Modeling Seasonal Influenza: A Comparison of Temperature-Driven vs. Seasonal-Driven Transmission Models

Shuozishan Wang

December 3, 2025

Abstract

Seasonal influenza in temperate regions shows a strong winter peak, but the mechanisms behind this pattern are still debated. Using weekly influenza-like illness (ILI) counts from CDC ILINet and NOAA temperature records, I assembled a weekly dataset for Washington State covering 2010–2019 and evaluated how strongly influenza transmission is tied to environmental temperature. I compared three models: a temperature-driven regression model, a cosine-based seasonal model, and a mechanistic SIR model in which temperature directly modulates the transmission rate. The regression results show that temperature explains substantially more variation than a fixed annual cycle, with lower AIC and lower train/test RMSE. The SIR model reproduces the general seasonal rise and fall but is less flexible and achieves higher error. Taken together, the findings indicate that temperature variation plays a more important role than a purely calendar-driven pattern in shaping influenza seasonality in Washington State.

1 Introduction

Seasonal influenza in temperate regions shows a clear and highly repeatable winter pattern, but the reasons behind this seasonality are still debated. One explanation is that influenza transmission is sensitive to environmental conditions such as temperature or humidity, which may affect viral survival, host immune response, or the amount of time people spend indoors. Another view is that influenza follows an internal yearly rhythm driven by social and behavioral factors, school calendars, holidays, and population mixing patterns, regardless of weather.

In this project, I investigate which mechanism better explains influenza patterns in Washington State using weekly influenza-like illness (ILI) data from 2010–2019. I evaluate three models for the time-varying transmission rate $\beta(t)$. The first two are statistical models inspired by class material:

1. **Temperature-driven model**, where transmission increases as temperatures drop,
2. **Pure seasonal model**, based solely on a yearly cosine cycle.
3. **Mechanistic SIR model**, where temperature influences $\beta(t)$ inside a full SIR system of ODEs.

By comparing these three approaches, using AIC, RMSE, and visual inspection, the goal is to assess whether influenza seasonality in Washington is more strongly tied to environmental temperature changes or to a built-in calendar-based cycle.

2 Data Resources

This project uses two publicly available datasets covering influenza activity and temperature variation in Washington State from 2010–2019. After cleaning and aligning both sources, they were merged into a unified weekly time series suitable for model fitting.

2.1 CDC ILINet Influenza Data

Weekly influenza-like illness (ILI) counts were obtained from the CDC's ILINet surveillance system, which compiles outpatient reports of patients presenting with ILI symptoms. Weeks between October 2010 and September 2019 were retained. A weekly date index was reconstructed from the `YEAR` and `WEEK` fields using a Sunday-based convention.

2.2 NOAA Temperature Records

Daily minimum and maximum temperatures were taken from NOAA's GHCN dataset. A daily average was computed as

$$T_{\text{avg}} = \tfrac{1}{2}\left(T_{\max} + T_{\min}\right),$$

and the series was resampled to Sunday-ending weekly averages for alignment with ILINet. Forward-filling was necessary because the raw station data contained several gaps during the 2012 and 2015 winters, which would have otherwise caused misalignment with the weekly ILI indices.

2.3 Merged Weekly Dataset

The influenza and temperature datasets were merged by week, producing 469 aligned weekly observations. Two derived variables were constructed for modeling:

$$\text{Temp\_anom}(t) = T_{\text{ref}} - T(t), \qquad \cos(2\pi t/52),$$

representing temperature anomalies and a calendar-based seasonal indicator, respectively. These serve as the main predictors for the two transmission models evaluated in this study.

3 Methods

3.1 Data Preprocessing

To prepare the inputs for both models, the ILINet flu data were matched to the correct week using a Sunday-based timestamp. NOAA temperature observations, which were provided daily, were averaged within each week to align with flu records. After merging the two sources, weeks with missing entries were removed.

I created two derived variables for the model fitting. The first is the temperature anomaly,

$$\text{Temp\_anom}(t) = T_{\text{ref}} - T(t),$$

which increases during colder weeks. The second is a seasonal cosine term,

$$s(t) = \cos\left(\frac{2\pi t}{52}\right),$$

which represents an idealized annual cycle and does not use any environmental data.

3.2 Transmission Model Framework

Both models start from the standard SIR system introduced in class:

$$\frac{dS}{dt} = -\beta(t)SI, \qquad \frac{dI}{dt} = \beta(t)SI - \gamma I, \qquad \frac{dR}{dt} = \gamma I.$$

While the full SIR system describes the non-linear dynamics of infection, fitting the differential equations directly requires complex parameter estimation for $S(t)$ and $I(t)$ that is beyond the scope of this comparison. Therefore, this study adopts a simplified phenomenological approach. We assume that for the purpose of identifying environmental drivers, the observed weekly ILI count serves as a linear proxy for the effective transmission intensity $\beta(t)$. This allows us to use Ordinary Least Squares (OLS) to statistically compare the predictive power of temperature versus seasonal forcing, rather than simulating the full epidemic curve.

