

# Forecasting Influenza Dynamics in the US: A Comparative Analysis Using the Nsub-Epidemic Model and the StatModPredict, A User-Friendly R-Shiny Interface with Statistical Models.

Mustapha Alhassan<sup>1</sup> and Clement Ampong<sup>1</sup>

<sup>1</sup>Georgia State University, Department of Mathematics and Statistics, Atlanta, GA 30303,  
United States

December 2024

## Introduction

Influenza remains a significant public health challenge in the United States, with its seasonal and regional variability necessitating robust forecasting systems. Accurate predictions of influenza trends are critical for enabling timely healthcare interventions, optimizing medical resources, and reducing the disease's societal impact [1, 2]. Reliable forecasts empower public health officials to respond proactively, minimizing morbidity and mortality associated with influenza outbreaks.

This study utilizes data collected through reported influenza cases and wastewater surveillance to evaluate predictive models' performance in forecasting influenza dynamics across the U.S. The dataset, spanning 26 weeks (May 4, 2024, to October 26, 2024), includes decimal-normalized case counts organized by five regions: National, Midwest, Northeast, South, and West. Wastewater surveillance provides an innovative early-warning tool, detecting viral activity within communities even before symptomatic cases are reported [2]. The viral activity levels from wastewater data are categorized based on infection risk, offering a detailed understanding of community transmission dynamics:

- **Minimal:** Up to 1.6
- **Low:** Greater than 1.6 and up to 4.5
- **Moderate:** Greater than 4.5 and up to 12.2
- **High:** Greater than 12.2 and up to 20.1
- **Very High:** Greater than 20.1

To evaluate forecasting accuracy, two complementary frameworks were employed: the *n-sub epidemic model* and the *STAT MOD PREDICT toolbox*. These frameworks assessed a range of models, including *ARIMA* (AutoRegressive Integrated Moving Average), *GLM* (Generalized Linear Models), *GAM* (Generalized Additive Models), *Prophet*, and various ranked and weighted ensemble approaches [3, 4]. Each model was analyzed for its ability to predict influenza case trajectories and infection risks accurately.

## Literature Review

Influenza forecasting is critical for managing public health responses and minimizing the disease's impact. Various statistical models, such as ARIMA, GLM, and GAM, have been widely employed for predicting influenza trends. ARIMA models are particularly effective for short-term forecasts due to their ability to capture temporal patterns. GLM and GAM models enhance flexibility by accommodating non-linear trends and additional covariates, such as vaccination rates and environmental factors [3].

Ensemble modeling approaches, such as Chowell's *n*-subepidemic framework, further improve forecasting accuracy by combining predictions from multiple models [4]. These methods are particularly effective in addressing the variability in regional epidemic patterns. For example, the SubEpiPredict toolbox has been specifically designed for this framework, allowing for the fitting and forecasting of epidemic growth trajectories using an ensemble of sub-epidemic models [7].

Frameworks like the STAT MOD PREDICT toolbox systematically evaluate the performance of multiple models, enabling the selection of the most accurate and reliable forecasting tools [3]. Additionally, the QuantDiffForecast toolbox has been used for parameter estimation and forecasting based on systems of ordinary differential equations, offering quantified uncertainty in its predictions [5]. The GrowthPredict toolbox also provides robust tools for forecasting growth trajectories in epidemics using phenomenological models, contributing significantly to forecasting capabilities [6].

## Research Questions

To investigate the effectiveness and reliability of various forecasting models, the study addresses the following research questions:

- What are the most accurate and reliable predictive models for forecasting influenza dynamics across different regions in the United States?
- How can predictive models support timely public health interventions and resource allocation to mitigate the impact of influenza outbreaks?

## Methodology

This study employs a systematic approach to forecast influenza dynamics in the United States, leveraging historical data and advanced modeling techniques. The methodology is structured as follows:

1. **Data Collection:** Historical influenza case data were derived from wastewater surveillance datasets provided by the Centers for Disease Control and Prevention (CDC) [2]. The dataset includes region-specific, decimal-normalized case counts spanning 26 weeks, from May 4, 2024, to October 26, 2024. The regions analyzed include National, Midwest, Northeast, South, East, and West, representing collections of states with varying demographic and environmental characteristics influencing influenza trends.

2. **Modeling Influenza Outbreaks:** The influenza data were modeled using two primary approaches:

- The *n*-sub epidemic model, which is specifically designed to capture the temporal progression of influenza outbreaks across multiple subpopulations [4].
- The StatModPredict toolbox, which offers a suite of statistical techniques for predicting time-series data [3].

These models were applied to the CDC influenza data to generate forecasts of case counts and assess their performance in capturing outbreak dynamics.

3. **Model Performance Evaluation:** To evaluate and compare the performance of the *n-sub epidemic model* and the StatModPredict toolbox, several error metrics and evaluation criteria were employed:

- Mean Squared Error (**MSE**): Measures the average squared difference between observed and predicted values, penalizing larger errors more heavily.
- Mean Absolute Error (**MAE**): Provides an average magnitude of prediction errors without considering their direction.
- 95% Prediction Interval (**95% PI**): Assesses the model's ability to capture the variability in the data within a specified confidence interval.
- Weighted Interval Score (**WIS**): Combines prediction interval coverage and sharpness to evaluate the accuracy and reliability of the forecasts.

4. **StatModPredict toolbox Forecasting and Model Specifications:**

We specify the models by selecting the parameters

- Calibration period length: 10 weeks
- Forecasting horizon length: 5 weeks
- Employed a 95% prediction interval for the models
- Used all four available models while keeping the default settings.

5. ***n-sub epidemic model* Model Fitting and Forecasting Specifications:**

We specify the models by selecting the parameters

- Error Distribution: Normal distribution
- Estimated method used: Nonlinear Least Square
- Used smoothing Factor of 1
- Used temporal resolution (DT=7), since we dealing with weekly data.
- Used 300 bootstrap realization to characterize parameter uncertainty
- Used 10 number of start points (number of initial guesses for parameter estimation procedure using MultiStart)
- Used all seven (7) available models.

By comparing these metrics, insights into the relative strengths and weaknesses of the two modeling approaches were derived, aiding in the identification of the most effective forecasting method.

## Forecasting Results

We generated a 5-week forecast for the weekly influenza cases in the USA from September 28, 2024, to October 26, 2024. The performance of the top-ranking models, ensemble models, ARIMA, GLM, GAM, and Prophet models were evaluated individually to assess their forecasting accuracy, reliability, and overall effectiveness.

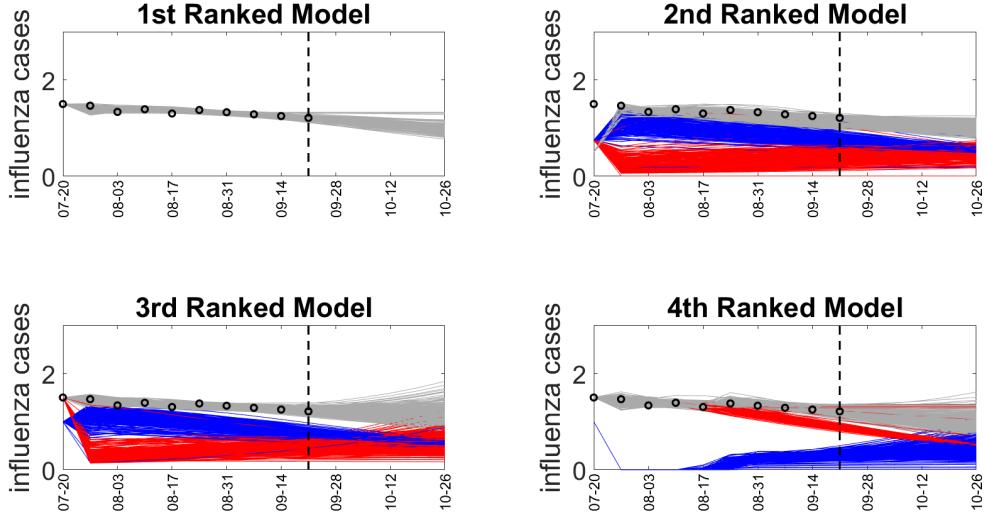
Note that the tables in this section provide a brief overview of the model performance, with a more comprehensive analysis discussed in the **discussion section**.

The colors indicated in the tables represent the performance of models based on key metrics:

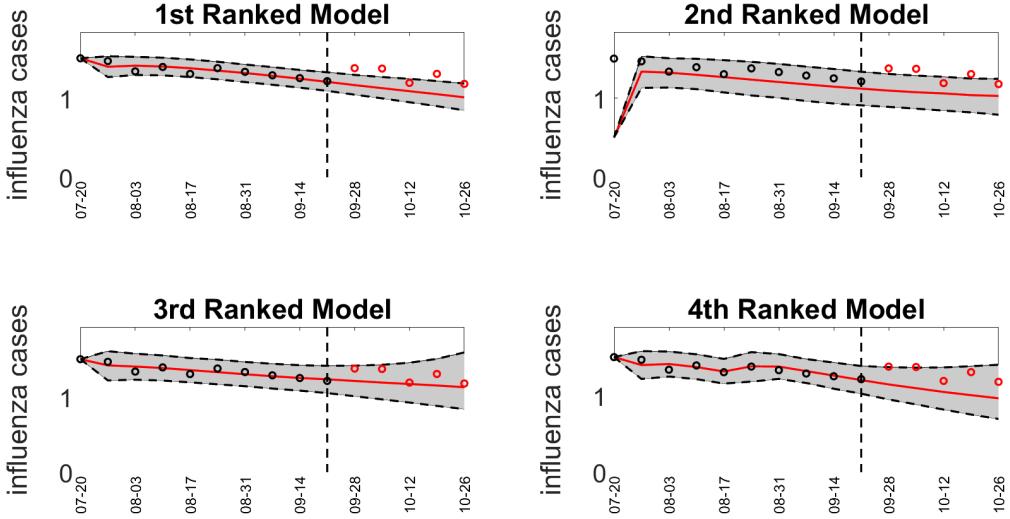
- **Blue:** Indicates the best-performing model with the highest accuracy and reliability.
- **Green:** Represents models that are close contenders, performing well but slightly below the best model.
- **Red:** Highlights the worst-performing models, characterized by high errors or low reliability.

