Flu Cases in the United States

Abstract

We analyze weekly U.S. flu case counts using both classical time-series methods and a mechanistic partially observed Markov process model. Focusing on 2016–2019, we find strong yearly seasonality from decomposition and frequency analysis. We then fit ARMA-type baselines and a seasonal SEIRS model, with parameter estimation by local and global search. Overall, the global SEIRS fit improves substantially over the local fit and provides a mechanistic description of the seasonal outbreak pattern.

1 Introduction

Flu is a recurring seasonal disease in the United States, with weekly case counts that vary across years. In this project, we analyze weekly U.S. flu data from 1997 to 2021.Our goal is to assess whether classical time-series models and partially observed Markov process (POMP) models can capture the dynamics of weekly flu cases. This is useful because flu data show strong seasonality, short-term dependence, and epidemic-like fluctuations (Hyndman and Khandakar 2008). We begin with exploratory analysis, STL decomposition, and frequency-domain methods to identify seasonal structure (Cleveland et al. 1990). We then fit ARMA and SARMA models and compare them using AIC and residual diagnostics (Hyndman and Khandakar 2008). We also build a POMP model based on compartmental epidemic dynamics. POMP models link a latent stochastic transmission process to noisy observed data (King, Nguyen, and Ionides 2016). Parameters are estimated with local and global search using iterated filtering (Ionides, Bret'o, and King 2011). Finally, we compare the statistical and mechanistic approaches in explaining the U.S. weekly flu series.

2 Methods

2.1 Data and preprocessing

We use weekly U.S. flu case counts and restrict the main modeling analysis to 2016–2019. This provides several seasonal cycles while keeping the analysis computationally manageable.

2.2 Exploratory and statistical time-series methods

We begin with exploratory plots, STL decomposition, and periodograms to study seasonality. We then fit ARMA and seasonal ARMA models using log transformation, differencing, autocorrelation diagnostics, and AIC-based comparison. ARMA-type models are included mainly as statistical benchmarks. They can summarize autocorrelation in the transformed series, but they do not represent epidemic mechanisms such as transmission, depletion of susceptibles, waning immunity, or seasonally varying infection pressure. For this reason, our main scientific model is the seasonal SEIRS POMP model, while ARMA and SARMA are used primarily for comparison.

2.3 Mechanistic POMP modeling

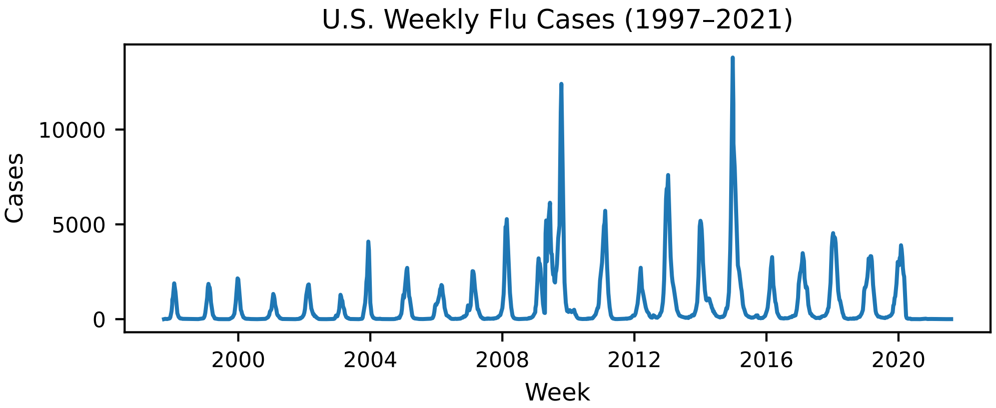
We use a seasonal SEIRS model within a partially observed Markov process framework. Reported weekly cases are linked to latent incidence through a negative-binomial observation model, and parameters are estimated using particle filtering and iterated filtering with local and global search.

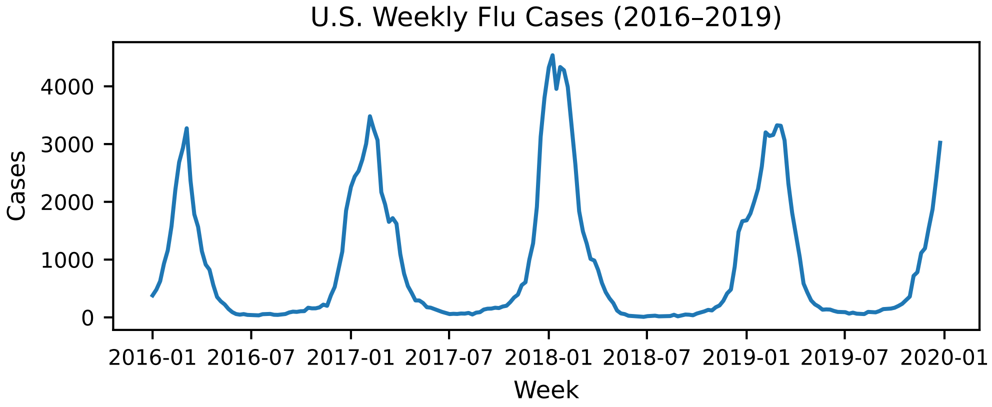
3 Results

3.1 Exploratory analysis

The dataset contains weekly total flu case counts in the United States from 1997 to 2021. We first visualize the full series to understand its long-term behavior, seasonal structure, and variation in outbreak intensity across years.

