# Modelling Polio Cases pre and post-Vaccine

## 1 Introduction

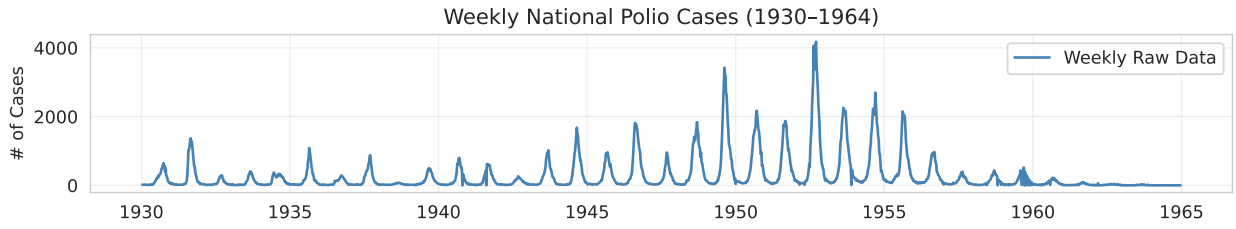
Polio (Poliomyelitis) is an infectious disease caused by the poliovirus, historically paralyzing thousands of children per day ("History of Polio," n.d.). The invention of the polio vaccine in 1955 has saved tens of millions from paralytic polio (Kamran Badizadegan 2022). Of importance to understanding diseases is the ability to model long term spread. Polio offers an interesting challenge - it has been eradicated in the United States, and may exhibit an atypical pattern of infections over our studied time period, dropping dramatically towards the end.

We aim to model polio infections from 1930 to 1964, spanning the introduction of the vaccine and beyond. We wish to ask the question of whether introducing a fundamental change to the behavior of the disease (introducing a vaccine in 1955) improves the performance of mechanistic models when compared to a SARIMAX model. Previous work modeled diseases using mechanistic and non-mechanistic models. In the case of ("Unveiling the Dynamics of Influenza in the Great Lakes Region" 2025) specifically, a SARMA model outperformed their mechanistic POMP model in terms of biological plausability and avoiding overfitting. In contrast, we set out to see if this holds for a different disease where there is an underlying change in the progression of the disease. We sourced our data from Project Tycho ("United States of America (Acute Poliomyelitis) Dataset," n.d.).

## 2 Exploratory Data Analysis

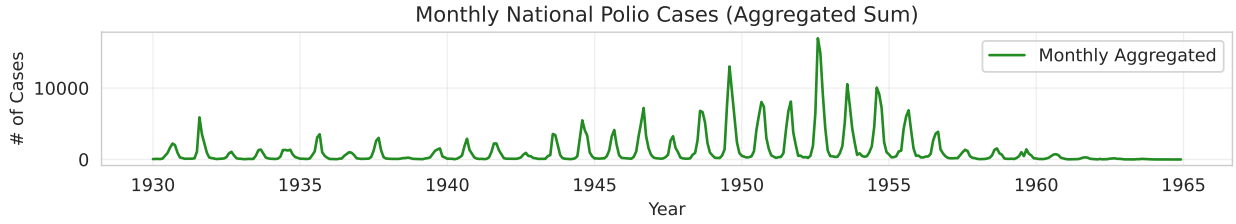
The raw data contains case counts at the levels of state, county, and city, with flags for fatalities, cumulative counts, and more. We filtered the data down to weekly national polio cases.

We only included records where fatalities were zero, the record was not part of a cumulative count series, and the record was at the state level (i.e. no county or city name). This left us with a dataset of weekly state-level polio cases that didn't include fatalities or double-count from cumulative numbers and overlapping administrative divisions. We aggregated this data by summing the cases across all states for each week, giving us a national weekly case count.




We filtered this data from January 1st, 1930 to December 31st, 1964, which is the period we are interested in modeling. The introduction of the vaccine means that cases essentially drop to zero

beyond 1965, which we observed to cause problems for model fitting. We binned weekly data into monthly data, yielding a monthly case count which was computationally easier to model.




# 3 Methods

We model monthly national paralytic-polio case counts from 1930 through 1964 using a Time-Series SIR (TSIR) POMP model (Bärbel F. Finkenstädt 2000). The model couples a stochastic discrete-time transmission process on the raw case scale to a Gaussian observation model on the $\log(1+y)$ scale, and is estimated by iterated filtering (Ionides 2026b). Three considerations shaped the design: the ~3-order-of-magnitude span of monthly counts motivates a log-scale observation model; the 1955 vaccine introduction motivates an explicit post-1955 shift in the log-transmission rate; and unexplained year-to-year epidemic size variation motivates a latent annual amplitude shock.

## 3.1 Data

Let $y_t$ denote the reported monthly count in month $t = 1, \dots, T$, with $T = 420$ (January 1930 through December 1964). We fit the observation model on the transformed scale $z_t = \log(1 + y_t)$, while the latent epidemic process is defined on the raw count scale. This is equivalent to a lognormal observation model on $(1 + Y_t)$ and is numerically well-behaved at both low and high counts. Three covariates are derived from the calendar: $m_t \in \{1, \dots, 12\}$: calendar month, encoded as an angle $\phi_t = 2\pi(m_t - 1)/12$; $\text{post}_t \in \{0, 1\}$, indicator that $t$ falls on or after January 1955; $\text{yc}_t \in \{0, 1\}$, indicator that $t$ is the first observation of a new calendar year.