The results are given below:

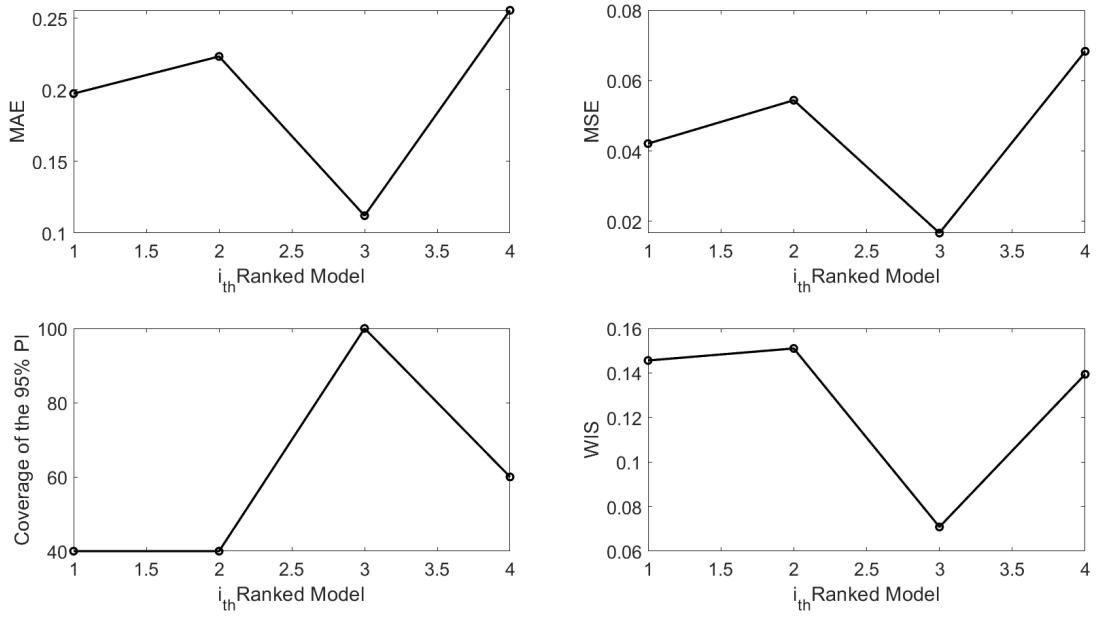
## Results for National



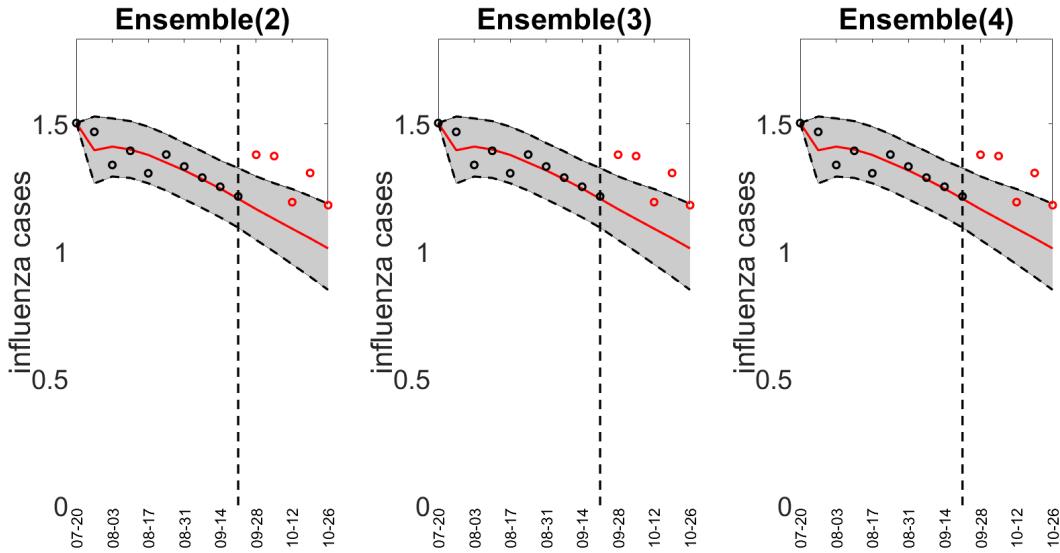
**Figure 1:** 5-week forecast sub-epidemic profiles derived from the top-ranking sub-epidemic models of the National weekly influenza cases. The blue and red curves represent different sub-epidemic waves generated with 300 bootstrap realizations, while the gray curves show the overall epidemic trajectory. The vertical line marks the start of the forecast, separating the calibration and forecast periods.



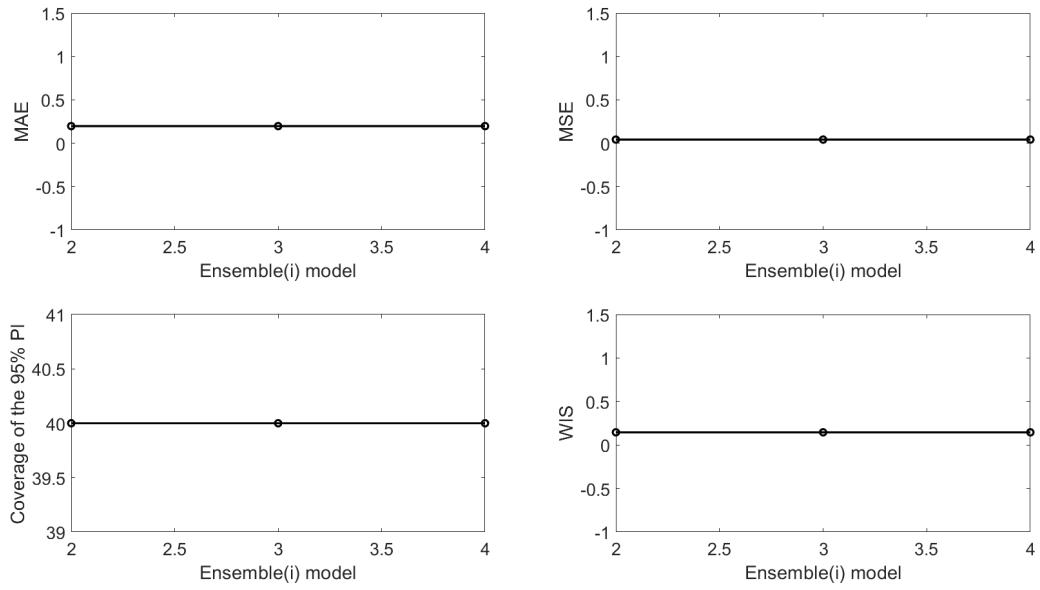
**Figure 2:** Shows 5-week forecasts from the top-ranking sub-epidemic models of the National weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. Circles indicate the observed data points.



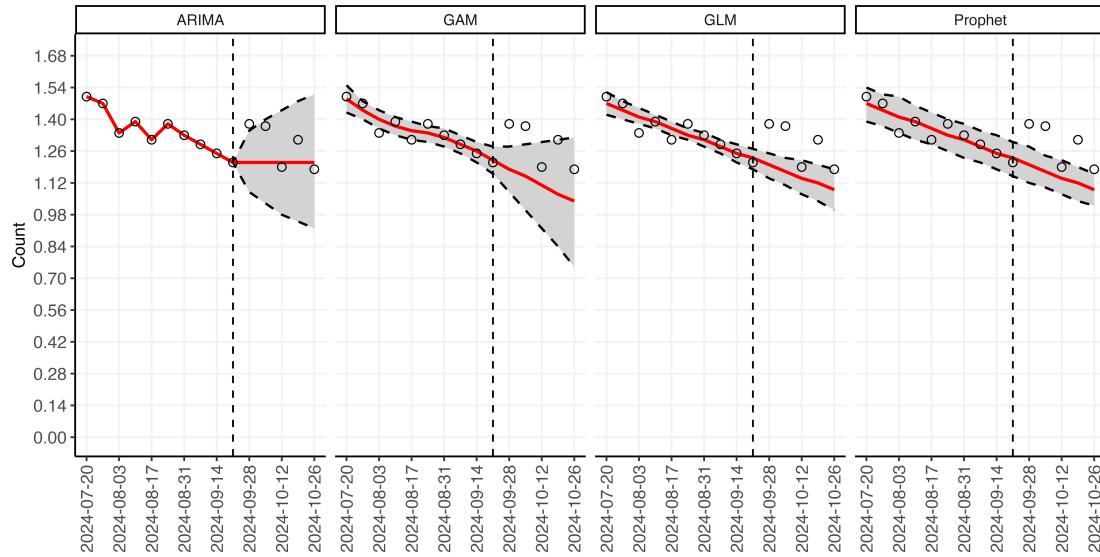
**Figure 3:** Performance metrics for 5-week forecasts based on the top-ranking sub-epidemic models for the National weekly influenza cases.



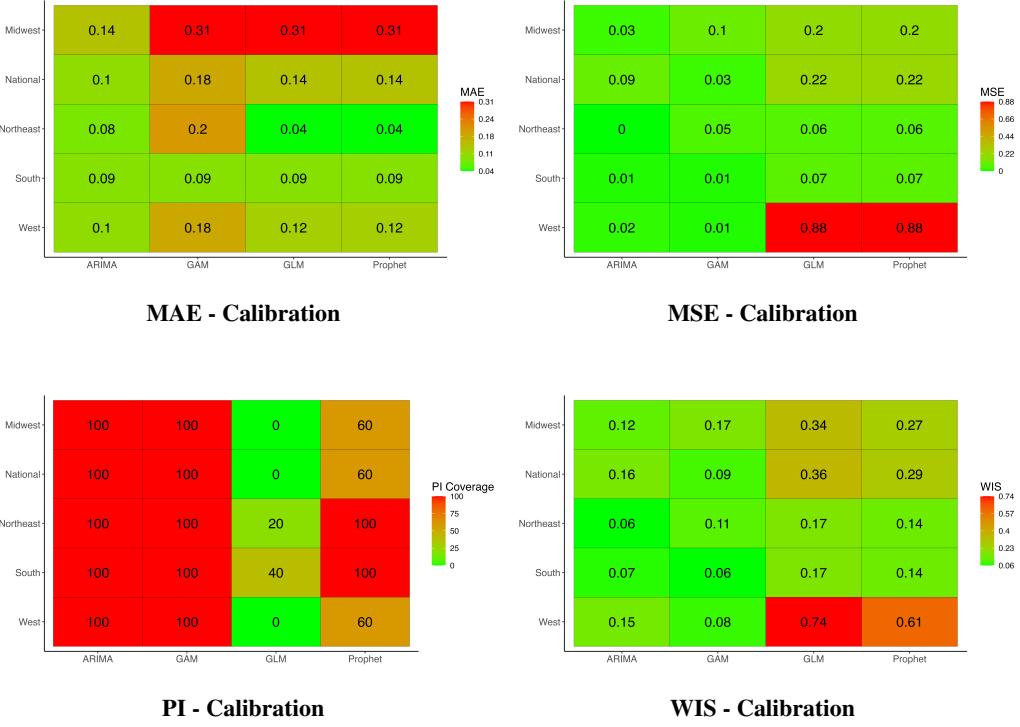
**Figure 4:** Displays 5-week sub-epidemic ensemble model forecasts for the National weekly influenza cases. The vertical line marks the start time of the forecast. The circles represent the data points, with model fits shown as solid lines and 95% prediction intervals indicated by shaded areas.



**Figure 5:** Performance metrics for 5-week forecasts based on the ensemble sub-epidemic models for the National weekly influenza cases.



**Figure 6:** Shows 5-week forecasts from the ARIMA, GLM, GAM AND THE PROPHET models of the National weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. The circles indicate the observed data points.