3.3 Model 1: Temperature–Driven Transmission

The first model tests the idea that influenza spreads more easily during colder periods. The transmission rate is written as

$$\beta(t) = \beta_0 + \alpha_T\big(T_{\text{ref}} - T(t)\big),$$

where β_0 is the baseline transmission level and α_T measures how strongly temperature anomalies affect transmission. This model was fitted using ordinary least squares (OLS) with the temperature–anomaly term as the predictor.

3.4 Model 2: Pure Seasonal Transmission

The second model assumes that influenza follows a built-in annual cycle, regardless of environmental temperature. In this case,

$$\beta(t) = \beta_0\left(1 + \delta\cos\left(\frac{2\pi t}{52}\right)\right),$$

where δ represents the strength of seasonal forcing. This model was also fitted using the OLS, using the cosine term as the explanatory variable.

3.5 Model 3: Temperature-Driven SIR ODE

To incorporate the course topics on differential equations, a mechanistic SIR model was also fit:

$$\beta(t) = \beta_0 + \alpha_T\big(T_{\text{ref}} - T(t)\big).$$

The SIR system was solved using `odeint` and the parameters $(\beta_0, \alpha_T, \gamma, I_0)$ were estimated by minimizing the RMSE between the simulated $I(t)$ and observed ILI counts.

3.6 Model Comparison

All models were trained on 2010–2017 and tested on 2017–2019. Performance was evaluated using

- AIC,
- train and test RMSE,
- visual comparison of fitted curves and observed data.

Lower AIC and RMSE indicate better agreement with the seasonal pattern of influenza activity.

4 Results

4.1 Exploratory Data Analysis

The weekly flu counts and temperature series show the expected opposite seasonal trends. Flu cases climb rapidly each winter, while temperatures drop to their lowest levels at roughly the same time in Figure 1. The close timing of these patterns suggests that colder weather might be linked to higher transmission.

The scatter plot in Figure 2 makes this even clearer. Almost all of the high ILI counts occur during weeks below about 45°F, while warmer weeks rarely show a noticeable spike. This gives an initial indication that temperature could be a meaningful predictor of influenza activity.

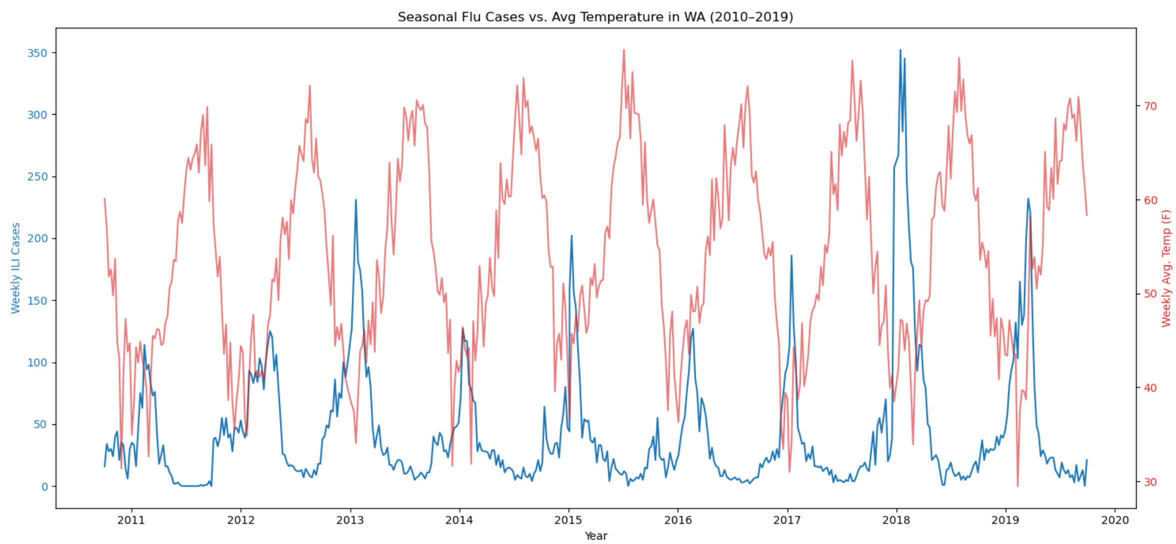

Figure 1: Weekly influenza-like illness (ILI) cases (blue) and weekly average temperature (red) in Washington State, 2010–2019.