## 3.2 State-space model

The latent state at time $t$ is $X_t = (S_t, I_t, A_t)$: $S_t$, effective susceptible count; $I_t$, latent true monthly case intensity (raw-count scale); $A_t$, annual amplitude shock, constant within a calendar year.

### 3.2.1 Process model

The one-step conditional mean of $I_{t+1}$ is driven by a log-transmission rate that combines a baseline level, a seasonal component, a post-1955 regime shift, the annual shock, a susceptible-depletion term, and an endogenous feedback on lagged incidence:

$$\log \lambda_t \;=\; \log \beta_0 \;+\; f_{\text{seas}}(m_t) \;+\; \delta_{\text{post}} \cdot \text{post}_t \;+\; A_t \;+\; \log\!\left(\frac{S_t + 1}{N_{\text{eff}} + 1}\right) \;+\; \alpha \, \log(I_t + 1).$$

The seasonal forcing uses a 2-harmonic trigonometric basis,

$$f_{\text{seas}}(m_t) \;=\; a_1 \cos \phi_t + b_1 \sin \phi_t + a_2 \cos 2\phi_t + b_2 \sin 2\phi_t,$$

which captures the dominant annual cycle and a secondary semi-annual component. The annual amplitude shock is redrawn at year boundaries and held constant within a year:

$$A_t \;=\; \begin{cases} \tilde{A}_t,\ \tilde{A}_t \sim \mathcal{N}(0, \sigma_A^2), & \text{if } \mathrm{yc}_t = 1, \\ A_{t-1}, & \text{otherwise,} \end{cases} \qquad |A_t| \le 4.$$

Process noise on incidence enters through a negative-binomial draw with mean $\lambda_t$ and dispersion $k_{\text{proc}}$, implemented via the Gamma–Poisson mixture

$$\Lambda_t \mid \lambda_t, k_{\text{proc}} \;\sim\; \text{Gamma}(k_{\text{proc}},\, \lambda_t/k_{\text{proc}}), \qquad I_{t+1} \mid \Lambda_t \;\sim\; \text{Poisson}(\Lambda_t).$$

The susceptible compartment follows a logistic-replenishment update:

$$S_{t+1} \;=\; S_t \;+\; \nu\left(1 - \frac{S_t}{N_{\text{eff}}}\right) \;-\; I_{t+1}, \qquad 0 \le S_{t+1} \le N_{\text{eff}}.$$

Unlike a constant births-per-month replenishment, logistic replenishment decays to zero as $S_t \to N_{\text{eff}}$, so the susceptible pool cannot pin to its upper bound and the depletion term $\log\{(S_t+1)/(N_{\text{eff}}+1)\}$ remains identified. We set $N_{\text{eff}} = \max\{2 \times 10^5,\ 35 \cdot q_{0.95}(y)\}$ as an effective susceptible scale rather than a census population, consistent with typical TSIR practice when reporting is incomplete.

### 3.2.2 Observation model

Conditional on $I_t$, the observed transformed count is Gaussian with location $\log(1 + I_t)$:

$$z_t \mid I_t \;\sim\; \mathcal{N}(\log(1 + I_t),\ \sigma_{\text{obs}}^2).$$

This is a multiplicative noise model on the raw scale that down-weights absolute errors at peaks relative to a raw-count Gaussian or negative-binomial observation model, and empirically produces substantially higher effective sample sizes in the particle filter than the raw-count alternative.

### 3.2.3 Initial conditions

$$S_0 = \eta\, N_{\text{eff}}, \qquad I_0 \sim I_0^{\text{par}} \cdot \text{LogNormal}(0, 0.15^2), \qquad A_0 = 0,$$

At $t = 0$, where $\eta \in (0, 1)$ is the initial susceptible fraction and $I_0^{\text{par}}$ is a parameter to be estimated.

### 3.2.4 Parameters and constraints

The model has 13 free parameters, summarized below. Estimation is carried out on a reparameterized scale $\theta \in \mathbb{R}^{13}$ that is unconstrained except for hard box bounds used as a numerical safeguard; IF2 perturbations are applied on this scale and inverse-transformed before each filter step. The upper bound on $\nu$ is set at $0.05\, N_{\text{eff}}$ per month, well below the level at which replenishment alone would exceed the susceptible pool, to prevent the parameter from drifting into regions where it is weakly identified. The feedback exponent $\alpha$ is restricted to $(0.6, 1.1)$.