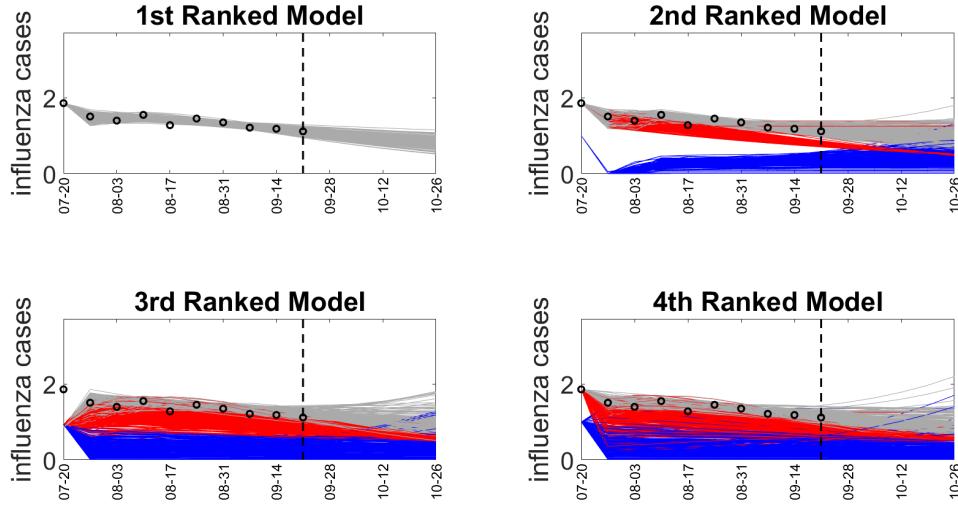


National				
Model	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.19	0.04	40	0.15
2 <sup>nd</sup> ranked model	0.22	0.05	40	0.15
3 <sup>rd</sup> ranked model	0.11	0.02	100	0.07
4 <sup>th</sup> ranked model	0.26	0.07	60	0.14
Weighted ensemble (2) model	0.19	0.04	40	0.15
Weighted ensemble (3) model	0.19	0.04	40	0.15
Weighted ensemble (4) model	0.19	0.04	40	0.15
ARIMA Model	0.10	0.01	80	0.06
GAM	0.18	0.03	60	0.12
GLM	0.14	0.02	40	0.12
Prophet	0.14	0.02	20	0.12

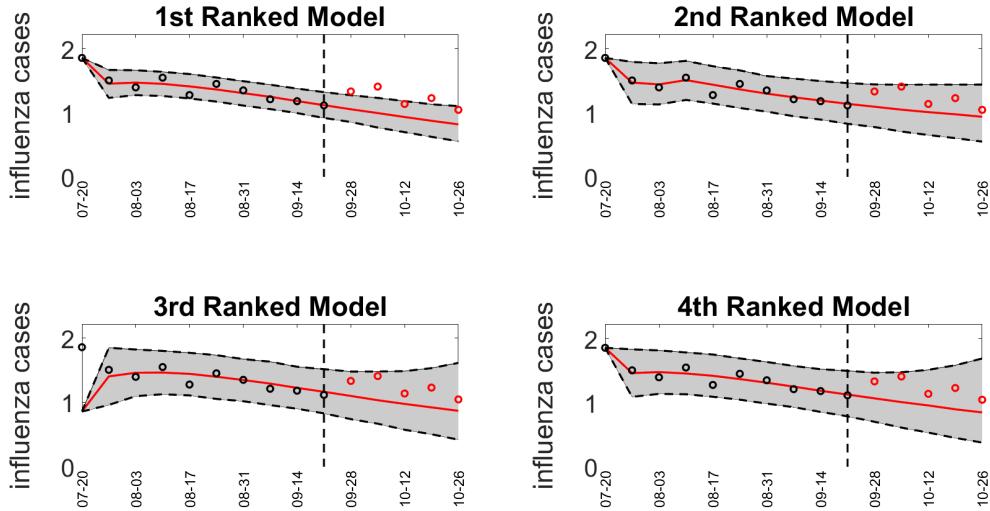
**Table 1:** 5-weeks forecasting performance metrics for National.

The figures in **Table 1** compare the performance metrics of various models for forecasting weekly influenza cases. The MAE is lowest for the ARIMA model (0.10), indicating the best accuracy, followed closely by the 3rd ranked model (0.11). Similarly, the ARIMA model also achieves the lowest MSE (0.01), reflecting superior precision. While the 3rd ranked model achieves 100% coverage of the 95% prediction interval, the ARIMA model shows a narrower but reliable 80% coverage. Lastly, the WIS score highlights the ARIMA model's superior calibration (0.06), slightly outperforming the 3rd ranked model (0.07).

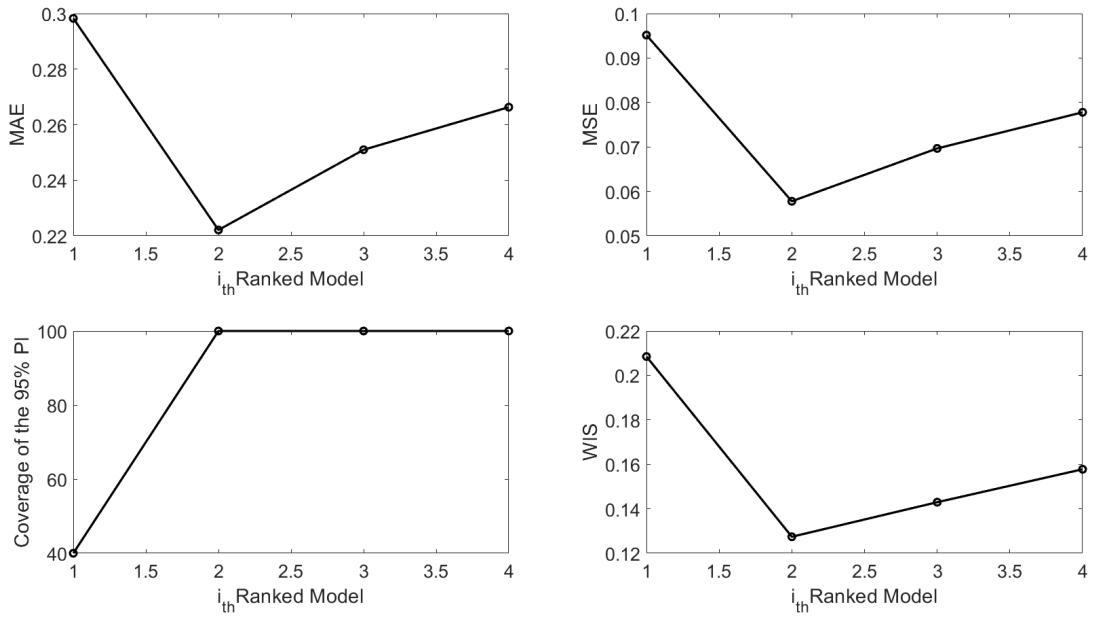
## Results for Midwest



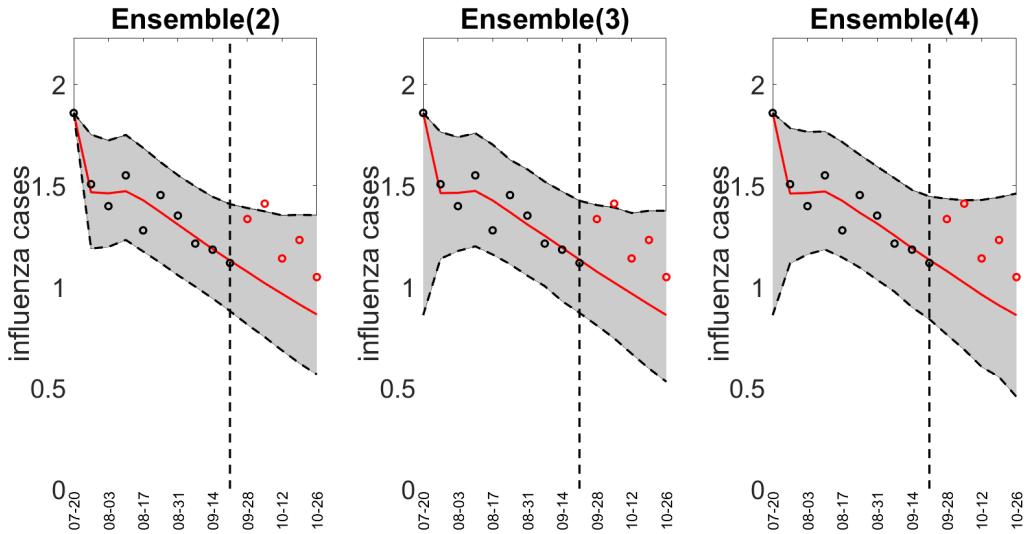
**Figure 7:** 5-week forecast sub-epidemic profiles derived from the top-ranking sub-epidemic models of the Midwest weekly influenza cases. The blue and red curves represent different sub-epidemic waves generated with 300 bootstrap realizations, while the gray curves show the overall epidemic trajectory. The vertical line marks the start of the forecast, separating the calibration and forecast periods.



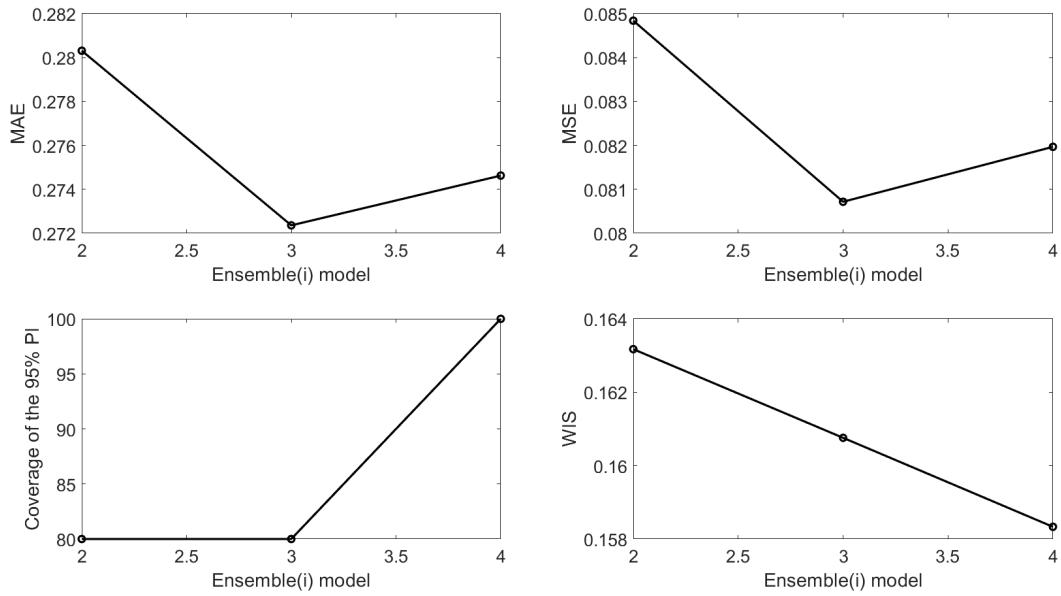
**Figure 8:** Shows 5-week forecasts from the top-ranking sub-epidemic models of the Midwest weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. Circles indicate the observed data points.



