS histogram has a wider spread of R_0 estimates, indicating greater variability and uncertainty. Also, the MLE histogram displays a more concentrated distribution, reflecting less variability and greater consistency. Overall, the NLS histogram suggests higher uncertainty in estimating R_0, while the MLE histogram indicates a more reliable estimation.

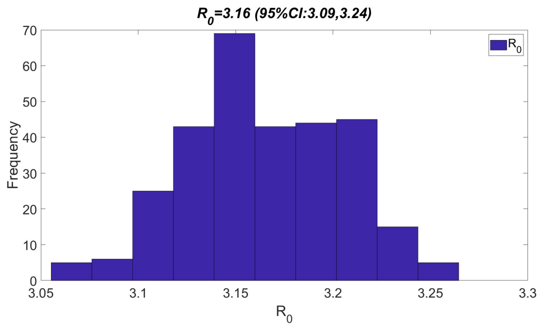


Model fit by NLS using Poisson error structure

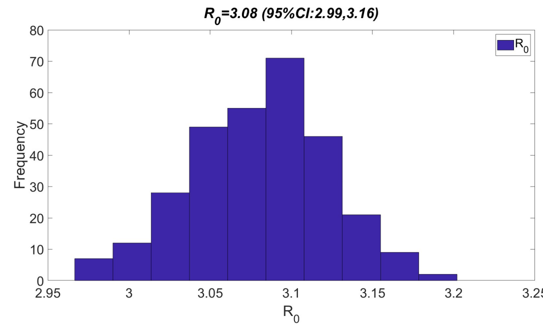


Model fit by MLE using Poisson error structure

These graphs represent the estimated transmission rate β using NLS and MLE methods. The lines show how well each method fits the observed data of the 1918 influenza pandemic in San Francisco. Both methods follow the data closely, with minor differences in fit quality.

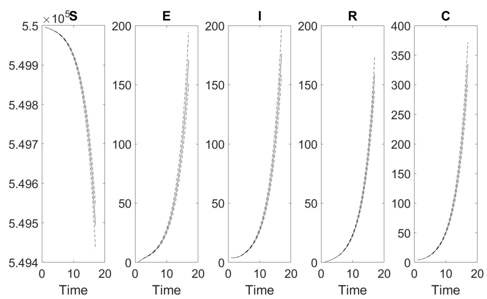


Basic reproduction number for NLS by Poisson error structure

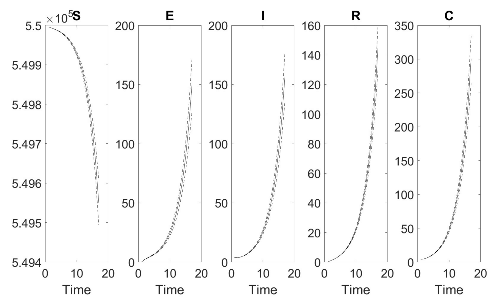


Basic reproduction number for MLE by Poisson error structure

These graphs demonstrate the basic reproduction number, R_0, calculated from the NLS and MLE methods. R_0 is a crucial epidemiological metric indicating the expected number of secondary infections from one case. Both methods provide similar estimates, but the MLE model shows a slightly higher range.



Bootstrap realization by NLS using Poisson error structure

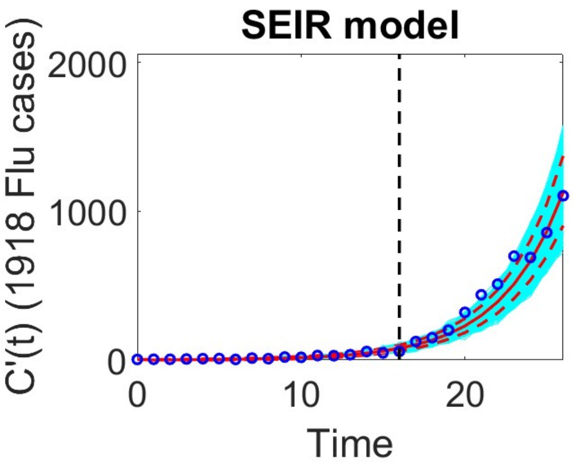


Bootstrap realization by MLE using Poisson error structure

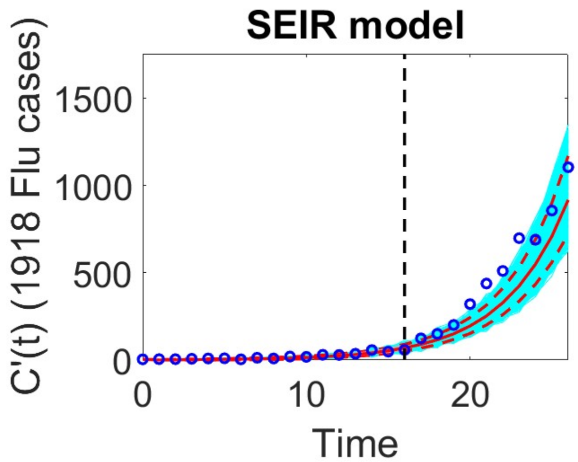
These graphs show the results from bootstrapping, a method used to quantify the uncertainty in parameter estimation. Each realization represents a possible outcome based on resampled data. The spread of these lines indicates the variability in the transmission rate estimates, with MLE showing slightly more concentrated outcomes than NLS.