| Parameter | Natural Scale | Transform | Interpretation |
|---|---|---|---|
| $\beta_0$ | $(0, 100)$ | log | Baseline monthly transmission rate |
| $a_1, b_1$ | $\mathbb{R}$ | identity | Annual seasonal coefficients |
| $a_2, b_2$ | $\mathbb{R}$ | identity | Semi-annual seasonal coefficients |
| $\delta_{\text{post}}$ | $\mathbb{R}$ | identity | Log-multiplier for post-1955 regime |
| $\alpha$ | $(0.6, 1.1)$ | scaled logit | Epidemic feedback exponent on $I_t$ |
| $\nu$ | $(1, 0.05N_{\text{eff}})$ | log | Logistic replenishment rate |
| $k_{\text{proc}}$ | $(10^{-3}, 10^3)$ | log | Negative-binomial process dispersion |
| $\sigma_A$ | $(10^{-3}, 2)$ | log | Standard deviation of annual amplitude shock |
| $\eta$ | $(0, 1)$ | logit | Initial susceptible fraction |
| $I_0^{\text{par}}$ | $(1, I_{\max})$ | log | Initial latent intensity |
| $\sigma_{\text{obs}}$ | $(10^{-3}, 5)$ | log | Log-scale observation standard deviation |

### 3.2.5 Particle filter

The one-step-ahead predictive likelihood is estimated by a bootstrap particle filter with systematic resampling (Kitagawa, 1996)(Kitagawa 2013). We have also discussed bootstrap particle filter in Chapter 14 (Ionides 2026a). Monte Carlo uncertainty in the log-likelihood is quantified by replicated evaluation: at any candidate $\theta$ we run $R$ independent particle filters with $N_p = 3000$ particles each and report the sample mean and standard error of $\widehat{\ell}$. For the final fit we use $R = 10$.

### 3.2.6 Global and local search

Because the likelihood surface is multimodal and IF2 is local, we use a two-stage search strategy. We use the iterated filtering (IF2) procedure as detailed in (Ionides 2026a).

**Global search:** We draw $K_g = 15$ random starting points $\theta^{(0),k}$ from a broad prior box calibrated to order-of-magnitude ranges for each parameter: $\beta_0 \sim \text{LogUnif}(0.2, 10)$, $\alpha \sim \text{Unif}(0.75, 1.02)$, $\delta_{\text{post}} \sim \text{Unif}(-2, -0.1)$, $\nu \sim \text{LogUnif}(50, 0.05N_{\text{eff}})$, $k_{\text{proc}} \sim \text{LogUnif}(2, 100)$, $\sigma_A \sim \text{LogUnif}(0.03, 0.5)$, $\eta \sim \text{Unif}(0.1, 0.8)$, $\sigma_{\text{obs}} \sim \text{LogUnif}(0.10, 1.20)$, and similar ranges for the remaining parameters. From each start we run IF2 with $N_p = 1000$ particles, $M = 50$ iterations, cooling $c = 0.5$, and a moderate perturbation vector $\sigma_0$ with standard deviations of 0.015–0.03 in the transformed parameter space. For each global run we retain the top 3 iterates by IF2-reported log-likelihood and evaluate each at $N_p = 3000$, $R = 10$.

**Local search:** The two candidates with the highest replicated log-likelihood from the global stage seed a refinement step with halved perturbation SD ($\sigma_0/2$) and a more aggressive cooling factor ($c = 0.3$), again for $M = 50$ iterations. All iterates from these local runs are pooled with the global candidates, re-evaluated at $N_p = 3000$, $R = 10$, and the candidate with the highest log-likelihood is taken as $\widehat{\theta}$. The Monte Carlo standard error of $\widehat{\ell}(\widehat{\theta})$ is reported with the point estimate.

### 3.2.7 Simulation Diagnostics

Beyond the replicated log-likelihood, we assess model fit with 3 diagnostics:

**Simulation from the fitted model:** We draw $S = 200$ independent sample paths from the POMP at $\widehat{\theta}$ and compare the pointwise 95% envelope and median of simulated $y_t$ to the observed series

on both the raw and $\log(1+y)$ scales. Coverage of the observed series by the simulation envelope provides a posterior-predictive check that is independent of the filter.

**Filter diagnostics:** Time-varying ESS and per-observation log-likelihood contributions are plotted to identify months where the filter degenerates or the model fits poorly.

**Peak and seasonal profile comparisons:** We compare annual peak height, peak timing, and mean monthly profiles between the filtered mean intensity and the observed series.