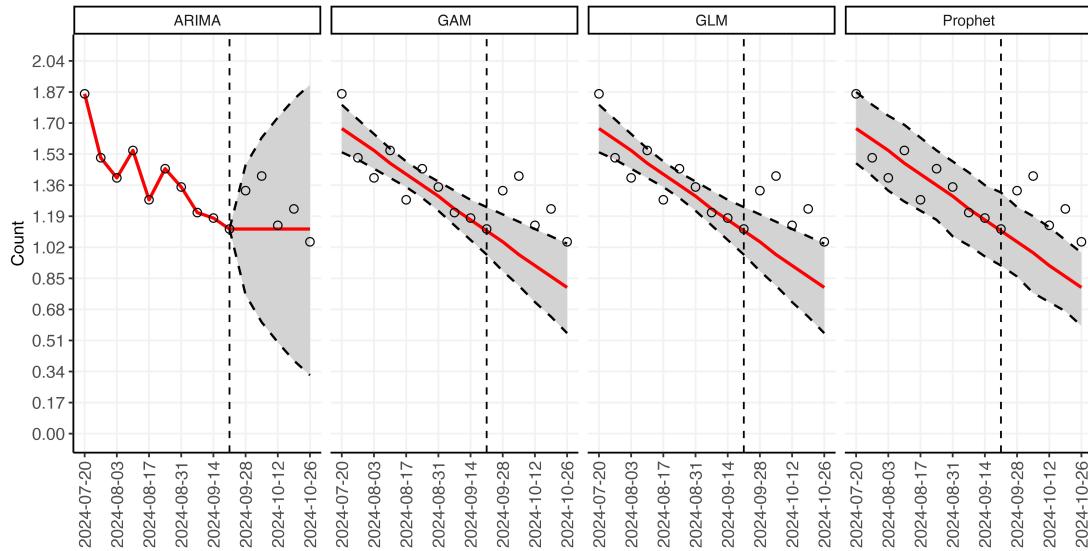
**Figure 9:** Performance metrics for 5-week forecasts based on the top-ranking sub-epidemic models for the Midwest weekly influenza cases.



**Figure 10:** Displays 5-week sub-epidemic ensemble model forecasts for the Midwest weekly influenza cases. The vertical line marks the start time of the forecast. The circles represent the data points, with model fits shown as solid lines and 95% prediction intervals indicated by shaded areas.



**Figure 11:** Performance metrics for 5-week forecasts based on the ensemble sub-epidemic models for the Midwest weekly influenza cases.



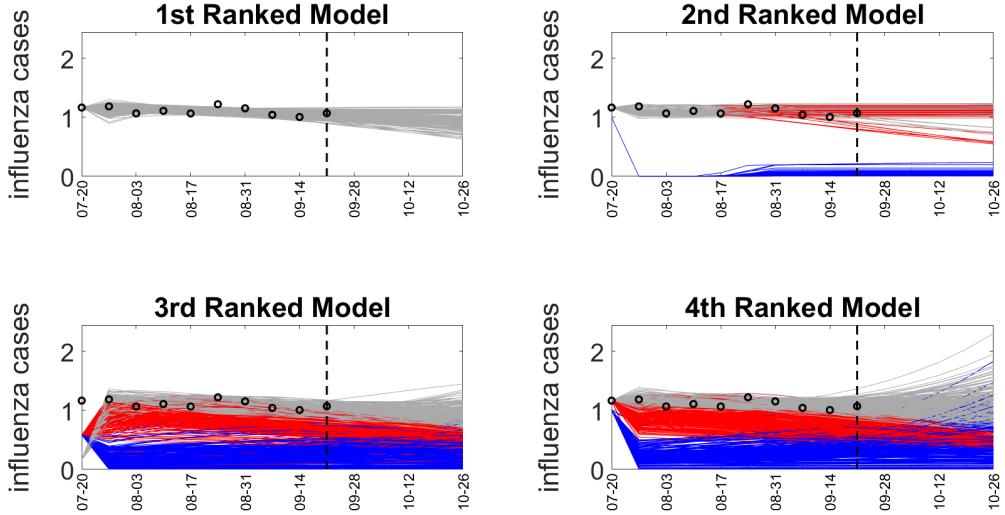
**Figure 12:** Shows 5-week forecasts from the ARIMA, GLM, GAM AND THE PROPHET models of the Midwest weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. The circles indicate the observed data points.

Midwest				
Model	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.29	0.09	40	0.22
2 <sup>nd</sup> ranked model	0.22	0.06	100	0.12
3 <sup>rd</sup> ranked model	0.24	0.06	100	0.14
4 <sup>th</sup> ranked model	0.28	0.09	100	0.17
Weighted ensemble (2) model	0.27	0.08	80	0.16
Weighted ensemble (3) model	0.28	0.09	100	0.16
Weighted ensemble (4) model	0.26	0.07	100	0.16
ARIMA Model	0.14	0.03	100	0.11
GAM	0.31	0.10	0	0.23
GLM	0.31	0.10	0	0.23
Prophet	0.31	0.10	0	0.23

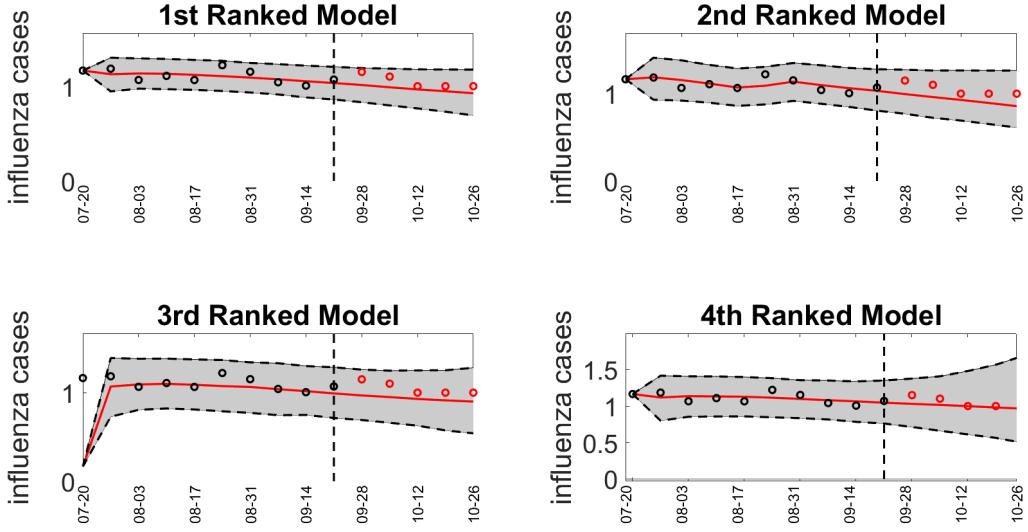
**Table 2:** 5-weeks forecasting performance metrics for Midwest

The figures in **Table 2** compare the performance metrics of various models for forecasting weekly influenza cases in the Midwest. The ARIMA model achieves the best accuracy with the lowest MAE (0.14) and MSE (0.03), reflecting superior precision. Both the ARIMA model and the 2nd ranked model achieve 100% coverage of the 95% prediction interval, demonstrating strong reliability, while other models show lower or no coverage. Finally, the ARIMA model has the lowest WIS (0.11), indicating superior calibration, closely followed by the 2nd ranked model (0.12).

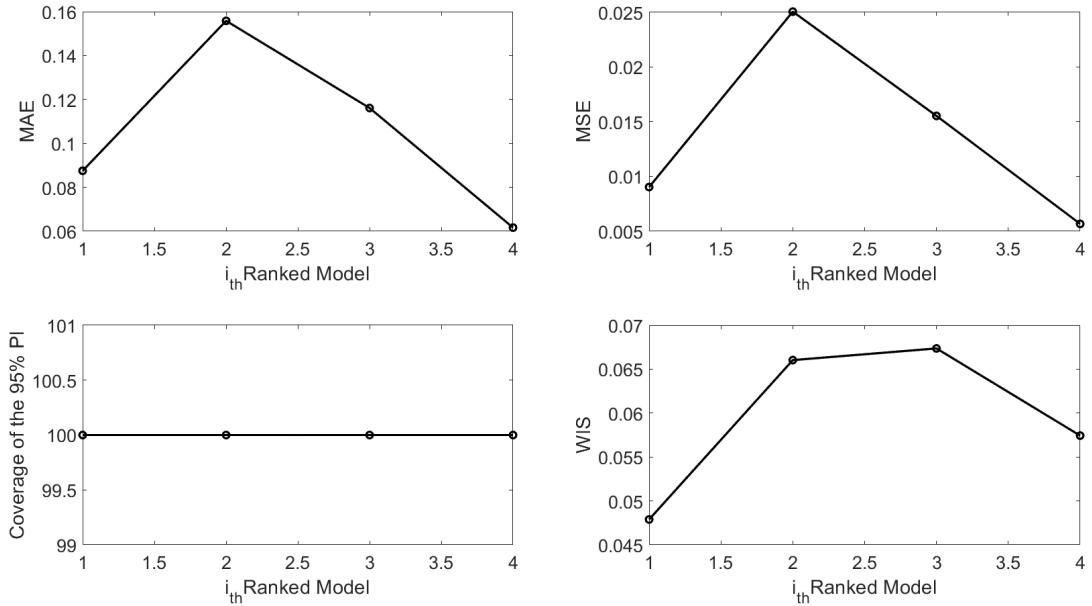
## Results for Northeast



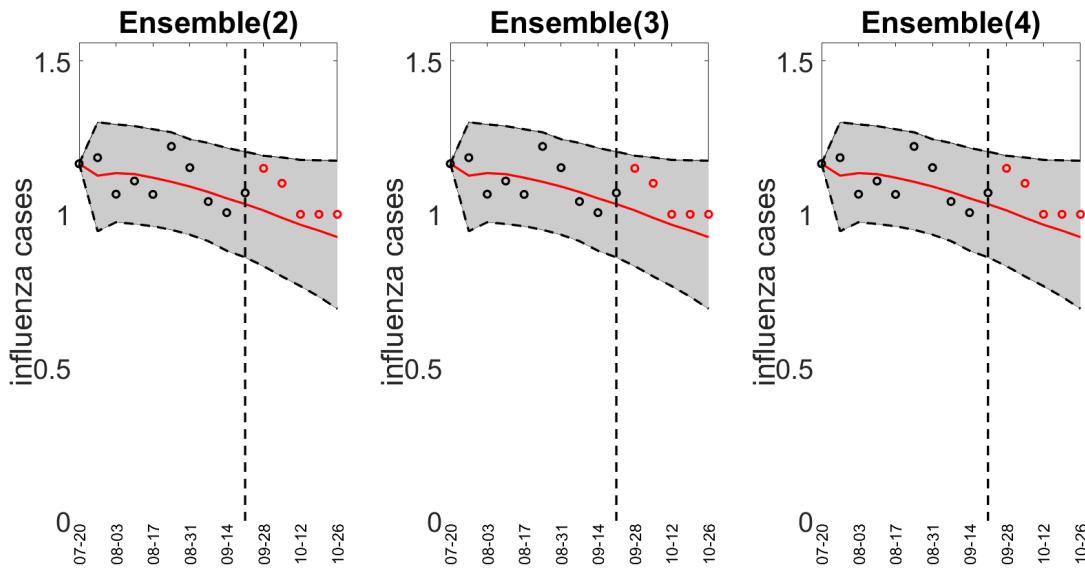
**Figure 13:** 5-week forecast sub-epidemic profiles derived from the top-ranking sub-epidemic models of the Northeast weekly influenza cases. The blue and red curves represent different sub-epidemic waves generated with 300 bootstrap realizations, while the gray curves show the overall epidemic trajectory. The vertical line marks the start of the forecast, separating the calibration and forecast periods.