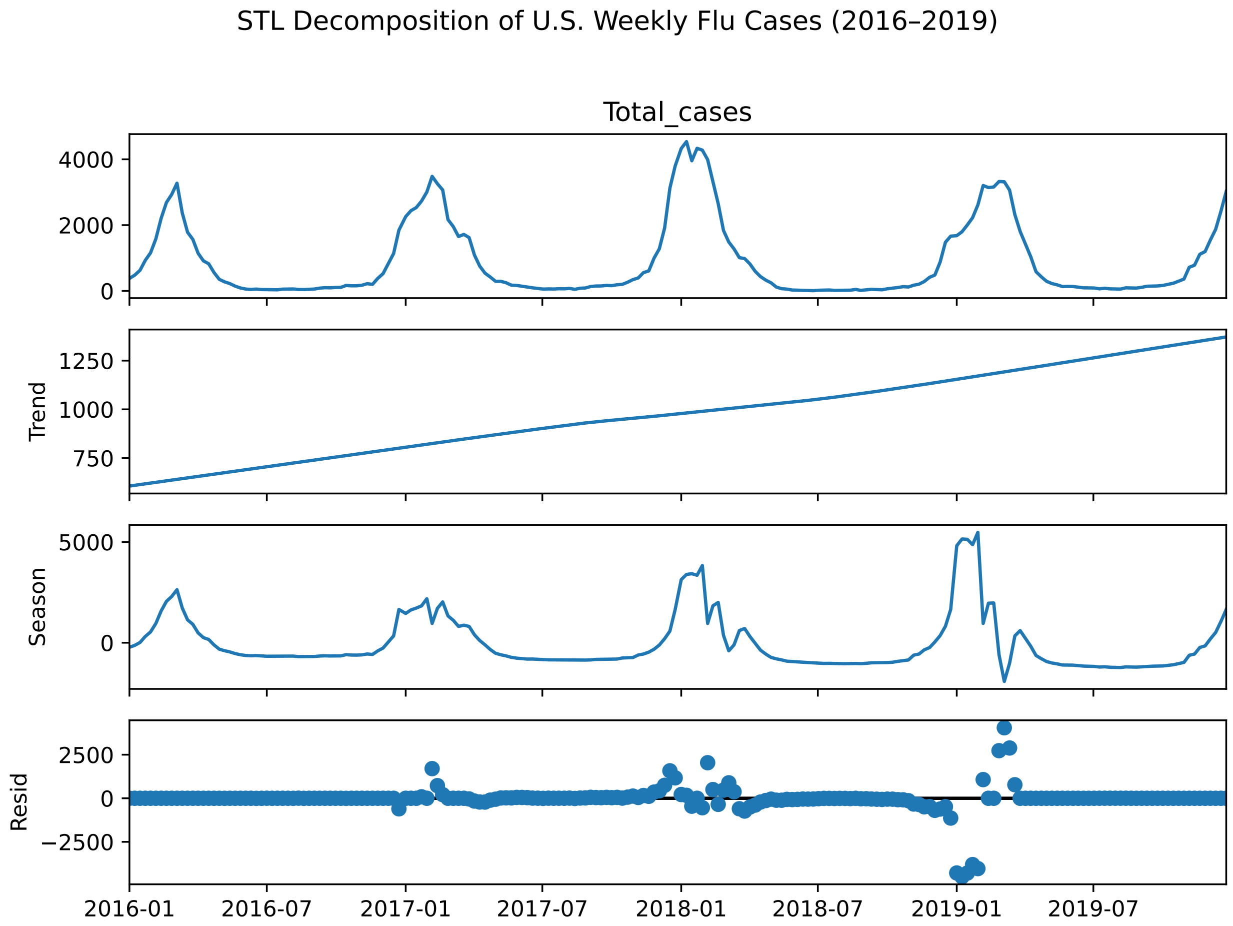
For the main modeling analysis, we focus on 2016 to 2019. This shorter window still captures multiple seasonal cycles while keeping filtering and parameter search computationally manageable. It also avoids the unusual disruption in 2020 and 2021 during the COVID-19 period. We therefore present both the full series and the shortened series.





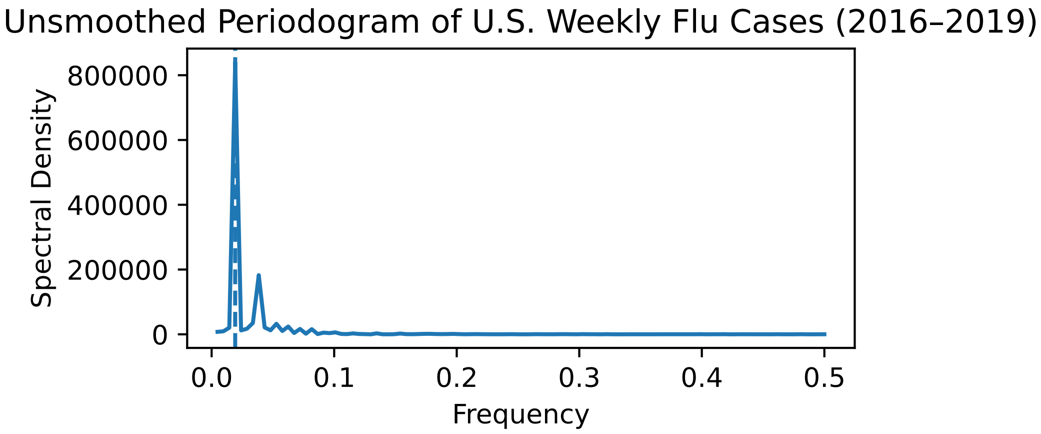
3.2 Decomposition

From the decomposition plot below, we observe a seasonal pattern that likely follows a yearly cycle. The trend changes gradually over time, while a few larger residual fluctuations remain near peak seasons.