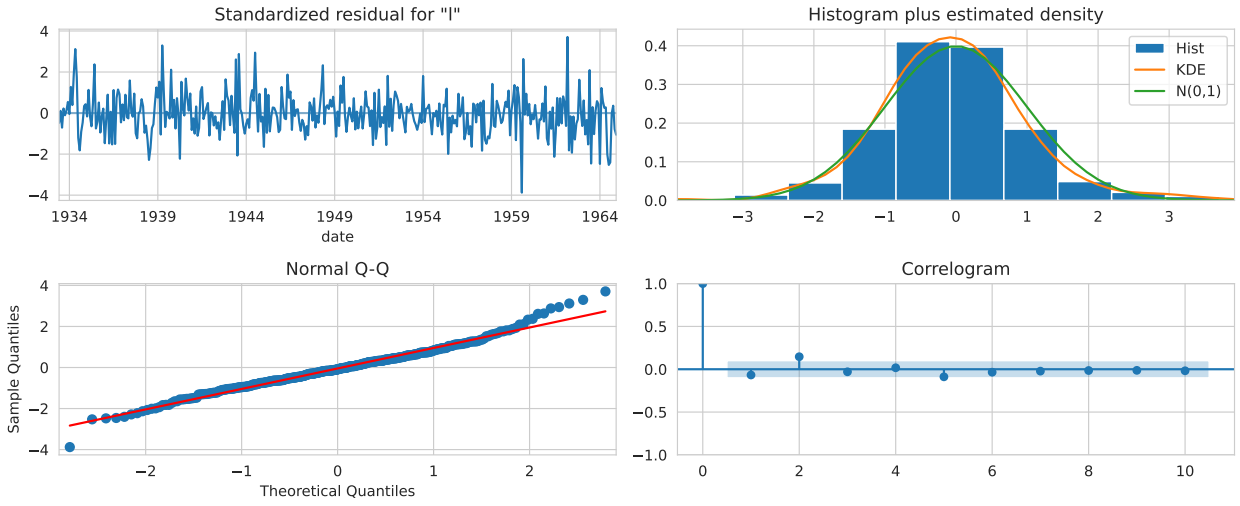
# 4 Results

## 4.1 SARIMAX Benchmarking

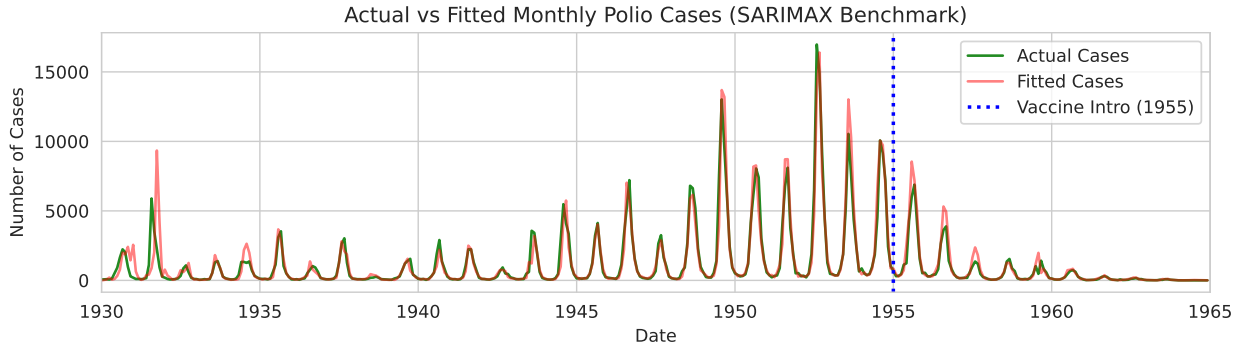
We used a SARIMAX model as a benchmark for our POMP model. The data was transformed to the log scale to stabilize variance and prevent negative predictions for case counts. We conducted a grid search over 288 different SARIMAX configurations, summmarized below. An exogenous indicator variable was added at 1955 to capture the regime change from vaccine introduction.

| Parameter | Type | Range / Value | Description |
|---|---|---|---|
| $p$ | Non-Seasonal AR | $\{0,1,2,3\}$ | Autoregressive order |
| $d$ | Non-Seasonal Diff | $\{0,1\}$ | Degree of differencing |
| $q$ | Non-Seasonal MA | $\{0,1,2,3\}$ | Moving average order |
| $P$ | Seasonal AR | $\{0,1,2\}$ | Seasonal autoregressive order |
| $D$ | Seasonal Diff | 1 | Seasonal differencing |
| $Q$ | Seasonal MA | $\{0,1,2\}$ | Seasonal moving average order |
| $m$ | Seasonal Period | 12 | Monthly seasonality |
| $I$ | Exogenous Variable | post-1955 | Vaccine introduction indicator |

We identified SARIMAX(3, 1, 3) x (1, 1, 2, 12) as the best model based on BIC, which was the lowest among all 288 models at 401.82. The diagnostics are shown below.



The Q-Q plot has outliers, but everything else seems well behaved; the KDE appears close to normal; there doesn't appear to be significant autocorrelation from the correlogram. Running a Ljung-Box test on the residuals confirms, however, that there is significant autocorrelation at both 12 and 24 lags, with p-values of 4.30e-09 and 1.13e-06 respectively. While this could be a sign of model misspecification, we believe that our search was thorough. Disease outbreaks are complex and non-linear, and this may be a shortcoming of SARIMAX.




The fitted values (red) appear to capture the overall trend and seasonality. It has a tendency to overpredict, and misses the timing of some of the early peaks. For a comparison with our POMP model, we calculated several metrics. The overall log likelihood of the model on the log scale is -168.25. The mean absolute error (MAE) and root mean squared error (RMSE) on the raw scale are 319.31 and 759.20 cases, respectively. The MAE and RMSE on the log scale are 0.3458 and 0.5138, respectively. Finally, the mean absolute error of the predicted peak value across years is 549.57 cases, and the mean absolute error of the predicted peak timing across years is 23.80 days. At a monthly resolution, this is a substantial error.