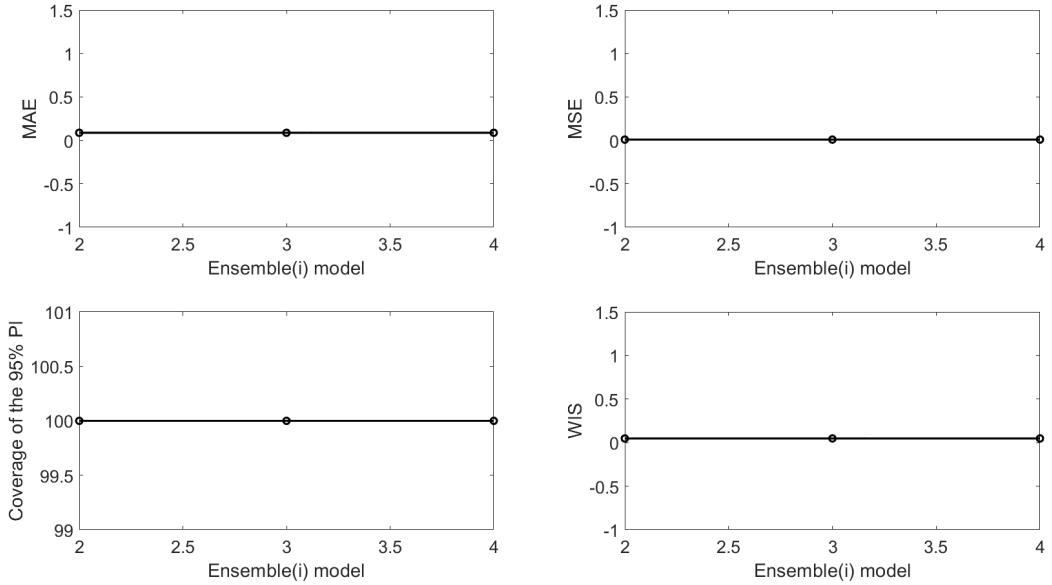
**Figure 14:** Shows 5-week forecasts from the top-ranking sub-epidemic models of the Northeast weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. Circles indicate the observed data points.



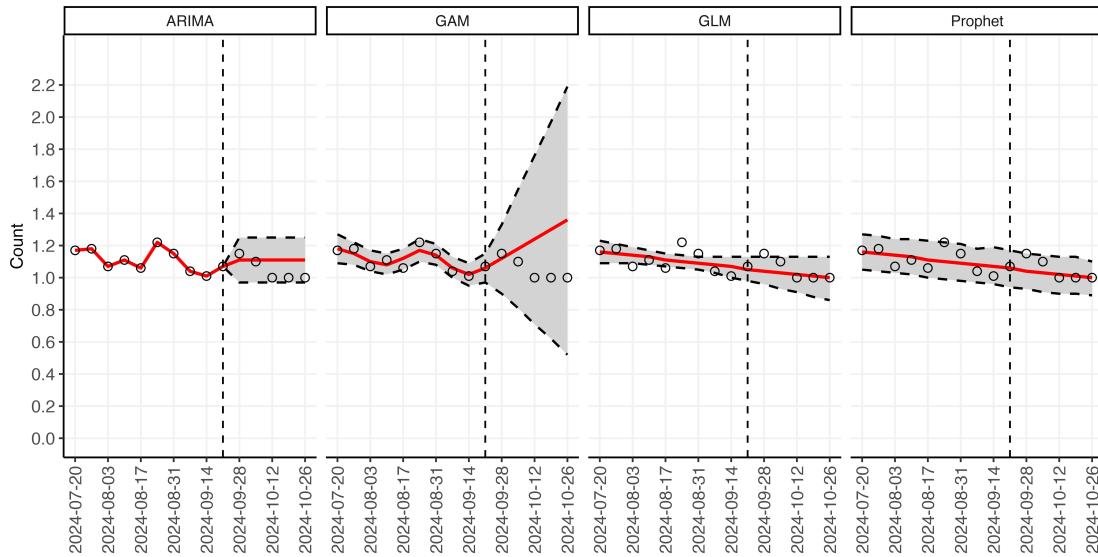
**Figure 15:** Performance metrics for 5-week forecasts based on the top-ranking sub-epidemic models for the Northeast weekly influenza cases.



**Figure 16:** Displays 5-week sub-epidemic ensemble model forecasts for the Northeast weekly influenza cases. The vertical line marks the start time of the forecast. The circles represent the data points, with model fits shown as solid lines and 95% prediction intervals indicated by shaded areas.



**Figure 17:** Performance metrics for 5-week forecasts based on the ensemble sub-epidemic models for the Northeast weekly influenza cases.



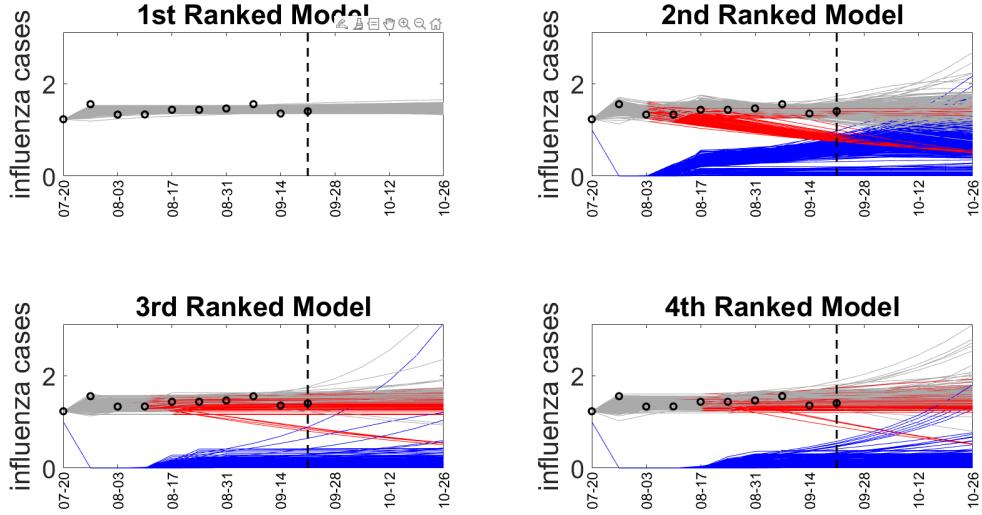
**Figure 18:** Shows 5-week forecasts from the ARIMA, GLM, GAM AND THE PROPHET models of the Northeast weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. The circles indicate the observed data points.

Northeast				
Model	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.09	0.01	100	0.05
2 <sup>nd</sup> ranked model	0.16	0.03	100	0.07
3 <sup>rd</sup> ranked model	0.11	0.02	100	0.07
4 <sup>th</sup> ranked model	0.06	0.01	100	0.06
Weighted ensemble (2) model	0.09	0.01	100	0.05
Weighted ensemble (3) model	0.09	0.01	100	0.05
Weighted ensemble (4) model	0.09	0.01	100	0.05
ARIMA Model	0.08	0.01	100	0.05
GAM	0.20	0.06	100	0.11
GLM	0.04	0	80	0.03
Prophet	0.04	0	100	0.03

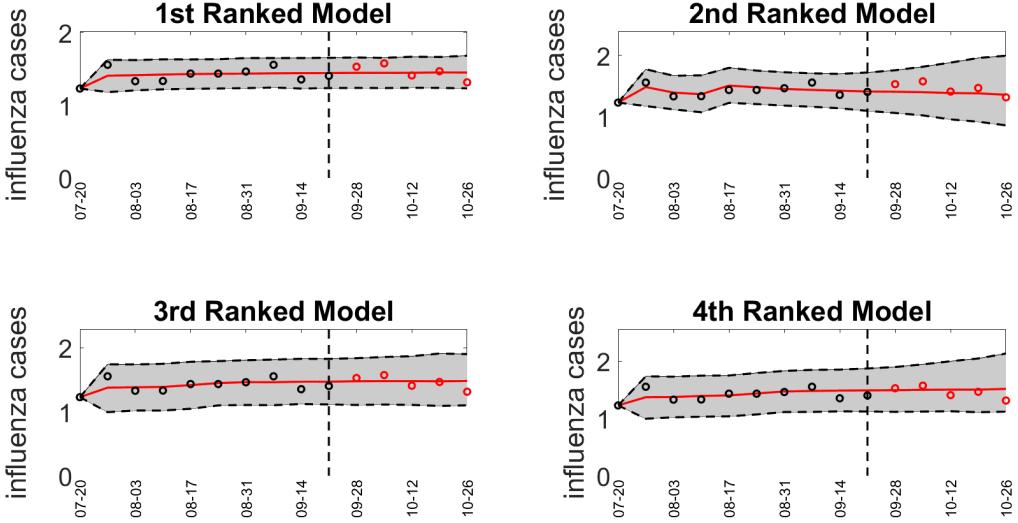
**Table 3:** 5-weeks forecasting performance metrics for Northeast

The figures in **Table 3** compare the performance metrics of various models for forecasting weekly influenza cases in the Northeast. The ARIMA model demonstrates the best overall performance with the lowest MAE (0.08) and MSE (0.01), reflecting exceptional accuracy and precision. It achieves 100% coverage of the 95% prediction interval, matching the reliability of other top-performing models like the 1st ranked model and weighted ensemble models. Additionally, the ARIMA model achieves one of the lowest WIS scores (0.05), indicating excellent calibration, comparable to the 1st ranked model and the weighted ensembles.

