# Short Term Influenza Forecasting in the Hospital Environment Using a Bayesian Kalman Filter

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Shankar Kaleeswaran Mani, M.S.

Graduate Program in Public Health: Biostatistics

The Ohio State University

2024

Master's Examination Committee:

Dr. Grzegorz Rempala, Advisor

Dr. Eben Kenah

Dr. Fernanda Schumacher

## Abstract

Accurate forecasting of weekly number of influenza (flu) lab tests and positive cases is vital for hospitals to provide adequate patient care at the right time. It also helps prevent shortages or overages of staffs and supplies. In this paper we present a practical implementation of a Bayesian Kalman filter to forecast weekly flu test and positive cases in a hospital environment. By integrating real time hospital data, this framework offers a robust methodology for accurately predicting flu volume one to four weeks out with a reasonable accuracy.

## Acknowledgments

I would like to express my sincere gratitude to my advisors Dr. Rempala, Dr. Schumacher, and Dr. Kenah for their expertise and guidance that led to the completion of this thesis. Their constant guidance and encouragement throughout this experience have deeply impacted my curiosity and growth as a researcher and as a person. I am also grateful to my employer, OhioHealth, for supporting and providing me with the clinical data that contributed to the completion of this thesis. Additionally, I extend heartfelt thanks to my wife, Kalpita Mehta, and my daughter, Trisha Mani, for their unwavering support and understanding throughout this journey. Their love and encouragement have been a constant source of strength and motivation.

# Vita

| | |
|---|---|
| 2003-2007 | B.E. Electronics and Communication, Anna University |
| 2007-2010 | Assistant Systems Engineer, Tata Consultancy Services |
| 2010-2012 | M.S. Operations Research and Supply Chain, Case Western Reserve University |
| 2012-2015 | Business Intelligence Analyst, Cleveland Clinic Foundation |
| 2015-present | Data Science Advisor, OhioHealth |

# Fields of Study

Major Field: Public Health (Biostatistics)

# Table of Contents

**Page**

# List of Figures

# Chapter 1: Introduction

Influenza places a significant burden on the US health care system annually [3]. According to [5] (CDC), between 2010 and 2023, about 9.3 million to 41 million influenza illnesses occured annually, resulting in 100,000-710,000 annual hospitalizations and 4,900-51,000 annual deaths. For the most recent seasonal year (2022-2023), CDC estimates about 31 million illnesses with 360,000 hospitalizations and 21,000 deaths.

During an influenza seasonal year, the provision of optimal patient care within a hospital system requires effective preparations for incoming influenza patients [6, 8]. Hence, accurate seasonal influenza volume estimates are paramount to facilitate resource allocation, budgeting, and delivering quality patient care. This underscores the importance of developing a predictive model that enables hospitals to anticipate and accommodate fluctuations in weekly case volumes. By accurately forecasting weekly influenza cases one to four weeks out, our objective is to help hospital system strategize resource allocation, preemptively adjust staffing levels, optimize operational workflows to accommodate patient surges effectively and ensure an adequate supply of beds and laboratory test specimens.

The data used for this project is collected from OhioHealth, a mid-sized nonprofit hospital system in central Ohio. We extracted data on the weekly count of all

laboratory tests performed and the resulting positive cases, from September 2016 to August 2023. This period was chosen since the health system implemented a new Electronic Medical Record (EMR) system mid 2015, and reliable data was available only beginning September 2016 for all locations.

For our analysis, we defined a seasonal influenza year as starting from September of that calendar year and ending in August of the subsequent year. The data includes all patients who had an influenza lab test order at one of the hospital locations and whether they tested positive [21, 7]. In total, we had seasonal influenza data for 8 different years (September 2016 to August 2023).

Compartmental epidemic models [12], such as the Susceptible-Infected-Recovered (SIR) model, offer valuable insights into the dynamics of infectious diseases, they are often geared towards understanding broader epidemiological trends, such as peak times and transmission rates, etc. Recent approaches linking SIR to survival analysis, as in [14, 13] and [19], were instrumental during the COVID-19 epidemic.

For further insights into the Extended Kalman Filter (EKF) methodology employed in influenza modeling study, readers are encouraged to refer to the work by [22]. Additionally, for a comprehensive understanding of nationwide influenza forecasting, the CDC's publications on influenza Forecasting [15] provide valuable context and insights into the recent development. For the immediate decision-making needs within hospital systems, where timely care for patients is critical [17], our focus lies in short-term prediction using readily available hospital data by taking advantage of the Kalman filter methodology.