3.2.1 Frequency Analysis

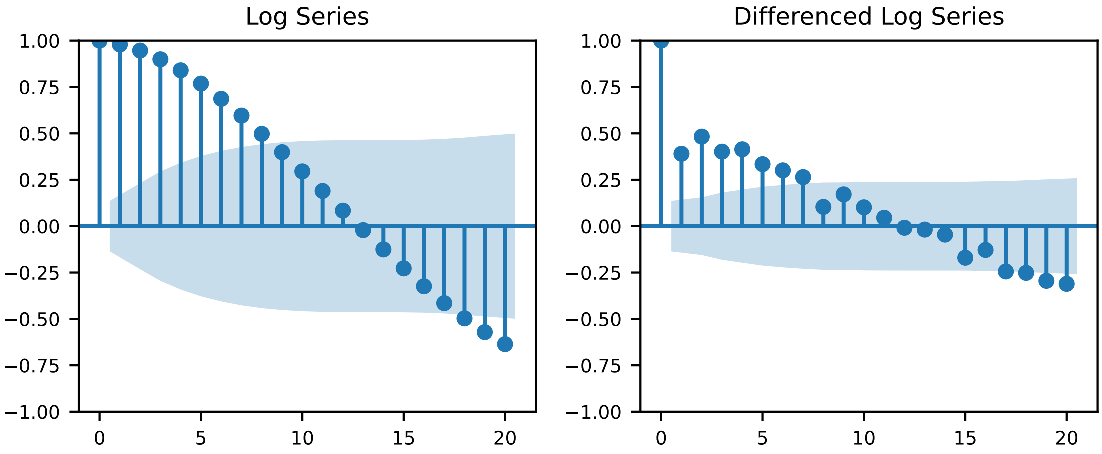
In this part, we analyze the data in the frequency domain to investigate seasonality. From the periodogram below, the dominant frequency is about 0.0192. This corresponds to a period of $T = \frac{1}{\omega} = \frac{1}{0.0192} \approx 52$ weeks. This suggests that the main seasonal pattern in the weekly flu data occurs approximately once every year.



3.3 ARMA and SARMA models

We next fit ARMA models to the restricted 2016–2019 flu series as a baseline time series approach. Since the earlier analysis showed strong yearly seasonality, the ARMA model serves mainly as a reference before moving to seasonal or mechanistic models.

To stabilize the variance, we apply a log transformation of the form log(1 + x), and then examine the autocorrelation structure before fitting the model.



The ACF of the log-transformed series shows slow decay, so the series is still not stationary after transformation. After first differencing, the autocorrelations are reduced, though some seasonal dependence remains.

We therefore use the differenced log series for baseline ARMA modeling.

3.3.1 ARMA

An ARMA(p, q) model can be written as

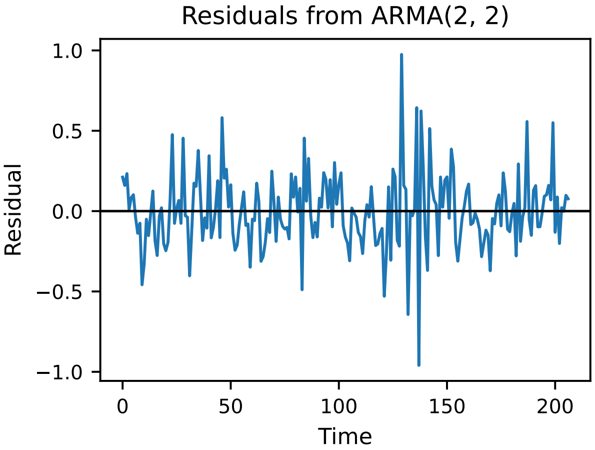
$$\phi(B)(Y_t - \mu) = \psi(B)\varepsilon_t.$$

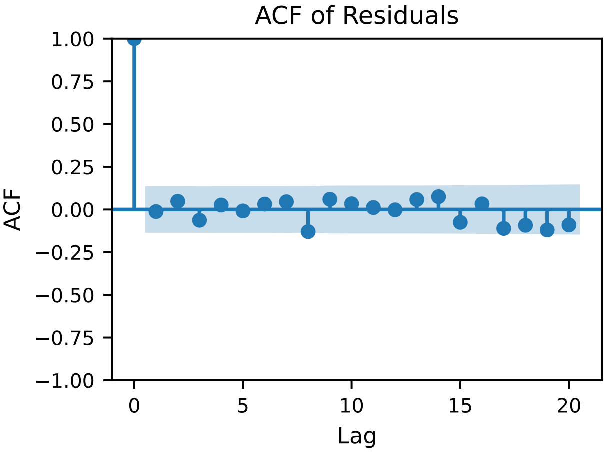
We compare candidate models using

$$\mathrm{AIC} = -2\ell(\hat{\theta}) + 2D,$$

and choose the model with the smallest AIC.