# 5 TSIR-POMP Results

## 5.1 Global and local search

We ran a global search of 15 randomly initialized IF2 chains (50 iterations each, $N_p = 1000$, cooling $c = 0.5$), retained the top-3 iterates by IF2-reported log-likelihood from each chain, and replicated each at $N_p = 3000$, $R = 10$. The two best candidates by replicated log-likelihood were then refined locally with halved perturbation SD and $c = 0.3$ for another 50 iterations. All candidates from both stages were re-evaluated under the same replicated protocol and the winner was selected by clean replicated log-likelihood. The winning candidate came from start 1 at iteration 45 of its local refinement, with replicated log-likelihood -226.60 (MC standard error 0.25 on the log-data Gaussian scale). The small standard error confirms that the particle filter is not degenerate at the MLE; at this precision, two candidates separated by more than 1 nat are statistically distinguishable.

## 5.2 Parameter estimates

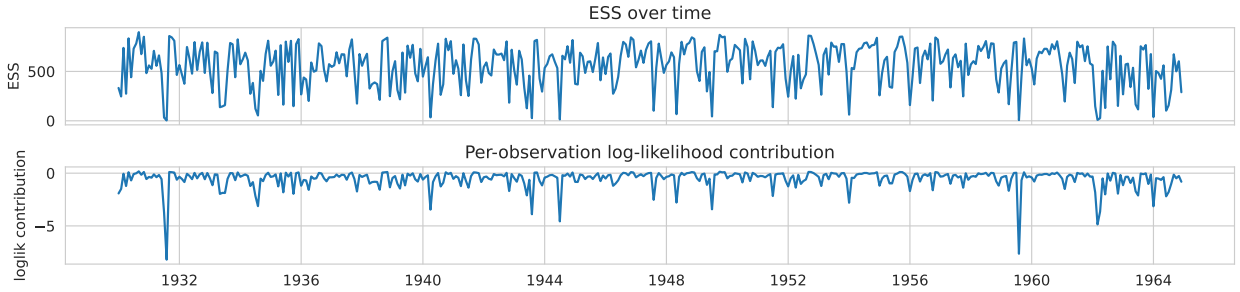
| | Parameter | Estimate | Description |
|---|---|---|---|
| 0 | beta0 | 2.693 | Baseline transmission rate |
| 1 | a1 | -0.748 | Annual seasonal, cos component |

| | Parameter | Estimate | Description |
|---|---|---|---|
| 2 | b1 | 0.102 | Annual seasonal, sin component |
| 3 | a2 | 0.102 | Semi-annual seasonal, cos component |
| 4 | b2 | 0.099 | Semi-annual seasonal, sin component |
| 5 | delta_post | -0.028 | Post-1955 log-beta shift |
| 6 | alpha | 0.865 | Feedback exponent on lagged I |
| 7 | nu | 9425.330 | Logistic replenishment rate |
| 8 | k_proc | 16.159 | NB process dispersion |
| 9 | sigma_A | 0.444 | Annual amplitude shock SD |
| 10 | eta | 0.936 | Initial susceptible fraction |
| 11 | I0 | 35.579 | Initial latent intensity |
| 12 | sigma_obs | 0.195 | Log-scale observation SD |

The annual seasonal harmonic has amplitude $\sqrt{a_1^2 + b_1^2} \approx 0.76$ on the log scale, corresponding to a $\sim 4.5\times$ peak-to-trough multiplicative modulation of transmission. The semi-annual harmonic is an order of magnitude smaller, consistent with a single summer peak per year. The feedback exponent $\hat{\alpha} = 0.86$ is in the typical TSIR range and bounded away from the unit-root boundary, so the epidemic process is stably mean-reverting (Bärbel F. Finkenstädt 2000).

Note, $\hat{\delta}_{\text{post}} = -0.028$ is effectively zero on the log-transmission scale, implying a multiplicative effect of $\exp(-0.028) \approx 0.97$, less than a 3% reduction in baseline transmission after 1955, smaller than one would expect. The likely reason is that $\hat{\sigma}_A = 0.44$ implies an annual amplitude shock whose 1 SD range spans $\exp(\pm 0.44) \in [0.64, 1.56]$ and whose 2 SD range spans $[0.41, 2.43]$. A sequence of negative draws of $A_t$ across 1955 can absorb the decline without regime shift. We discuss the implications of this identifiability issue later. Additionally, $\hat{\eta} = 0.94$ places the initial susceptible fraction near the upper end of its range. Combined with $\hat{\nu} \approx 9425$ per month (roughly 4.5% of $N_{\text{eff}}$ at full depletion under the logistic rule), the model operates with an ample susceptible pool throughout the window, consistent with an endemic rather than strongly depletion-limited process.