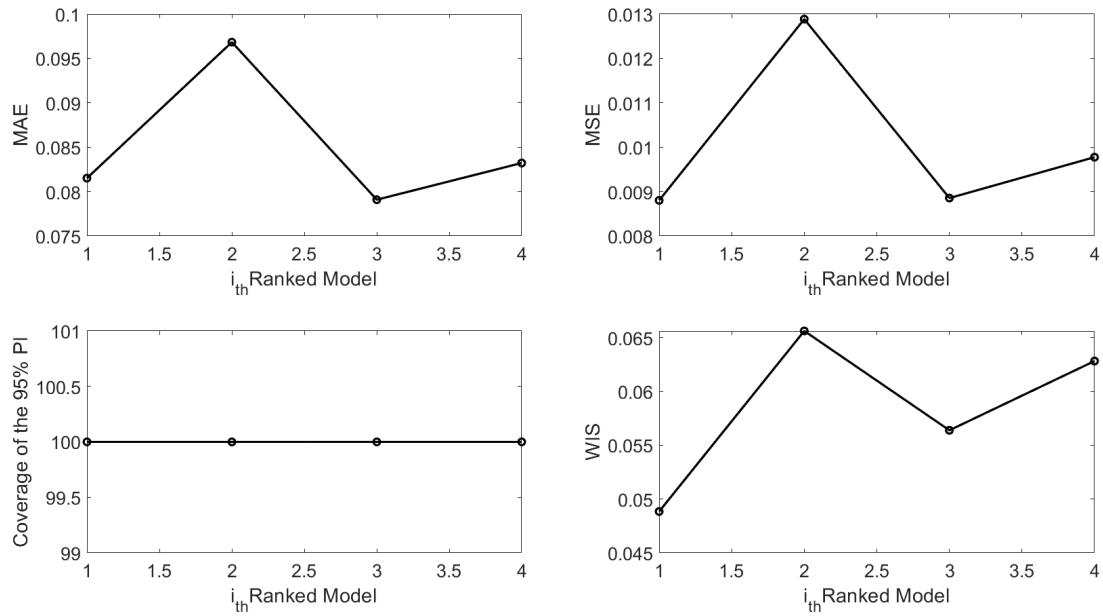
## Results for South



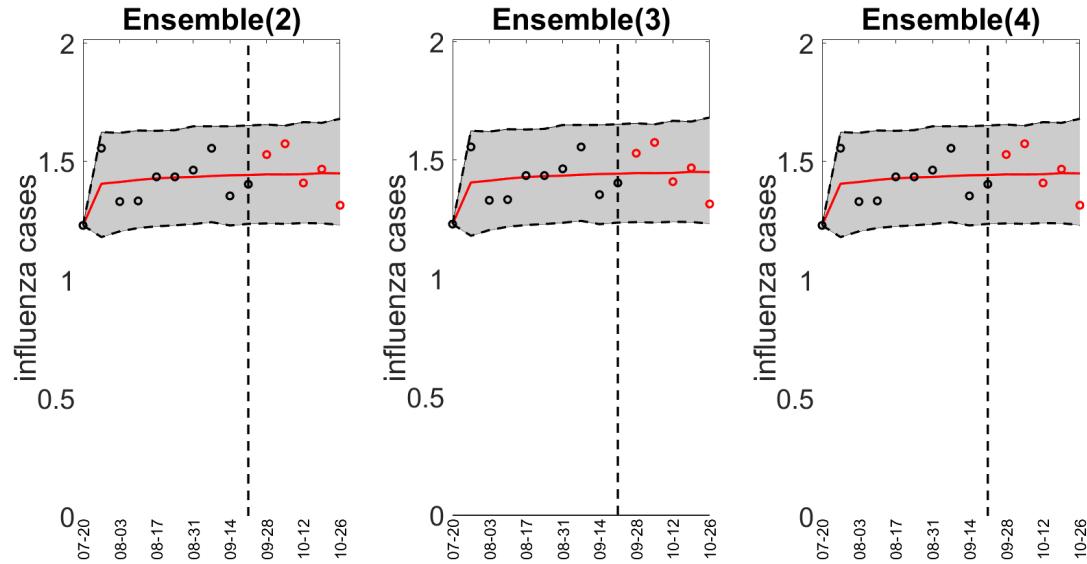
**Figure 19:** 5-week forecast sub-epidemic profiles derived from the top-ranking sub-epidemic models of the South weekly influenza cases. The blue and red curves represent different sub-epidemic waves generated with 300 bootstrap realizations, while the gray curves show the overall epidemic trajectory. The vertical line marks the start of the forecast, separating the calibration and forecast periods.



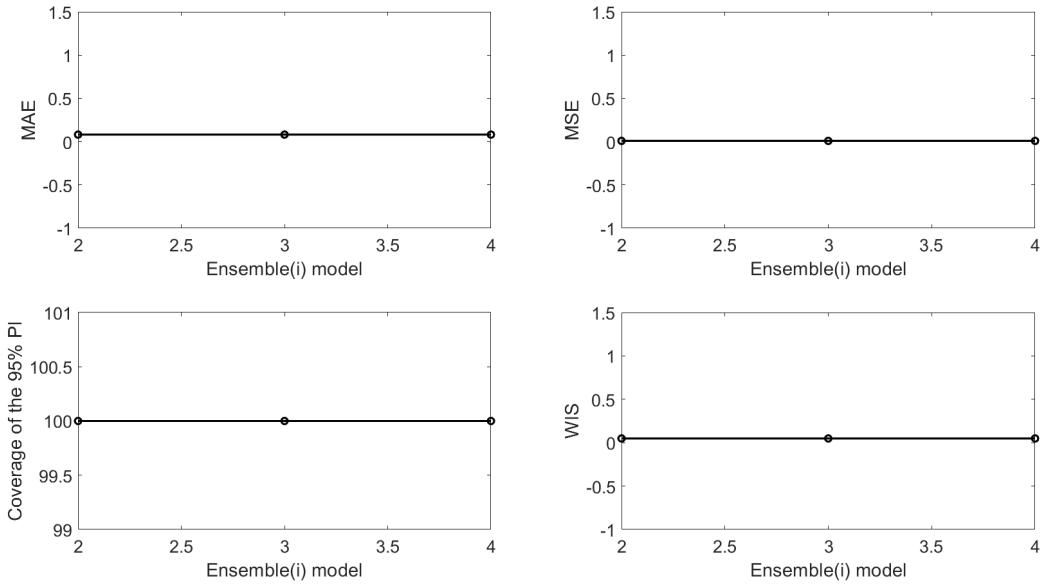
**Figure 20:** Shows 5-week forecasts from the top-ranking sub-epidemic models of the South weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. Circles indicate the observed data points.



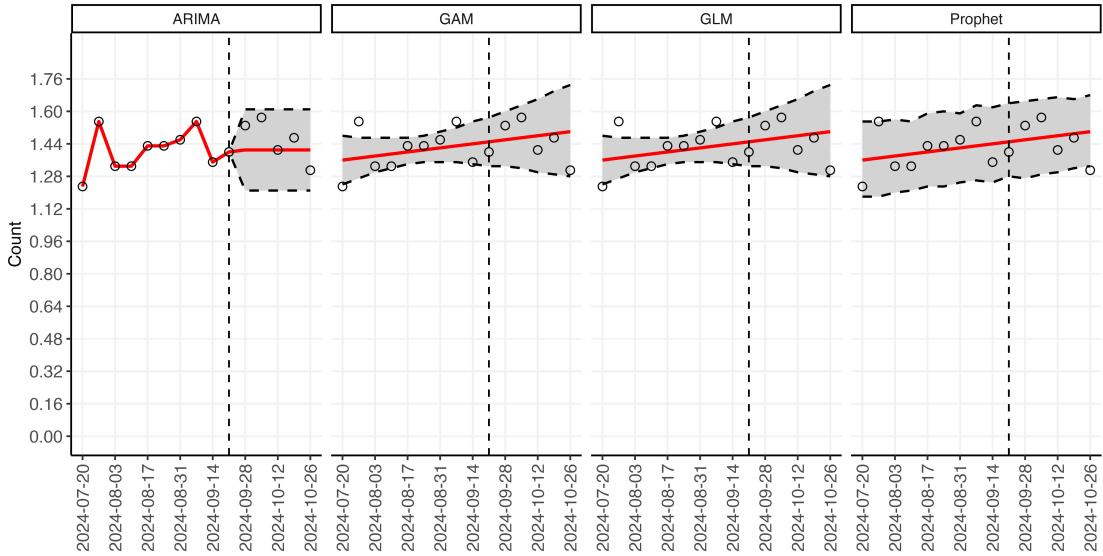
**Figure 21:** Performance metrics for 5-week forecasts based on the top-ranking sub-epidemic models for the South weekly influenza cases.



**Figure 22:** Displays 5-week sub-epidemic ensemble model forecasts for the South weekly influenza cases. The vertical line marks the start time of the forecast. The circles represent the data points, with model fits shown as solid lines and 95% prediction intervals indicated by shaded areas.



**Figure 23:** Performance metrics for 5-week forecasts based on the ensemble sub-epidemic models for the South weekly influenza cases.



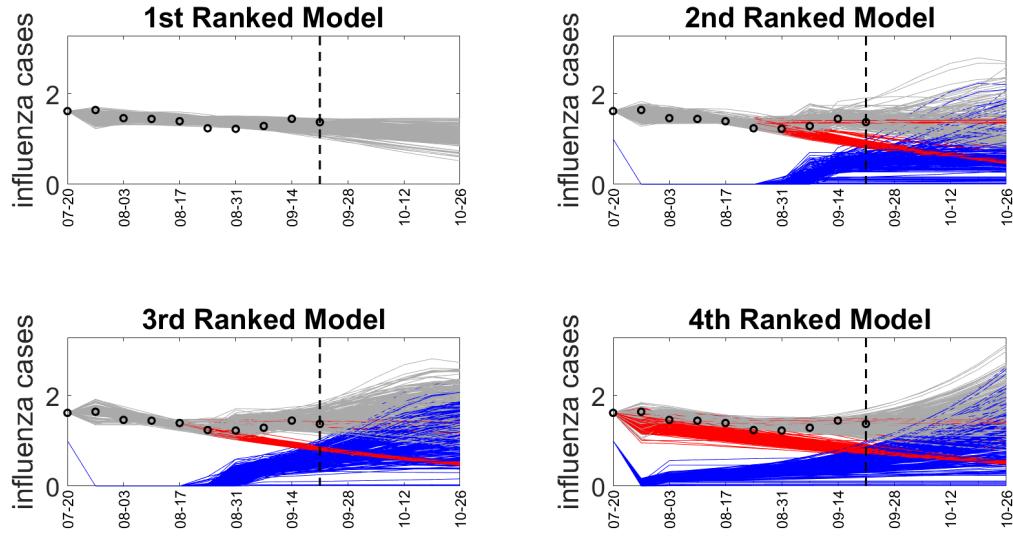
**Figure 24:** Shows 5-week forecasts from the ARIMA, GLM, GAM AND THE PROPHET models of the South weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. The circles indicate the observed data points.

South				
Model	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.08	0.01	100	0.05
2 <sup>nd</sup> ranked model	0.09	0.01	100	0.07
3 <sup>rd</sup> ranked model	0.08	0.01	100	0.06
4 <sup>th</sup> ranked model	0.08	0.01	100	0.06
Weighted ensemble (2) model	0.08	0.01	100	0.05
Weighted ensemble (3) model	0.08	0.01	100	0.05
Weighted ensemble (4) model	0.08	0.01	100	0.05
ARIMA Model	0.09	0.01	100	0.05
GAM	0.18	0.01	100	0.05
GLM	0.12	0.01	100	0.05
Prophet	0.12	0.01	80	0.06

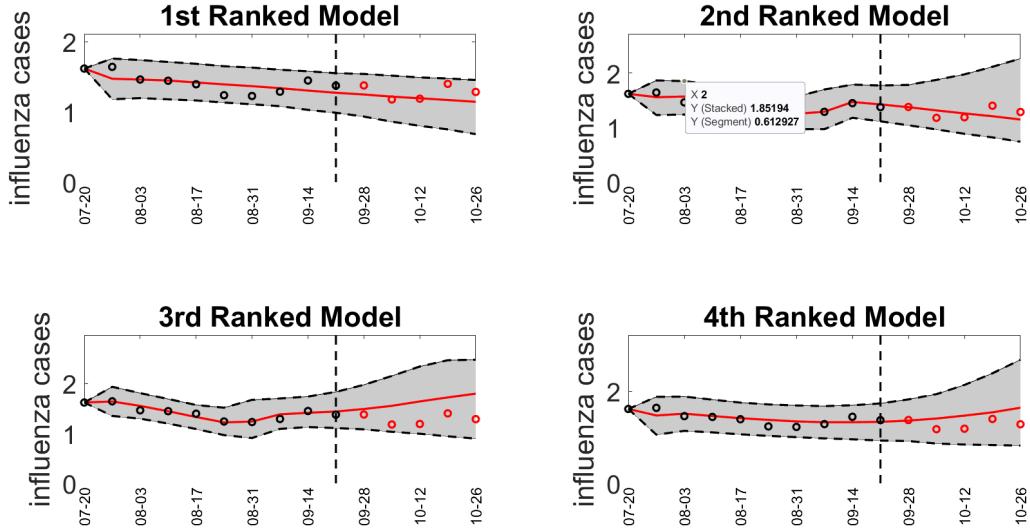
**Table 4:** 5-weeks forecasting performance metrics for South

The figures in **Table 4** compare the performance metrics of various models for forecasting weekly influenza cases in the South. The 1st ranked model and the weighted ensemble models exhibit the best overall performance with the lowest MAE (0.08), MSE (0.01), and WIS (0.05), reflecting exceptional accuracy, precision, and calibration. These models also achieve 100% coverage of the 95% prediction interval, demonstrating strong reliability. The ARIMA model closely follows, with a slightly higher MAE (0.09) but identical MSE, 95% PI coverage, and WIS, maintaining competitive performance.