The rest of the manuscript is organized as follows. Chapter 2 gives a gentle introduction to both linear Kalman filter and Bayesian Kalman filter methodologies to estimate parameters for forecasting weekly influenza lab tests and positive cases. It details the development of our final models, including the derivation of informative priors Section 2.4 from other seasonal years for the Markov Chain Monte Carlo (MCMC) estimation. In Chapter 3, we assess the performance of our final models by comparing actual versus predicted values over a forecasting horizon of 1-4 weeks, evaluating performance metrics such as Mean Absolute Deviation (MAD) and coverage probability. Finally, we highlight the strengths and weaknesses of our model in the Chapter 4, along with suggestions for further improvements.

# Chapter 2: Methods

In this paper, we propose a bivariate linear Kalman filter to simultaneously model and predict weekly total influenza tests ($T$) and the resulting positive cases ($P$) in the OhioHealth hospital system. The assumption is that both $T$ and $P$ are influenced by the unknown true incidence of influenza in the source population. By adjusting for this hidden state using a Kalman filter, we hope to extract a meaningful signal from noisy data.

While the dynamics of infectious disease epidemics, such as influenza, are inherently nonlinear and subject to complex interactions, our focus lies on short-term predictions (1 to 4 weeks), where it is reasonable to approximate the influenza dynamics as locally linear. Because the Kalman filter is an optimal filter for a linear Gaussian process, it is expected to provide a good reconstruction of the means of $T$ and $P$ using aggregated weekly data. The use of a Bayesian framework allows us to borrow information from other seasonal years to create informative priors for the MCMC estimation process, [10].

## 2.1 Linear Kalman Filter

The Kalman filter, which was first introduced by Rudolph Emil Kalman, is a method for state estimation in a dynamic system [11]. It has since been used in many areas, such as robotics motion planning, object navigation, signal processing, and wireless networks to extract signals from noisy data. At a basic level, the Kalman filter estimates the true state of a system at a current step (say $t$) by combining the estimate made at the previous step $(t-1)$ for the current step $(t)$ with the currently observed value (at $t$) to better approximate the true state of the system at $t$. The current observed value is assumed to be noisy (e.g., due to faulty devices) and need a way to get a precise estimate of its true value. The aim of a filter is to find optimal values for the parameters by reducing the deviance between the predicted state and observed value.

A Kalman filter has two steps that operate at each time step: the predict step and the update step 2.1. The state equation predicts the value at the next time step with uncertainty estimated through an uncertainty extrapolation equation. After observing the value at the next time step (which is assumed to include some error), this estimate is then updated according to a factor called the Kalman gain. The filter then uses the updated predictions and the state equation to predict the next step value, and the system goes on in a loop. The filter has been extensively discussed in the literature, and details can be found in [18], for example.

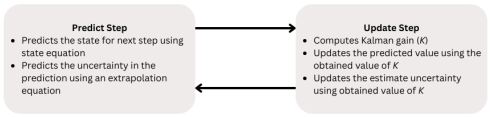


Figure 2.1: Predict and update step in a Kalman filter.

Consider a bivariate state space model [SSM, 1] whose state and observation equation is linear with multivariate normally distributed error, which can be defined as follows:

- *State Equation:*

$$S_t = FS_{t-1} + E_{t-1}, \tag{2.1}$$

where

$$E_{t-1} \sim \text{MVN}(0, \Sigma_s),\ \Sigma_s = \begin{pmatrix} \sigma_T^2 & \rho_s \sigma_T \sigma_P \\ \rho_s \sigma_T \sigma_P & \sigma_P^2 \end{pmatrix},$$
$$F = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix},\ \text{and}\ S_t = \begin{pmatrix} T_t \\ P_t \end{pmatrix}.$$

- *Observation Equation:*

$$O_t = HS_t + V_t, \tag{2.2}$$

where

$$V_t \sim \text{MVN}(0, \Sigma_o), \;\; \Sigma_o = \begin{pmatrix} \nu_T^2 & \rho_o \nu_T \nu_P \\ \rho_o \nu_T \nu_P & \nu_P^2 \end{pmatrix}, O_t = \begin{pmatrix} T_{o,t} \\ P_{o,t} \end{pmatrix},$$

and $H$ is a data transformation matrix (often an identity matrix).

Following [2] and [16], the set of equations that are needed in a Kalman filter to perform the predict and update step are given by:

1. Predict Step:

   - *Estimate next state value:*

   $$\hat{S}_{t+1,t} = F\hat{S}_{t,t}; \tag{2.3}$$

   - *Estimate next state variance:*

   $$\hat{P}_{t+1,t} = F\hat{P}_{t,t}F^\top + \Sigma_s; \tag{2.4}$$

   where $\hat{P}_{t,t}$ is defined recursively in equation (2.7).