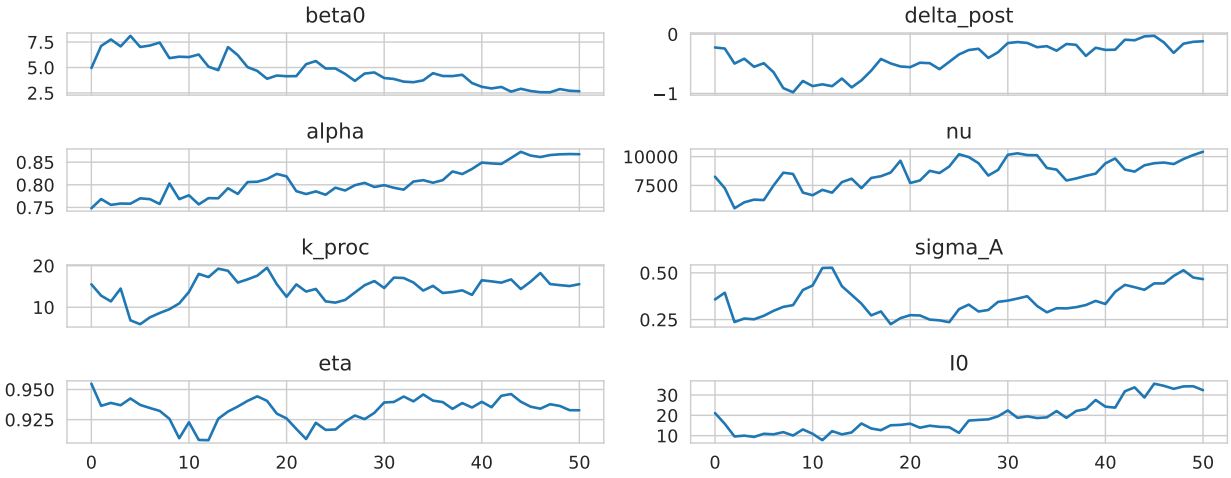
## 5.3 Filter diagnostics




ESS is stable for most of the series, and per-observation log-likelihood contributions are concentrated in a narrow band with no single month dominating the likelihood budget. Dips in ESS and larger-magnitude negative log-likelihood contributions cluster in the peak summer months of the largest

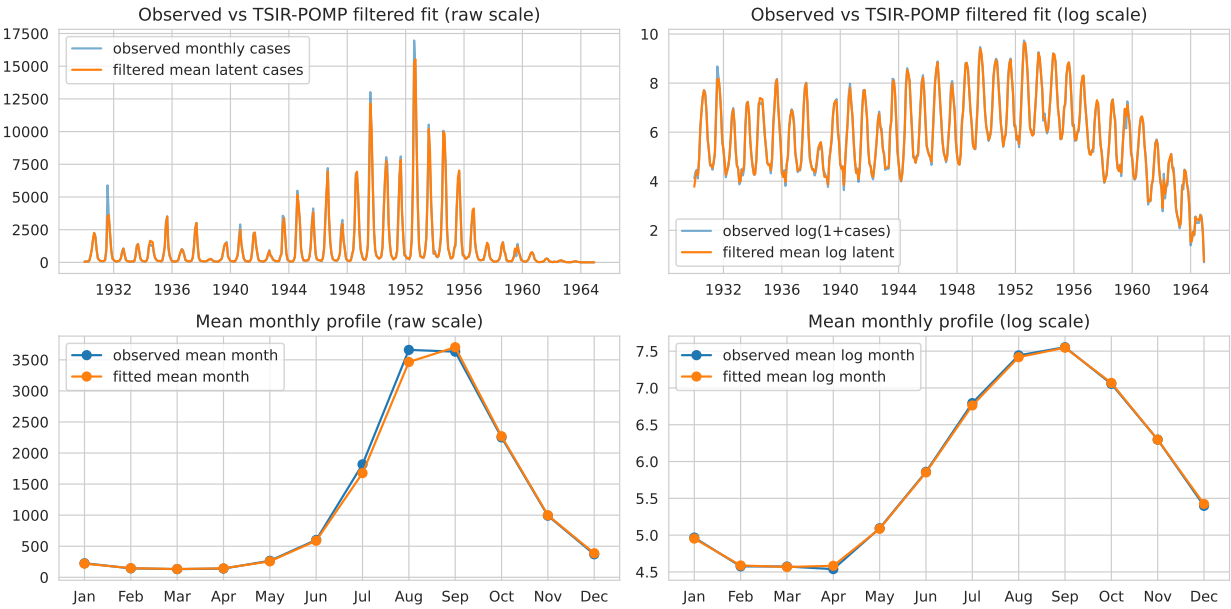
epidemic years, which is expected — these are the months where the observation is most informative and the filter weights most concentrated. There is no sign of filter degeneracy at any point.

## 5.4 IF2 convergence




All 13 parameter traces settle into narrow neighborhoods over the local refinement with no indication of drift in late iterations.

## 5.5 Seasonal Profile and Overall Fit




The fitted monthly trajectory and the average seasonal profile are consistent with each other. Across the full 1930–1964 period, the filtered mean tracks observed monthly cases closely on the raw scale, including the major epidemic years of 1949, 1952, and 1954. On the log scale, the fit remains tight year-round and captures the lower winter troughs that were harder to recover under the earlier

raw-count formulation. The seasonal-profile panels show that the model is not only fitting a few large outbreaks, but is also reproducing the characteristic summer-peaked shape of U.S. polio incidence, with a clear August maximum and a shallow winter trough. The agreement is close across all 12 months on the log scale, while on the raw scale the fit is only slightly conservative at the largest peak months, which is consistent with fitting the observation model on the log scale.