The AIC table for models with $p, q \in \{0, 1, 2\}$ suggests that ARMA(2,2) provides the best fit among the candidates considered, with the smallest AIC value of -7.012. We therefore select ARMA(2,2) as the baseline non-seasonal model for the differenced log series.





```
Ljung-Box test (lag 20): statistic = 20.348, p-value = 0.436
```

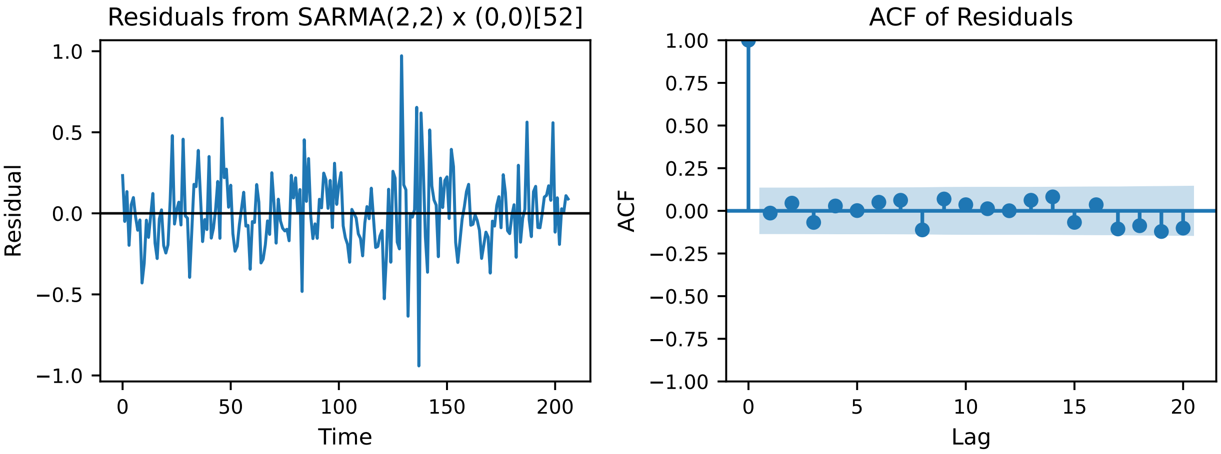
The residual diagnostics suggest that ARMA(2,2) captures much of the short-term dependence in the transformed series. The residual ACF is small, and the Ljung–Box test (p-value = 0.436) does not indicate remaining autocorrelation. Overall, ARMA(2,2) provides a reasonable baseline fit, though the strong seasonality in the data suggests that a seasonal model may be more appropriate.

3.3.2 SARMA

We next consider SARMA models to account for the strong yearly seasonality in the weekly flu series. For weekly data, the general SARMA$(p, q) \times (P, Q)_{52}$ model is

$$\phi(B)\,\Phi(B^{52})\,(Y_t - \mu) = \psi(B)\,\Psi(B^{52})\,\varepsilon_t.$$

Since ARMA$(2, 2)$ gave the best baseline AIC, we fix $p = 2$ and $q = 2$ and search over seasonal orders P and Q. The seasonal AIC comparison selected seasonal order $(P, Q) = (0, 0)$, with best SARMA AIC = -5.4, indicating that adding seasonal terms did not improve the fit over the baseline ARMA model.



```
Ljung-Box test (lag 20): statistic = 20.826, p-value = 0.407
```

Although the original weekly flu series shows clear yearly seasonality, the models are fitted to the first-differenced log series, where much of that structure has already been reduced. In addition, the restricted 2016–2019 window contains only a few yearly cycles, making seasonal parameters harder to estimate reliably. As a result, adding seasonal terms did not improve the fit enough to justify the extra model complexity.

3.4 SEIRS Model

We model the weekly flu data using a seasonal SEIRS model within the partially observed Markov process framework. The model allows individuals to move through susceptible, exposed, infectious, and recovered states, with waning immunity returning recovered individuals to the susceptible class. A seasonal transmission rate is included to capture the yearly pattern in flu activity.

3.4.1 Model Definition

The compartment transitions are:

- $S(t) \to E(t)$: new infections
- $E(t) \to I(t)$: progression from exposed to infectious
- $I(t) \to R(t)$: recovery
- $R(t) \to S(t)$: waning immunity

The transition rates are

$$\frac{dN_{SE}}{dt} = \beta(t)\frac{S(t)I(t)}{N}$$

$$\frac{dN_{EI}}{dt} = \mu_{EI}E(t)$$

$$\frac{dN_{IR}}{dt} = \mu_{IR}I(t)$$

$$\frac{dN_{RS}}{dt} = \mu_{RS} R(t)$$

where N is the total population size.

Seasonality is modeled through

$$\beta(t) = \beta_0 \left[1 + \text{amp} \cos\left(\frac{2\pi(t + \text{phase})}{52} \right) \right],$$

where β_0 is the baseline transmission rate, amp controls the strength of seasonal variation, and phase shifts the timing of the seasonal peak.