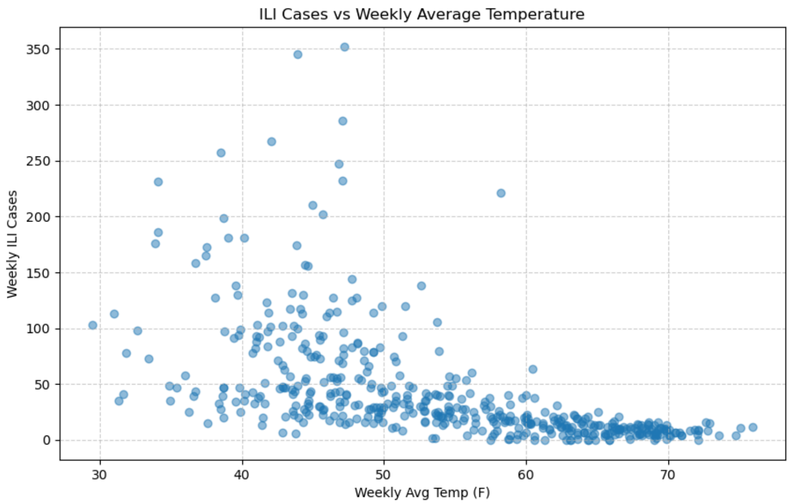

Figure 2: Scatter plot of weekly ILI cases versus weekly average temperature. Higher influenza activity occurs predominantly at lower temperatures.

4.2 Model Fitting Results

Model 1: Temperature-Driven Transmission. The temperature-based model

$$\widehat{\mathrm{ILI}}(t) = \beta_0 + \alpha_T \, \mathrm{Temp\_anom}(t)$$

produced a statistically significant estimate for α_T ($p < 10^{-15}$), indicating that colder weeks correspond to higher predicted flu activity. This model explained 39% of the variance ($R^2 = 0.391$) and achieved the lowest AIC among the tested models.

Model 2: Pure Seasonal Forcing. The seasonal cosine model

$$\widehat{\mathrm{ILI}}(t) = \beta_0 + \delta \cos(2\pi t/52)$$

captured the annual oscillation but explained only 11.6% of the variance ($R^2 = 0.116$). Its peak heights were consistently lower and smoother than observed ILI peaks, suggesting limited flexibility in representing real outbreak shapes.

4.3 Full Model Comparison Including SIR Dynamics

I fitted an SIR ODE model in which temperature modulates the transmission rate. Although the model reproduced the general rise and fall of each season, it was much less flexible than the linear models: the fitted curve remained well below the large winter peaks and produced a smoother trajectory than the observed data. This is reflected in its higher errors, with a training RMSE of 52.6 and a test RMSE of 88.8.

Table 1 summarizes all three models. The temperature-driven linear model clearly performed best across every metric. The cosine model provided a coarse seasonal pattern but lacked accuracy, and the mechanistic SIR model was too rigid to capture the sharp peak behavior seen in Washington's flu data.

Table 1: Performance comparison across the three fitted models. Lower values indicate better fit.

Model	Type	Train RMSE	Test RMSE	Description
Model 1	Linear regression	30.1	63.0	Temperature-driven
Model 2	Linear regression	36.3	68.8	Seasonal cosine
Model 3	SIR ODE model	52.6	88.8	Temp-driven transmission rate

4.4 Visual Comparison of Fitted Curves

Figure 3 overlays all three fitted curves with the observed ILI data. The temperature-linear model aligns most closely with peak timing and magnitude. The cosine model captures the annual rhythm but underestimates peak intensity. The SIR ODE model tracks only the broad shape of each season and fails to match sharp outbreaks.