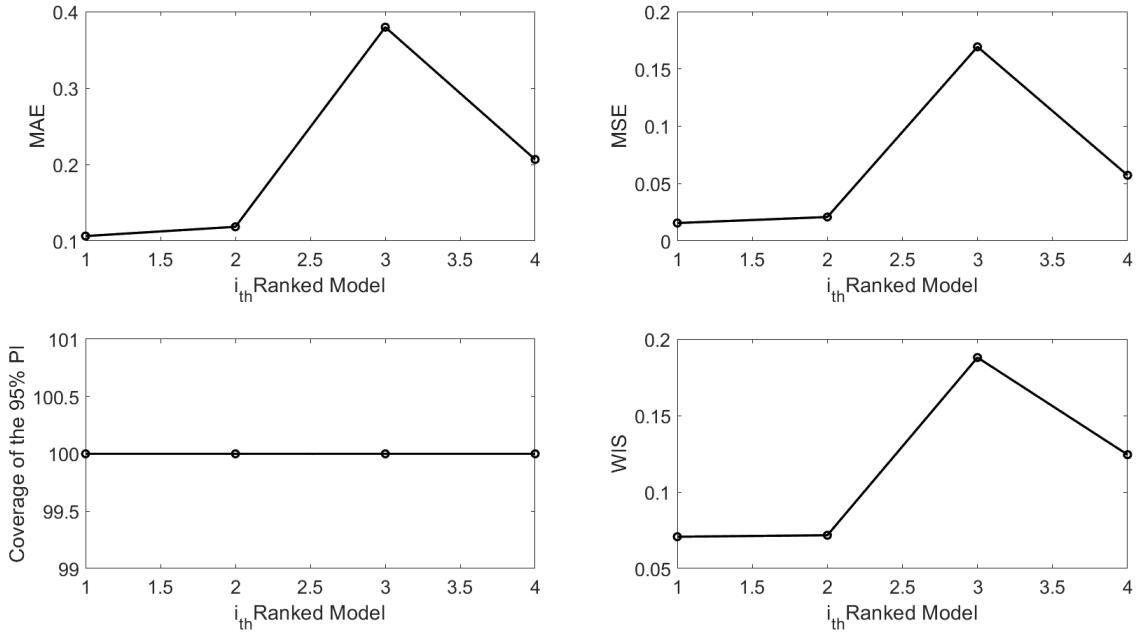
## Results for West



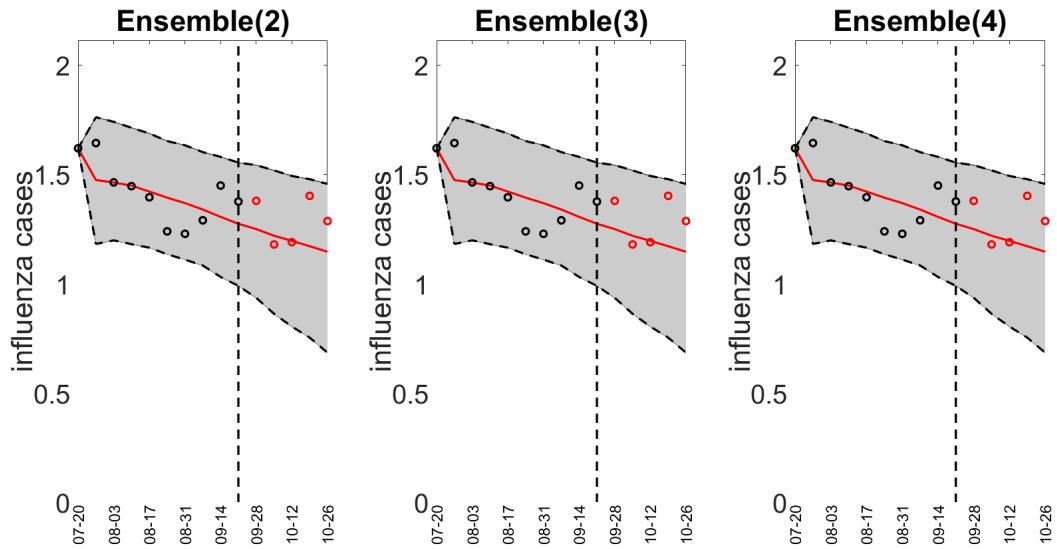
**Figure 25:** 5-week forecast sub-epidemic profiles derived from the top-ranking sub-epidemic models of the West weekly influenza cases. The blue and red curves represent different sub-epidemic waves generated with 300 bootstrap realizations, while the gray curves show the overall epidemic trajectory. The vertical line marks the start of the forecast, separating the calibration and forecast periods.



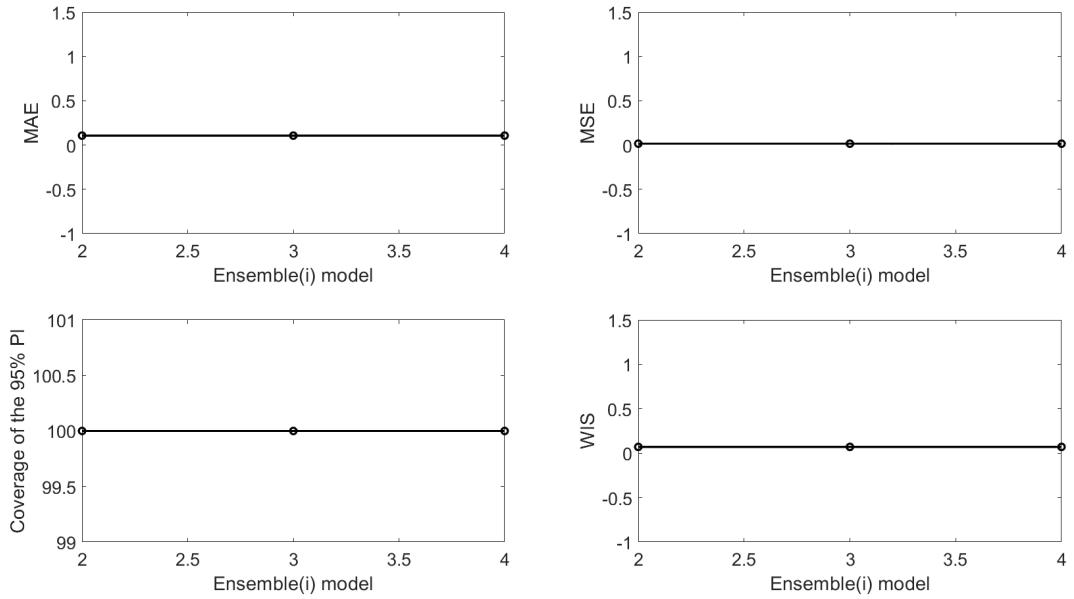
**Figure 26:** Shows 5-week forecasts from the top-ranking sub-epidemic models of the West weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. Circles indicate the observed data points.



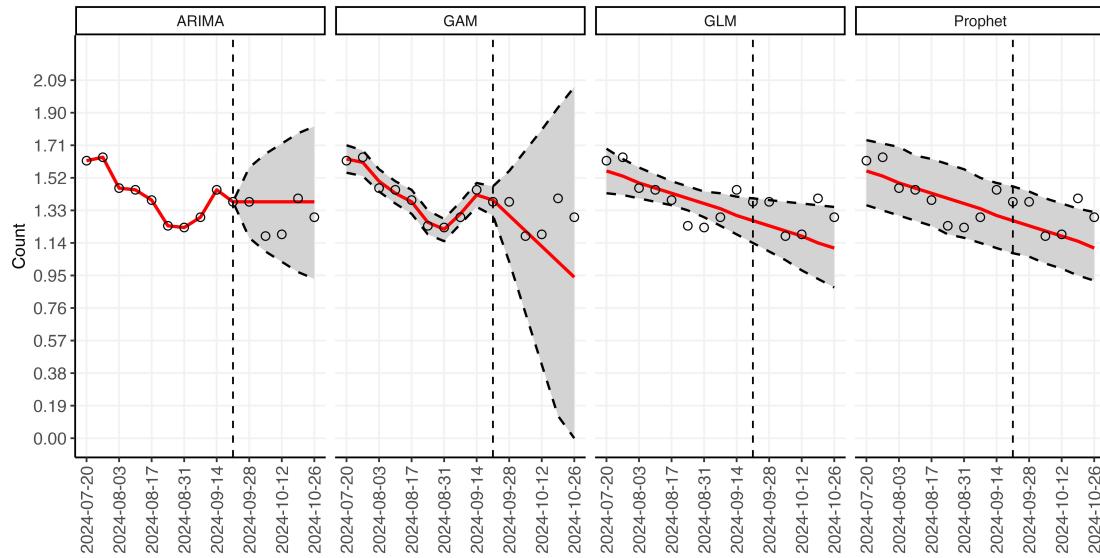
**Figure 27:** Performance metrics for 5-week forecasts based on the top-ranking sub-epidemic models for the West weekly influenza cases.



**Figure 28:** Displays 5-week sub-epidemic ensemble model forecasts for the West weekly influenza cases. The vertical line marks the start time of the forecast. The circles represent the data points, with model fits shown as solid lines and 95% prediction intervals indicated by shaded areas.



**Figure 29:** Performance metrics for 5-week forecasts based on the ensemble sub-epidemic models for the West weekly influenza cases.



**Figure 30:** Shows 5-week forecasts from the ARIMA, GLM, GAM AND THE PROPHET models of the West weekly influenza cases. The 95% prediction interval is represented by a shaded area around the solid line, with a vertical line marking the start of the forecast to separate the calibration and forecast periods. The circles indicate the observed data points.

West				
Model	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.11	0.02	100	0.07
2 <sup>nd</sup> ranked model	0.12	0.02	100	0.07
3 <sup>rd</sup> ranked model	0.38	0.17	100	0.19
4 <sup>th</sup> ranked model	0.21	0.06	100	0.12
Weighted ensemble (2) model	0.11	0.02	100	0.07
Weighted ensemble (3) model	0.11	0.02	100	0.07
Weighted ensemble (4) model	0.11	0.02	100	0.07
ARIMA Model	0.10	0.02	100	0.07
GAM	0.18	0.05	100	0.11
GLM	0.12	0.02	80	0.08
Prophet	0.12	0.02	80	0.08

**Table 5:** 5-weeks forecasting performance metrics for West

The figures in **Table 5** compare the performance metrics of various models for forecasting weekly influenza cases in the West. The ARIMA model demonstrates the best overall performance with the lowest MAE (0.10) and competitive MSE (0.02), along with 10% coverage of the 95% prediction interval, indicating strong accuracy and reliability. The 1st ranked model and weighted ensemble models also show strong performance, with identical MAE (0.11), MSE (0.02), and WIS (0.07), and full prediction interval coverage. Other models, such as GAM, GLM, and Prophet, perform slightly worse, with higher error metrics and less reliable prediction intervals.