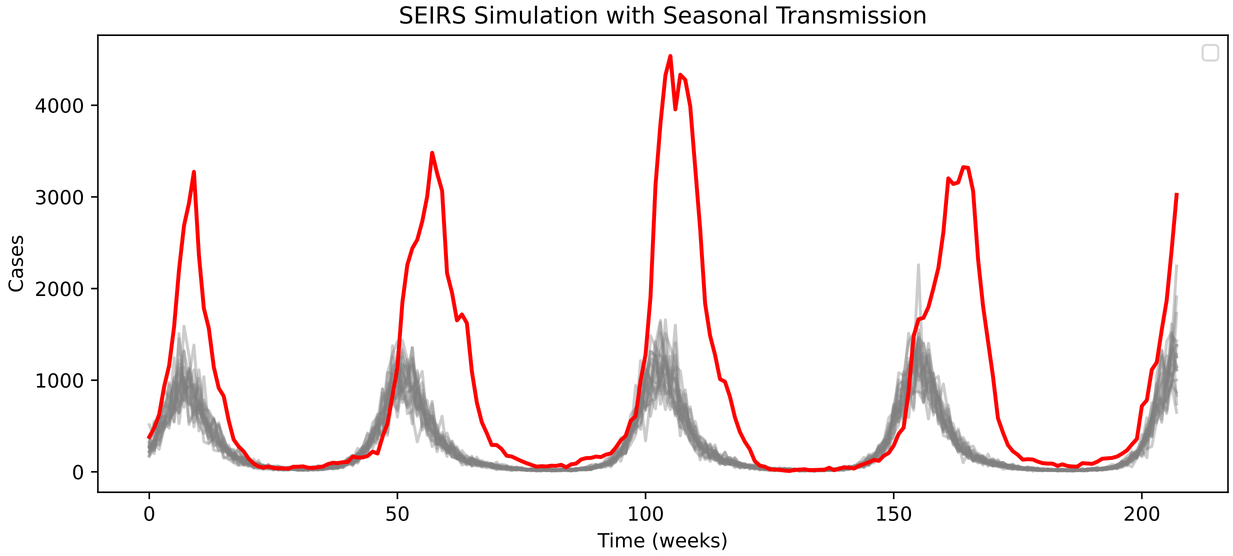
Using a discrete-time approximation, the transitions over a small interval Δt are modeled as

$$\Delta N_{SE} \sim \text{Binomial}\left(S(t),\ 1 - \exp\left(-\beta(t)\frac{I(t)}{N}\Delta t \right) \right)$$
$$\Delta N_{EI} \sim \text{Binomial}\left(E(t),\ 1 - \exp(-\mu_{EI}\Delta t) \right)$$
$$\Delta N_{IR} \sim \text{Binomial}\left(I(t),\ 1 - \exp(-\mu_{IR}\Delta t) \right)$$
$$\Delta N_{RS} \sim \text{Binomial}\left(R(t),\ 1 - \exp(-\mu_{RS}\Delta t) \right)$$

The observed weekly cases are linked to the latent infections through

$$Y_t \mid H_t \sim \text{NegBin}(\text{mean} = \rho H_t,\ \text{dispersion} = k),$$

where ρ is the reporting fraction and k controls overdispersion.

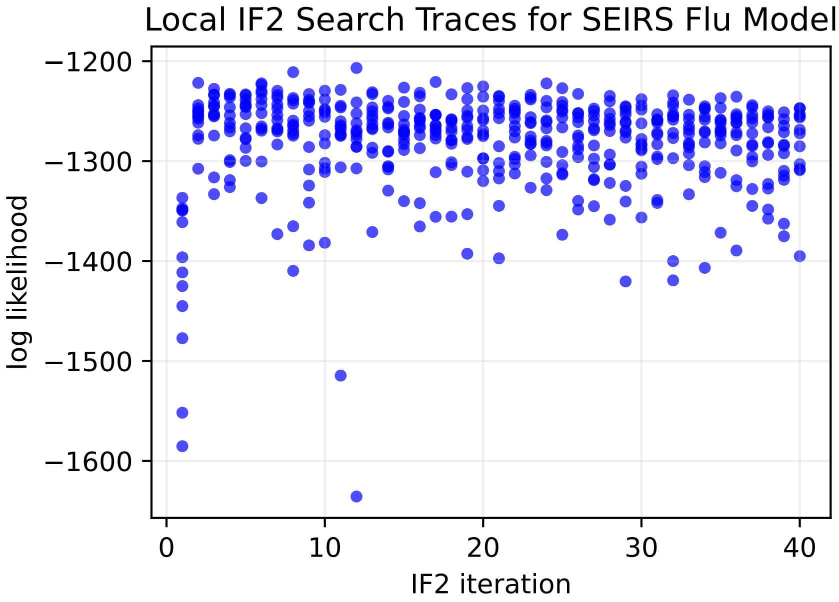


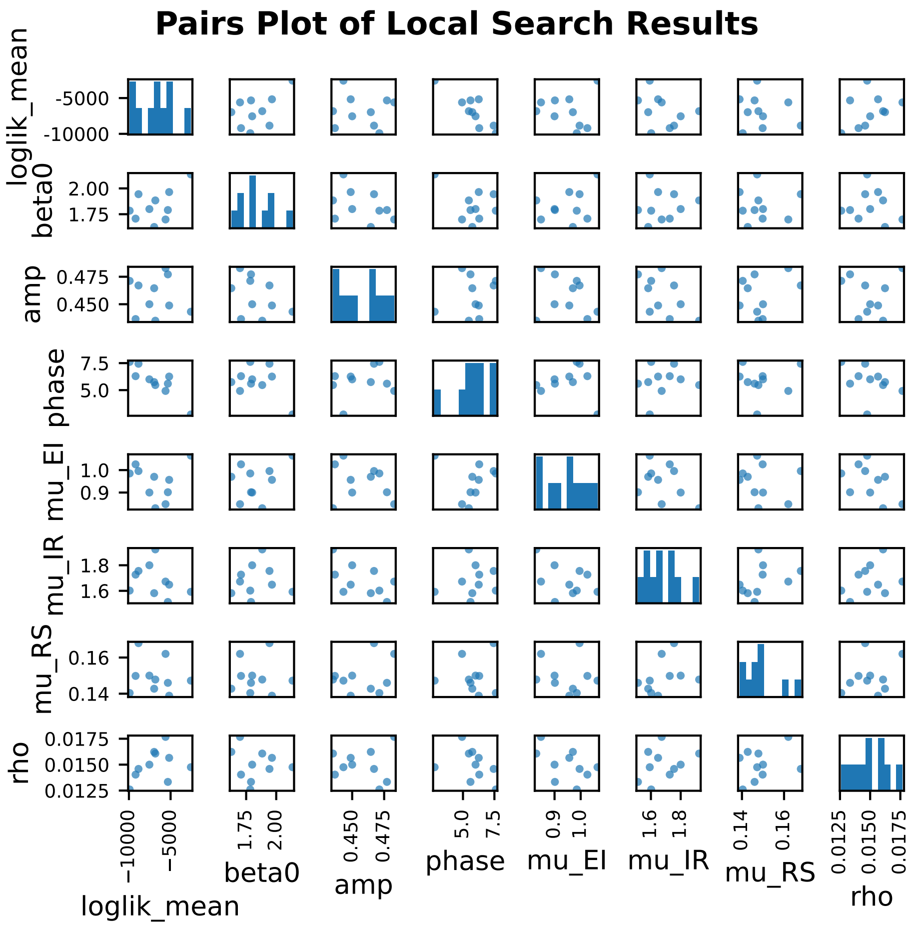
The SEIRS simulations capture the main yearly outbreak pattern and seasonal timing in the data, although some observed peaks are still underestimated. This provides a reasonable mechanistic starting point before formal parameter estimation.