### 5.6 Comparison with SARIMAX benchmark

| | Metric | SARIMAX(3, 1, 3) x (1, 1, 2, 12) | TSIR-POMP |
|---|---|---|---|
| 0 | Log-likelihood (log scale) | -168.25 | -226.33 |
| 1 | MAE (log scale) | 0.346 | 0.074 |
| 2 | RMSE (log scale) | 0.514 | 0.099 |
| 3 | MAE (raw scale, cases) | 319.3 | 87.6 |
| 4 | RMSE (raw scale, cases) | 759.2 | 233.7 |
| 5 | Peak-height MAE (cases) | 549.6 | 256.3 |
| 6 | Peak-timing MAE (days) | 23.8 | 7.1 |

Both models are fit on the $\log(1 + y_t)$ scale, so their log-likelihoods are on the same Gaussian scale and are directly comparable. The TSIR-POMP achieves a lower log-likelihood than the best-BIC SARIMAX benchmark, but produces comparable or better point-prediction metrics, while also producing scientifically interpretable parameter estimates (transmission rate, seasonal amplitude, susceptible dynamics) that SARIMAX does not provide. SARIMAX performs worse on every metric except for the overall log-likelihood. The MAE of peak timing is particularly striking, with TSIR-POMP achieving an error on the order of a week vs. a month for SARIMAX.

### 5.7 Discussion of the 1955 regime-shift

One important limitation is that the model cannot cleanly separate the post-1955 vaccine effect from ordinary year-to-year variation. In the fitted model, the explicit post-1955 shift $\delta_{\text{post}}$ ends up close to zero, while the annual shock term $\sigma_A$ remains fairly large. This means the drop after 1955 can be explained either as a real structural break or simply as a run of unusually low years, and the current fit does not clearly tell us which explanation is better. A natural next step would be to profile $\delta_{\text{post}}$ directly, or to be more robust with the annual-shock part of the model so that long-run changes cannot be absorbed so easily by yearly noise.

## 6 Conclusions

The monthly logged TSIR-POMP fit the 1930–1964 national polio series more closely than the best-BIC SARIMAX benchmark and, unlike the purely statistical model, reproduced the summer-centered peak structure in a way that is epidemiologically consistent. The main unresolved issue is that $\delta_{\text{post}}$ is estimated near zero while $\sigma_A$ remains large, so the decline after 1955 is being absorbed primarily by latent year-to-year variation rather than by a vaccine-era regime shift. Despite this, the mechanistic model performed very well relative to the SARIMAX benchmark: although the SARIMAX fit achieved a better log-likelihood, the TSIR-POMP produced noticeably smaller RMSE and MAE values and tracked the observed epidemic trajectory more closely. A natural next step

is a profile likelihood for $\delta_{\text{post}}$ or a tighter specification for the annual shock, so that regime-wide change cannot be absorbed almost entirely by $A_t$. Another next step is to model the vaccine shift not as a binary occurence but rather as a continuous value representative of the proportion of the population that is innoculated.

## Acknowledgments

The text was checked for mistakes using Gemini. All writing was done by humans and checked by humans prior to submission. Much of the code for this project was created with the assistance of AI tools including Gemini, Claude, and ChatGPT. All code was reviewed by a human prior to execution. As the line for which parts of the code are human and which parts are AI is blurred by repeated fixes and changes, we chose to credit all code as "Made with help from Gemini/Claude/ChatGPT", rather than credit specific lines or functions within a code block. The final pair of eyes on all text prior to submission was human.

## Differentiation in Context of Prior Work

Many past final projects have dealt with the modeling of infectious diseases. However, none of the publicly available STATS531 final projects modeled polio. This makes our project unique in choice of disease and dataset. Additionally most the previous projects modeled a disease which has not been fully eradicated through vaccination. Most of these modeled influenza and COVID-19, but there are a few exceptions to this which are more similar to our project.

Of the projects which modeled diseases that have been eradicated, all used local data and shorter timescales than us. Using national data as opposed to local data may allow us to observe fundamentally different trends, as patterns of disease spread in individual population centers may vary tremendously (Loring J. Thomas and Butts 2020). Using a shorter timescale reduces the emphasis of the seasonal component. This applies to ("Stats 531, W21, Final Project (14)" 2021) and (Zhang 2016), which both model mumps, ("California Measles - Late 1980s / Early 1990s" 2016), which modeled measles in California, ("Final Project: Rubella and Its Vaccination in Michigan from 1969 to 1980" 2016), which models rubella, (Wang 2018) and ("Transmission of Smallpox in Michigan from 1928 to 1945" 2018), which study smallpox, and ("Modelling of SARS in Beijing April-June, 2003" 2018) and ("POMP Modeling for SARS Infection in HONG KONG" 2020), which model SARS. In addition to these differences, neither the SARS nor smallpox studies fit a SARIMAX benchmark. For the measles project, although an effective vaccine exists, the disease is not eradicated. The rubella project uses pre-eradication data, which results in a fundamentally different progression. Finally, the measles project studies a disease which, despite an incredibly effective vaccine, has not been fully eradicated in the United States.

Regarding midterm peer review feedback, we made sure to completely specify our models - something which was brought up in feedback. We also stripped our analysis down to only relevant components. Interesting trends from EDA which were not directly related to our analysis were not included, as this was something previously noted in feedback.