Forecasting

We generated a 10-day forward projection using the best-fit model of the influenza incidence. The graphs of these projection are given below.

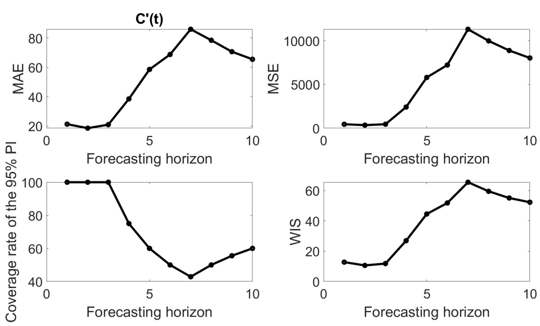


10-day model forecast by NLS

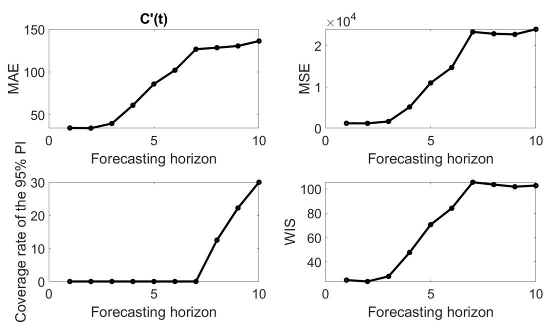


10-day model forecast by MLE

These graphs provide the 10-day forecast of the epidemic using NLS and MLE. Both forecasts project future incidence rates based on the best-fit models. While the forecasts are similar, the MLE forecast exhibits tighter uncertainty bounds compared to the NLS forecast, indicating better confidence in predictions.



Performance metric of NLS



Performance metric of MLE

These graphs compare the Mean Absolute Error (MAE), Mean Squared Error (MSE), and coverage of the 95% Prediction Interval (PI) for NLS and MLE models. MLE consistently outperforms NLS in terms of lower error metrics and better coverage of the 95% PI, though neither method reaches the ideal 95% PI coverage.

The table below presents the estimated transmission rate parameter β along with its confidence intervals from the bootstrapping procedure. Note that all the other parameters were fixed:

Method	Parameter	Estimate	CI
Maximum Likelihood Estimation (MLE)	β	0.77	0.75-0.79
Nonlinear Least Squares (NLS)	β	0.75	0.73-0.77

Performance Evaluation

The evaluation of forecast performance involving: Mean Absolute Error (MAE), Mean Squared Error (MSE), the coverage of the 95% Prediction Interval (PI), and the Weighted Interval Score (WIS) are presented in the tables below.

Metric	NLS	MLE
Mean Absolute Error (MAE)	5.696	5.737
Mean Squared Error (MSE)	48.342	60.445
Coverage of the 95% PI	47.059%	64.706%
Weighted Interval Score (WIS)	3.834	3.895

Comparison of performance calibration of NLS and MLE

Metric	NLS	MLE
Mean Absolute Error (MAE)	136.176	65.465
Mean Squared Error (MSE)	24013.770	8022.81567
Coverage of the 95% PI	30%	60%
Weighted Interval Score (WIS)	102.686	52.242

Comparison of Forecast Performance of NLS and MLE

From the tables, it is evident that the forecast performance indicates greater reliability in the intervals generated by Maximum Likelihood Estimation (MLE) compared to those produced by Nonlinear Least Squares (NLS).

Regarding potential improvements to the model's forecasting performance, While Maximum Likelihood Estimation (MLE) demonstrated better coverage of the 95% prediction interval compared to Nonlinear Least Squares (NLS), it still did not achieve the ideal 95%. To enhance this, refining the parameter estimation process by employing an alternative method could be beneficial. Also, the model could be improved by incorporating more time-series data for better fitting and parameter estimation.

Conclusion

In this work, parameter estimation and forecasting for the 1918 influenza pandemic were performed using NLS and MLE methods. Both approaches produced reasonable parameter estimates and forecasts, but MLE demonstrated superior performance in terms of fit accuracy and reliability. Although MLE provided better prediction interval coverage, further improvements could be made by enhancing the parameter estimation process and incorporating more comprehensive data. Ultimately, MLE is preferable for reliable forecasting in epidemic modeling, though both methods have their strengths depending on the specific application.