3.5 Local Search

We estimated the SEIRS parameters using a local search based on iterated filtering. The IF2 trace plot shows movement toward a higher-likelihood region, although some variability remains across replicates.

```
Best local loglik = -2565.89 (beta0=2.136, amp=0.443, phase=2.745, rho=0.0148, k=12.881)
```



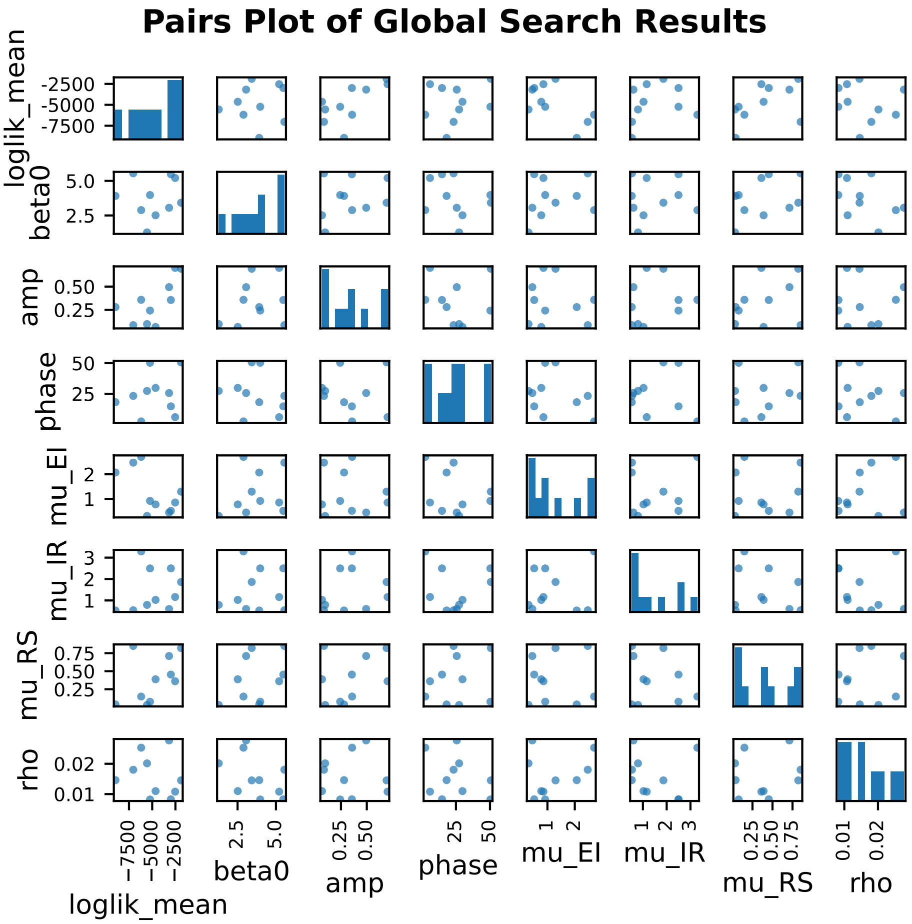


3.5.1 Global Search

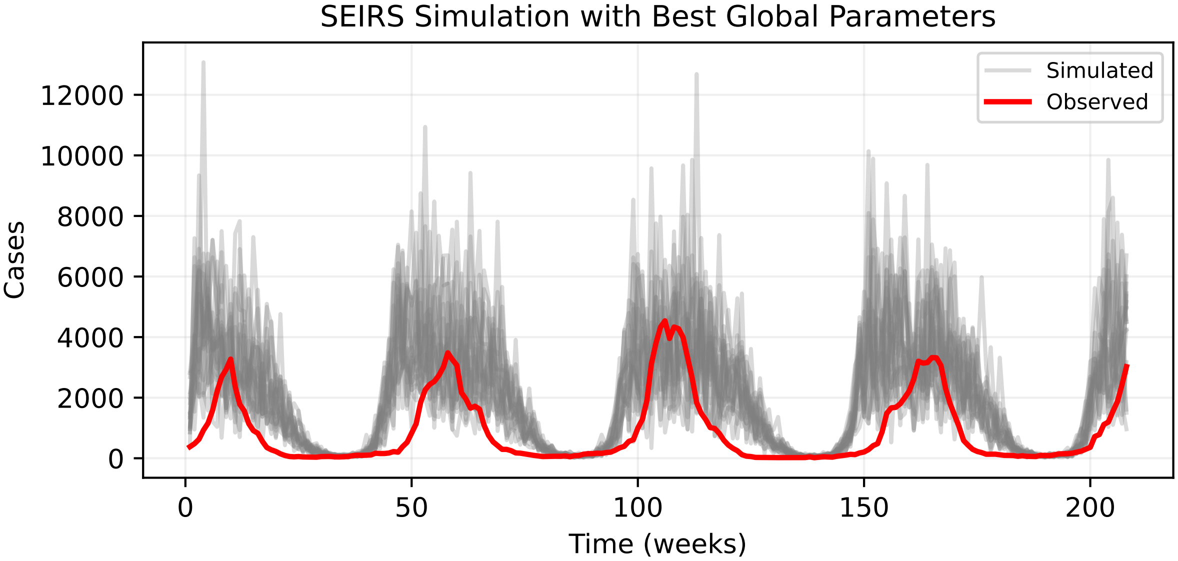
We next conducted a global search using multiple starting points and evaluated the fitted endpoints by particle filtering. The best global-search parameter set had beta0 = 3.422, amp = 0.691, phase = 50.634, mu_EI = 1.295, mu_IR = 1.861, mu_RS = 0.822, rho = 0.0145, and k = 5.021, with an estimated log likelihood of -1892.04.

This is a clear improvement over the best local-search log likelihood of -2565.59. The fitted simulations capture the main seasonal outbreak pattern, although they remain more variable than the observed series.

```
Best global fit: loglik=-1892.08, beta0=3.422, amp=0.691, phase=50.634, rho=0.0145, k=5.021
```



Pairs Plot of Global Search Results



SEIRS Simulation with Best Global Parameters

3.6 Model Comparison

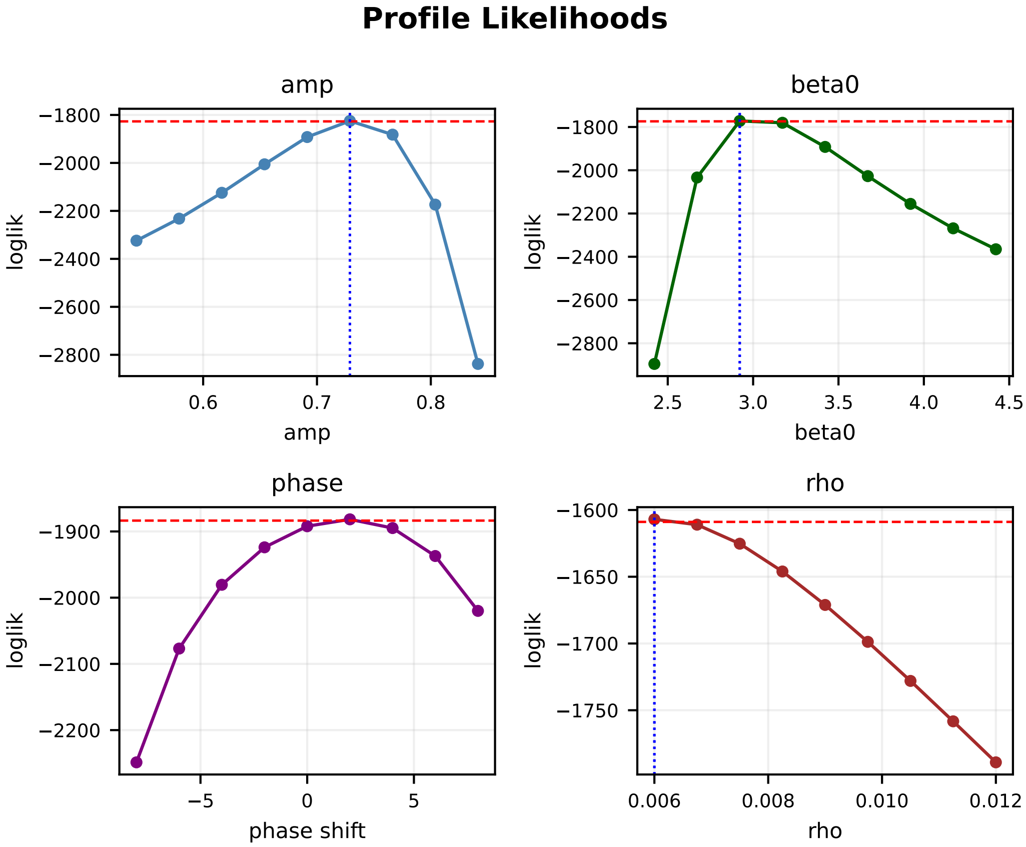
The ARMA(2,2) model provides a useful statistical baseline for the restricted flu series, but it does not directly represent epidemic transmission or seasonally changing susceptibility. The seasonal time-series analysis confirms a strong yearly pattern, with a dominant period of about 52 weeks.

The SEIRS model provides a mechanistic description through disease-state transitions and seasonal transmission. Local search improved the fit, and global search improved it further, raising the best log likelihood from -2565.89 to -1892.08.

Overall, the SEIRS model gives a more informative description of the epidemic dynamics, although its simulated trajectories remain more variable than the observed data. This over-variability suggests that, although the model captures the timing and broad seasonal structure of outbreaks, it does not fully reproduce the stability of the observed weekly series. It also suggests that some parameters may be only weakly identified, with different parameter combinations producing similar qualitative behavior. We therefore interpret the SEIRS fit mainly as a mechanistic explanation of seasonal flu dynamics rather than as a fully calibrated quantitative model.