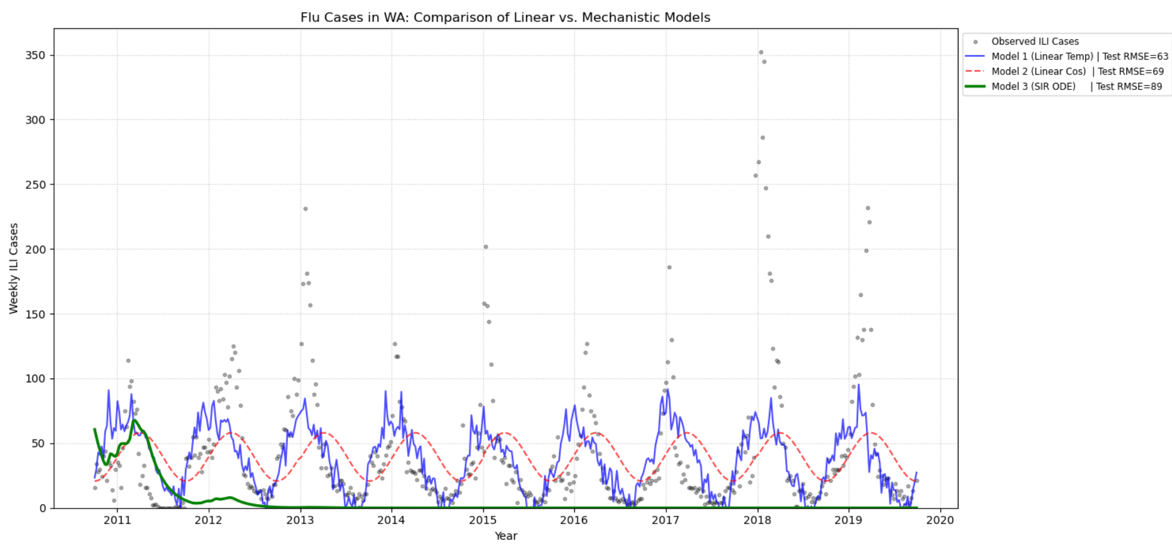

Figure 3: Observed weekly ILI counts compared with predictions from all models. Blue: temperature-driven linear model; red dashed: cosine seasonal model; green: SIR ODE model.

5 Discussion

The aim of this project was to understand whether the strong winter pattern of influenza in Washington State is better explained by changes in temperature or by an intrinsic yearly cycle. Looking across the exploratory figures and the performance of all three models, the temperature– based explanation consistently aligns more closely with the data.

The descriptive plots already suggest a strong link. The highest ILI counts occur almost exclusively during the coldest weeks of each year, and once temperatures rise above about 50°F, influenza activity drops sharply. The scatter plot reinforces this point: the large outbreaks are concentrated in a narrow band of cold temperatures, while warmer weeks produce almost no activity. This provides an early indication that temperature is at least strongly associated with transmission intensity.

The statistical results make the comparison clearer. Among the two linear regression models, the temperature-driven model performs markedly better, with a substantially lower AIC and smaller errors on both the training and test sets. Its fitted curve also follows the rise and decline of each season more closely. The cosine model captures only the broad "once-per-year" rhythm. It tends to smooth out the sharp winter peaks and never reaches the observed heights, suggesting that a fixed sinusoidal pattern is too rigid to capture the rapid changes that often occur when temperatures drop quickly.

Adding the temperature-driven SIR model helps connect these statistical observations with the differential equations studied in class. While the SIR model reproduces the general winter rise and summer decline, it is far less flexible than the regression models: the simulated infectious population is smoother, substantially underestimates the size of the winter peaks, and fails to match sudden mid-season surges. This is reflected in its higher RMSE values (52.6 train, 88.8 test). Part of this limitation is inherent to the basic SIR structure, which assumes fixed population compartments and cannot easily represent short-term changes due to behavior, school schedules, or the emergence of more transmissible viral strains.

Although the temperature-driven regression model performs best overall, it does not capture everything. The very large peaks in 2017–18 and 2018–19 remain underestimated. These seasons were unusually severe across the U.S. and were influenced by strain-specific factors and vaccine effectiveness—features not included in the present models. The residuals of both linear models also show strong autocorrelation, indicating that the week-to-week dynamics cannot be reduced to temperature alone. A more complete model would likely incorporate humidity, school calendars, mobility measures, or a richer mechanistic structure with explicit depletion of susceptibles.

Taken together, the results argue against the idea that influenza in Washington simply follows a fixed calendar cycle. Instead, temperature appears to play a measurable and meaningful role in determining outbreak intensity and timing. At the same time, the remaining unexplained variation highlights the limits of simplified models and points toward the need for more detailed transmission mechanisms in future work.

6 References

6.1 Appendices

1. Centers for Disease Control and Prevention (CDC). *ILINet: U.S. Outpatient Influenza-like Illness Surveillance Network*.

   Available at: `https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html`. Accessed 2025.

2. National Oceanic and Atmospheric Administration (NOAA). *Global Historical Climatology Network (GHCN) Daily Data*.

   Available at: `https://www.ncei.noaa.gov/cdo-web/search`. Accessed 2025.

3. Code Notebook Link

   Available at: `https://github.com/Doris040512/AMATH-383/blob/main/AMATH%20383%20Final%20Project%20(4).ipynb`

6.2 Citation

1. TUNG K. K. *Scaling Laws of Life, the Internet, and Social Networks.* Princeton University Press, 2007.

2. TUNG K. K *Modeling Change One Step at a Time*. In: *Princeton University Press. (2007)*