## Discussion

This section presents a comprehensive comparison of the model performance metrics, summarizing the evaluation of each model's forecasting accuracy and reliability based on the results presented in **Table 6** below.

Model	Midwest				National			
	MAE	MSE	95% PI	WIS	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.29	0.09	40	0.22	0.19	0.04	40	0.15
2 <sup>nd</sup> ranked model	0.22	0.06	100	0.12	0.22	0.05	40	0.15
3 <sup>rd</sup> ranked model	0.24	0.06	100	0.14	0.11	0.02	100	0.07
4 <sup>th</sup> ranked model	0.28	0.09	100	0.17	0.26	0.07	60	0.14
Weighted ensemble (2) model	0.27	0.08	80	0.16	0.19	0.04	40	0.15
Weighted ensemble (3) model	0.28	0.09	100	0.16	0.19	0.04	40	0.15
Weighted ensemble (4) model	0.26	0.07	100	0.16	0.19	0.04	40	0.15
ARIMA Model	0.14	0.03	100	0.11	0.10	0.01	80	0.06
GAM	0.31	0.10	0	0.23	0.18	0.03	60	0.12
GLM	0.31	0.10	0	0.23	0.14	0.02	40	0.12
Prophet	0.31	0.10	0	0.23	0.14	0.02	20	0.12

Model	Northeast				South			
	MAE	MSE	95% PI	WIS	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.09	0.01	100	0.05	0.08	0.01	100	0.05
2 <sup>nd</sup> ranked model	0.16	0.03	100	0.07	0.09	0.01	100	0.07
3 <sup>rd</sup> ranked model	0.11	0.02	100	0.07	0.08	0.01	100	0.06
4 <sup>th</sup> ranked model	0.06	0.01	100	0.06	0.08	0.01	100	0.06
Weighted ensemble (2) model	0.09	0.01	100	0.05	0.08	0.01	100	0.05
Weighted ensemble (3) model	0.09	0.01	100	0.05	0.08	0.01	100	0.05
Weighted ensemble (4) model	0.09	0.01	100	0.05	0.08	0.01	100	0.05
ARIMA Model	0.08	0.01	100	0.05	0.09	0.01	100	0.05
GAM	0.20	0.06	100	0.11	0.18	0.01	100	0.05
GLM	0.04	0	80	0.03	0.12	0.01	100	0.05
Prophet	0.04	0	100	0.03	0.12	0.01	80	0.06

Model	West			
	MAE	MSE	95% PI	WIS
1 <sup>st</sup> ranked model	0.11	0.02	100	0.07
2 <sup>nd</sup> ranked model	0.12	0.02	100	0.07
3 <sup>rd</sup> ranked model	0.38	0.17	100	0.19
4 <sup>th</sup> ranked model	0.21	0.06	100	0.12
Weighted ensemble (2) model	0.11	0.02	100	0.07
Weighted ensemble (3) model	0.11	0.02	100	0.07
Weighted ensemble (4) model	0.11	0.02	100	0.07
ARIMA Model	0.10	0.02	100	0.07
GAM	0.18	0.05	100	0.11
GLM	0.12	0.02	80	0.08
Prophet	0.12	0.02	80	0.08

**Table 6:** Comparison of Calibration Performance based on MAE, MSE, 95% PI, and WIS

## Model Performance Discussion

**Best Model Across Regions:** The **ARIMA model** consistently emerges as the best performer across all regions:

- **Midwest:** It achieves the lowest MAE (0.14), MSE (0.03), and WIS (0.11) while providing 100% coverage of the 95% prediction interval, indicating excellent accuracy, reliability, and calibration.
- **National:** ARIMA excels with the lowest MAE (0.10), MSE (0.01), and WIS (0.06) and maintains 80% prediction interval coverage, slightly less than some other models but still effective.
- **Northeast:** ARIMA ties with several weighted ensemble models and the 1st ranked model in achieving the lowest WIS (0.05), 100% 95% PI coverage, and a competitive MAE (0.08), showcasing consistent reliability and precision.
- **South:** Though the ARIMA model's MAE (0.09) is slightly higher than the 1st ranked model (0.08), its MSE (0.01), WIS (0.05), and 100% prediction interval coverage keep it on par with top performers.
- **West:** ARIMA is the standout performer with the lowest MAE (0.10), excellent MSE (0.02), and WIS (0.07), tied with the 1st ranked model and weighted ensembles. Its 100% coverage of the 95% PI confirms its reliability.

### Close Contenders:

- **Weighted Ensemble Models (2, 3, 4):** These models consistently perform close to ARIMA in most regions, particularly in terms of MSE and WIS.
  - In the **Northeast** and **South**, their MAE (0.08–0.09), MSE (0.01), and WIS (0.05) match the ARIMA model, showcasing their robustness.
  - In the **West**, they share the same strong performance metrics as ARIMA (MAE 0.11, MSE 0.02, WIS 0.07, and 100% PI coverage).
- **1st Ranked Model:** Performs well in specific regions:
  - In the **South**, it matches or outperforms ARIMA with the lowest MAE (0.08), MSE (0.01), and WIS (0.05), along with 100% PI coverage.
  - In the **Northeast**, its performance is comparable to ARIMA with similar metrics (MAE 0.09, MSE 0.01, WIS 0.05, and 100% PI coverage).
  - However, it underperforms slightly in the **Midwest** and **National** settings with higher MAE and WIS values compared to ARIMA.
- **2nd Ranked Model:** It is competitive in the **Midwest**, achieving low MAE (0.22) and WIS (0.12) with 100% PI coverage. In the **West**, it delivers solid performance but is outperformed by ARIMA and the weighted ensembles.

### Underperforming Models:

- **3rd Ranked Model:** Performs poorly in most regions except the **Midwest** (MAE 0.24, WIS 0.14) and **National** (MAE 0.11, WIS 0.07). Particularly in the **West**, it has the highest MAE (0.38), MSE (0.17), and WIS (0.19), making it less effective model.
- **4th Ranked Model:** Performs moderately across regions, achieving reasonable 95% PI coverage in the Northeast (100%), but with relatively high MAE (0.28) and WIS (0.17) in the Midwest. While it is not the worst-performing model, its higher error metrics make it less effective compared to the top-ranked models.

- **GAM, GLM, Prophet:** Across all regions, these models consistently show higher MAE, MSE, and WIS values compared to ARIMA and other top models.
  - In the **Midwest**, GAM and GLM have a high MAE (0.31) and WIS (0.23), indicating poor accuracy and calibration.
  - In the **South**, GLM and Prophet slightly improve (MAE 0.12), but their 95% PI coverage is lower (80% for Prophet) or non-existent (GLM), reducing reliability.
  - In the **West**, GAM performs poorly with a high WIS (0.11), while GLM and Prophet offer slightly better performance but remain behind ARIMA and the weighted ensembles.

Hence from the above analysis, we observe the following:

- **Best Model Overall:** The ARIMA model consistently leads in accuracy, precision, and reliability across regions.
- **Strong Contenders:** Weighted ensemble models and the 1st ranked model often rival ARIMA, particularly in the **Northeast** and **South**.
- **Weak Performers:** The 3rd ranked model underperforms significantly in the **West**, while GAM, GLM, and Prophet are less reliable and accurate across all regions.

## Insights and Public Health Implications

The detailed analysis of the *n*-sub-epidemic framework and the *STAT MOD PREDICT* toolbox models provides direct insights into how these tools can influence public health decisions. By comparing model performance across metrics such as MAE, MSE, 95% prediction interval (PI) coverage, and WIS, the analysis demonstrates the practical applications of these frameworks in improving epidemic forecasting and guiding effective public health responses.

High-performing models, such as the ARIMA model and Weighted Ensembles, support early detection of outbreaks, efficient resource allocation, and targeted interventions. The ability to capture complex epidemic dynamics ensures reliable forecasts that inform public health strategies, enabling authorities to respond proactively and minimize the impact of infectious diseases.

## Limitations

- **Data Quality and Availability:** Influenza case data relies on reporting systems that may have inconsistencies, delays, or under reporting, potentially affecting the accuracy of forecasts. Similarly, wastewater data may vary in sampling frequency, data collection methods, and regional coverage, limiting the generalizability of the findings.
- **Temporal Scope:** The study's time frame (26 weeks) limits its ability to assess long-term trends, seasonal variations, or the impact of evolving public health measures.

## Future Work

- **Extended Time Periods:** Future studies should consider extended time periods to capture a broader range of influenza dynamics, including seasonal and inter-annual variations.
- **Error Structures:** This analysis only utilized a normal error structure. Future work will explore alternative error structures, such as Poisson and negative binomial errors, to better model the variability and distribution of influenza case data.

## Conclusion

This study explored advanced predictive models using the *n-sub epidemic model* and the *STAT MOD PREDICT toolbox* to improve influenza forecasting across the United States. The analysis identified the *ARIMA model* as the most effective, consistently achieving the lowest error metrics and reliable prediction intervals, particularly at the national and regional levels. Models within the *n-sub epidemic model* framework, such as Weighted Ensembles, also performed well, excelling in regions with complex epidemic dynamics.

The findings emphasize the importance of accurate forecasting for public health decision-making, supporting early interventions, efficient resource allocation, and tailored regional strategies. By combining robust statistical modeling with innovative surveillance techniques, this work offers a reliable framework for mitigating influenza's public health impact.

## References

- [1] Centers for Disease Control and Prevention, "Influenza Surveillance Reports," Available: <https://www.cdc.gov/flu/weekly/overview.htm>, Accessed: Nov. 2024.
- [2] Centers for Disease Control and Prevention, "National Wastewater Surveillance System: Data Methods," Available: <https://www.cdc.gov/nwss/about-data.html#data-method>, Accessed: Nov. 2024.
- [3] A. M. Bleich, "STAT MOD PREDICT Toolbox," Available: <https://github.com/bleicham/StatModPredict>, Accessed: Nov. 2024.
- [4] G. Chowell, "Ensemble n-Subepidemic Framework for Forecasting Epidemics," Available: [https://github.com/gchowell/ensemble\\_n-subepidemic\\_framework](https://github.com/gchowell/ensemble_n-subepidemic_framework), Accessed: Nov. 2024.
- [5] G. Chowell, A. Bleichrodt, and R. Luo, "Parameter Estimation and Forecasting with Quantified Uncertainty for Ordinary Differential Equation Models Using QuantDiffForecast: A MATLAB Toolbox and Tutorial," *Statistics in Medicine*, vol. 43, pp. 1826–1848, 2024. DOI: <https://doi.org/10.1002/sim.10036>.
- [6] G. Chowell, A. Bleichrodt, S. Dahal et al., "GrowthPredict: A Toolbox for Fitting and Forecasting Growth Trajectories Using Phenomenological Models," *Scientific Reports*, vol. 14, no. 1630, 2024. DOI: <https://doi.org/10.1038/s41598-024-51852-8>.
- [7] G. Chowell, S. Dahal, A. Bleichrodt et al., "SubEpiPredict: A Toolbox for Fitting and Forecasting Growth Trajectories Using the Ensemble n-Sub-epidemic Framework," *Infectious Disease Modelling*, vol. 9, pp. 411-436, 2024. DOI: <https://doi.org/10.1016/j.idm.2024.02.001>.