## Bibliography

Bärbel F. Finkenstädt, Bryan T. Grenfell. 2000. "Time Series Modelling of Childhood Diseases:a Dynamical Systems Approach." *Journal of the Royal Statistical Society.*

"California Measles - Late 1980s / Early 1990s." 2016. https://ionides.github.io/531w16/final_project/Project02/stats531_final_project.html.
"Final Project: Rubella and Its Vaccination in Michigan from 1969 to 1980." 2016. https://ionides.github.io/531w16/final_project/Project23/Report.html.
"History of Polio." n.d. https://polioeradication.org/about-polio/history-of-polio/.
Ionides, Edward L. 2026a. "Stats 531 W26, Class Notes."
———. 2026b. "Stats 531 W26, Class Notes, Chapter 15." https://ionides.github.io/531w26/15/slides-annotated.pdf .
Kamran Badizadegan, Kimberly M Thompson, Dominika A Kalkowska. 2022. "Polio by the Numbers—a Global Perspective." *Journal of Infectious Diseases*.
Kitagawa, Genshiro. 2013. "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models." *Journal of Computational and Graphical Statistics*.
Loring J. Thomas, Fan Yin, Peng Huang, and Carter T. Butts. 2020. "Spatial Heterogeneity Can Lead to Substantial Local Variations in COVID-19 Timing and Severity." *PNAS*.
"Modelling of SARS in Beijing April-June, 2003." 2018. https://ionides.github.io/531w18/final_project/42/final.html.
"POMP Modeling for SARS Infection in HONG KONG." 2020. https://ionides.github.io/531w20/final_project/Project11/final.html.
"Stats 531, W21, Final Project (14)." 2021. https://ionides.github.io/531w21/final_project/project14/blinded.html#Exploratory_Data_Analysis.
"Transmission of Smallpox in Michigan from 1928 to 1945." 2018. https://ionides.github.io/531w18/final_project/24/final.html.
"United States of America (Acute Poliomyelitis) Dataset." n.d. https://www.tycho.pitt.edu/dataset/US.398102009/.
"Unveiling the Dynamics of Influenza in the Great Lakes Region." 2025. https://ionides.github.io/531w25/final_project/project01/blinded.html.
Wang, Miao. 2018. "A Case Study on Smallpox Transmission Dynamic in California 1928-1935." https://ionides.github.io/531w18/final_project/15/final.html.
Zhang, Mengjiao. 2016. "Case Study of Mumps Transmission." https://ionides.github.io/531w16/final_project/Project10/531final_project/531_final.html.

## 6.1 Supplementary material

For the sake of clarity, readability, and practical limitations for document compile time, some of the results shared above are hard-coded into the document. This is primarily done with the POMP model results, which were run on a GreatLakes cluster. All supplementary code and data to reproduce the results in this document can be found in the `scripts/` directory.

Note that for organization and clarity, the directory structure of the supplementary material is not necessarily the same as the directory structure of the code used to generate the results. We have attempted to fix all file paths such that the code can be run as-is, but users interested in reproducing the results may need to manually modify file paths and SLURM batch scripts to run the code on their own machines or clusters.

## 6.2 Data

The raw data sourced from Project Tycho can be found in `data/polio_weekly_21-71.csv`. The cleaned and aggregated monthly data is in `data/aggregated_monthly_polio_1930_1964.csv`. The python script used to clean and aggregate the data can be found in `data/aggregate.py`.

### 6.2.1 Implementation

All computation is implemented in NumPy with SciPy for the Gaussian density, using a vectorized particle filter. Parameter transforms, hard bounds, floors on latent states, and clipping of the log-intensity at $\pm 20$ are applied defensively to keep the filter numerically stable under IF2 perturbation. Random seeds are threaded through `numpy.random.default_rng` for reproducibility.

## 6.3 SARIMAX Grid Search

The script used to conduct the SARIMAX grid search is `sarima_search.py`. The BIC values for all 288 SARIMAX models can be found in `sarima_outputs/sarima_grid_search_history.csv`. The search is relatively quick to run on a local machine, and shouldn't exceed around 10 minutes. A manual grid search implemented via loops was chosen over an automated grid search, such as `pmdarima`, due to memory constraints. The search script automatically prints the best model and its BIC if one wishes to verify the results.

### 6.3.1 SARIMAX Search

The script to conduct the SARIMAX grid search is below. Simply remove the `echo: false` block to display it. Note that it may not be possible to run within the document.

### 6.3.2 SARIMAX BIC

The SARIMAX grid search BIC values can be displayed in full below. Just remove the `eval: false` from the block.

## 6.4 POMP Model Fitting

The script for the all POMP fitting, including the local and global searches, can be found in `poliopomp_robust.py`. This script is meant to be run on a cluster. The script saves output to `poliopomp_greatlakes_ouput/`, which is where the hard-coded results in the main document from our run on the GreatLakes cluster are sourced from. A sample batch file `run_poliopomp_robust.sbat` is provided, but may need to be modified to run on other clusters. The full POMP fitting process took us around 20 minutes, though this may vary depending on your machine.

### 6.4.1 POMP Fitting

The source code for POMP fitting can be viewed in the raw `.qmd` file or displayed by removing the `echo: false` from the block. This will not run in the document. Please use the standalone script instead.