3.7 Profile Likelihood

To study parameter uncertainty, we compute profile likelihoods for selected SEIRS parameters and examine how the fitted log likelihood changes as each parameter is varied.



The profile likelihoods for beta0 and amp show clear peaks, suggesting that these parameters are reasonably well identified. The phase profile is best interpreted near the seasonal boundary, while rho remains boundary-limited and is less reliably identified.

3.8 Conclusion

In this project, we analyzed the U.S. weekly flu series using both classical time-series methods and a mechanistic POMP framework. The exploratory analysis, STL decomposition, and frequency analysis all showed strong annual seasonality, indicating that yearly structure is a dominant feature of the data.

The ARMA model provided a useful statistical baseline for the transformed series, but it does not directly represent epidemic transmission or changing susceptibility. In contrast, the seasonal SEIRS model gave a mechanistic interpretation through disease-state transitions and seasonal transmission.

Parameter estimation substantially improved the SEIRS fit. Local search identified a promising region of parameter space, and global search improved the best log likelihood from -2565.89 to -1892.08. The fitted simulations also reproduced the main timing and general magnitude of the seasonal outbreaks.

Overall, the main conclusion is that the flu series is strongly seasonal and that the SEIRS model provides a richer explanation of the epidemic dynamics than the statistical baseline. However, the fitted simulations remain more variable than the observed data, so the final model should be viewed as a strong qualitative and mechanistic fit rather than a perfectly calibrated quantitative one. This likely reflects remaining model misspecification or partial parameter non-identifiability, where multiple parameter settings can generate similar seasonal patterns. As a result, the fitted parameter values should be interpreted cautiously, with greater emphasis on the model's ability to explain the seasonal mechanism than on exact numerical calibration.

Scholarship

This project was developed in the context of earlier STATS 531 influenza projects, especially Modeling Flu Cases in Oklahoma (Winter 2024 Project 5), Influenza cases in the Netherlands (Winter 2024 Project 16), Flu Cases in Michigan (Winter 2025 Project 3), and Influenza in Nova Scotia (Winter 2025 Project 14) (STATS/DATASCI 531 Winter 2024 2024b, 2024a; STATS/DATASCI 531 Winter 2025 2025a, 2025b). Like these earlier studies, our report combines exploratory time-series analysis with mechanistic epidemic modeling, but our contribution is different in focusing on weekly U.S. flu counts from 2016–2019 and in centering the final analysis on a seasonal SEIRS model fitted by likelihood-based methods rather than on descriptive analysis alone.

The peer reviews of earlier flu projects directly shaped the presentation of our report. Reviews of prior projects repeatedly noted that ARMA analysis for epidemic count data is often better handled on a log scale and should be treated mainly as a benchmark for mechanistic modeling rather than the main scientific result (STATS/DATASCI 531 Winter 2024 2024e, 2024d; STATS/DATASCI 531 Winter 2025 2025d, 2025c). Other recurring comments concerned weak in-text citation practice, unclear explanation of creative contribution, overstated conclusions(STATS/DATASCI 531 Winter

2024 2024c, 2024d; STATS/DATASCI 531 Winter 2025 2025d). In response, we use ARMA mainly as a baseline, keep the main emphasis on the fitted SEIRS model, state more clearly how our work relates to previous 531 flu projects, and aim for more cautious interpretation of what the fitted model does and does not explain.

Acknowledgments

AI assistance (ChatGPT) was used for adapting code, debugging, code cleanup and formatting support while preparing this Quarto report.

Bibliography

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." *Journal of Official Statistics* 6 (1): 3–33.

Hyndman, Rob J., and Yeasmin Khandakar. 2008. "Automatic Time Series Forecasting: The Forecast Package for R." *Journal of Statistical Software* 27 (3): 1–22. https://doi.org/10.18637/jss.v027.i03.

Ionides, Edward L., Carles Bret'o, and Aaron A. King. 2011. "Iterated Filtering." *The Annals of Statistics* 39 (3): 1776–1802. https://doi.org/10.1214/10-AOS886.

King, Aaron A., Dao Nguyen, and Edward L. Ionides. 2016. "Statistical Inference for Partially Observed Markov Processes via the R Package Pomp." *Journal of Statistical Software* 69 (12): 1–43. https://doi.org/10.18637/jss.v069.i12.

STATS/DATASCI 531 Winter 2024. 2024a. "Influenza Cases in the Netherlands." 2024. https://ionides.github.io/531w24/final_project/project16/blinded.html.

———. 2024b. "Modeling Flu Cases in Oklahoma." 2024. https://ionides.github.io/531w24/final_project/project05/blinded.html.

———. 2024c. "Review Comments on Project 10." 2024. https://ionides.github.io/531w24/final_project/project10/comments.html.

———. 2024d. "Review Comments on Project 16." 2024. https://ionides.github.io/531w24/final_project/project16/comments.html.

———. 2024e. "Review Comments on Project 5." 2024. https://ionides.github.io/531w24/final_project/project05/comments.html.

STATS/DATASCI 531 Winter 2025. 2025a. "Flu Cases in Michigan." 2025. https://ionides.github.io/531w25/final_project/project03/blinded.html.

———. 2025b. "Influenza in Nova Scotia." 2025. https://ionides.github.io/531w25/final_project/project14/blinded.html.

———. 2025c. "Review Comments on Final Project 14." 2025. https://ionides.github.io/531w25/final_project/project14/comments.html.

———. 2025d. "Review Comments on Final Project 3." 2025. https://ionides.github.io/531w25/final_project/project03/comments.html.