2. Update Step:

   - *Compute Kalman gain (K):*

   $$\hat{K}_t = \hat{P}_{t,t-1}H^\top (H\hat{P}_{t,t-1}H^\top + \Sigma_o)^{-1}; \tag{2.5}$$

   - *Update state estimate:*

   $$\hat{S}_{t,t} = \hat{S}_{t,t-1} + \hat{K}_t(O_t - H\hat{S}_{t,t-1}); \tag{2.6}$$

   - *Update state variance:*

   $$\hat{P}_{t,t} = (I - \hat{K}_t H)\hat{P}_{t,t-1}(I - \hat{K}_t H)^\top + \hat{K}_t \Sigma_o \hat{K}_t^\top. \tag{2.7}$$

## 2.2 Bayesian Kalman Filter

The setup of a bivariate Kalman Bayesian Markov Chain Monte Carlo (MCMC) model is shown in Algorithm 1.

---

**Algorithm 1** Bayesian Kalman Filter.

**Require:** Initialize step, at $t = 0$:

$$\hat{S}_{1,0} = S_{init},$$

$$\hat{P}_{1,0} = P_{init}.$$

**for** $t$ in 1 to $n$ **do**:

1. Compute Kalman gain using 2.5.
2. Perform state and variance update using update equations given in 2.6 and 2.7 to obtain $(\hat{S}_{t,t}, \hat{P}_{t,t})$;
3. Predict the next state value using updated estimates and transition equation such that,

$$\hat{S}_{t+1,t} \sim \text{MVN}(F\hat{S}_{t,t}, (F\hat{P}_{t,t}F^t + \Sigma_s));$$

4. Assume multivariate normal distribution for observations, such that

$$O_{t+1} \sim \text{MVN}(\hat{S}_{t+1,t}, \Sigma_o).$$

**end for**

---

The objective of the Algorithm 1 is to estimate the parameters ($F$, $\Sigma_s$ and $\Sigma_o$), such that predicted state values $\hat{S}_{t+1,t}$ provides us with estimates and uncertainty of the hidden states reflecting the true incidence means better than the observed observations $O_{t+1}$, which are considered to be noisy.

## 2.3 Final model for weekly lab test and positive case incidence

We analyzed and modeled each seasonal influenza year independently but borrowed information from other years to construct an informative prior for all the parameters

to be estimated through MCMC sampling. For example, when we analyzed 2016 seasonal data, information from 2017 to 2023 was used to generate an informative prior on all model parameters. Similarly, when we analyzed 2020 seasonal year data, we borrowed information from year 2016 to 2019 and 2021 to 2023. The rationale for this approach is that each influenza season has different dynamics [4]. By modeling each year independently, we allow the model to reflect this variation between years. By using an informative prior based on other years, we take advantage of the fact that some patterns are shared across influenza seasons. See Section 2.4 for more details on how the priors were constructed.

### 2.3.1 Influenza Lab Tests Model

To model the number of weekly lab tests, we considered an ensemble model which integrates the estimates from two different Bayesian Kalman filter setups. After iterating through various model specifications, we found that the ensemble of two distinct Kalman filters produced the best results in terms of Mean Absolute Deviation (MAD) and coverage probability metrics. Specifically, the first model in the ensemble excelled in predicting week 1 cases, while the second model performed better for week 4 predictions. Therefore, combining these two models enabled us to optimize predictions made 1 to 4 weeks ahead.

The first model employs a straightforward bivariate autoregressive of order 1 or AR(1) framework for total lab test and positive cases, similar to the one we considered for the bivariate state space model 2.1. Within this framework, we modeled the uncertainty by specifying unknown state and observation covariance matrices (i.e., the correlation of errors between the two series in each matrix is non-zero). The matrix H is the

identity matrix, and we estimate the $F$ matrix, the AR(1) parameter matrix, and the state and observation covariance matrices $\Sigma_S$ and $\Sigma_O$. The priors for the parameters are informative with information borrowed from other seasonal years.

In contrast, the second model assumes zero correlation between the two series, decoupling their respective dynamics. By assuming zero correlation between the two series, we treat each series as an independent normal distributed random variable, effectively decoupling their respective dynamics. The model has a first-order autoregressive parameter, $F$ matrix, and state and observation variance terms, $\Sigma_s$ and $\Sigma_o$, respectively to be estimated through MCMC sampling. Again, the priors are informative with the information borrowed from other seasonal influenza years.

We combined the two models within a Bayesian hierarchical framework to produce an ensemble estimate such that $M_1W_1 + M_2(1 - W_1)$. This hierarchical setup facilitates the integration of two models through an unknown weight parameter, $W_1$, governing the relative influence of each model. Critically, $W_1$ is not predetermined but instead inferred directly from the data with an informative beta prior, allowing the data to dictate the influence of each model on the final predictions.

### 2.3.2 Influenza Positive Cases Model

For the weekly number of positive influenza cases, we considered a bivariate Kalman filter on a difference scale ($d = 1$). The structure of this model is very similar to the model considered for total lab test Section 2.3.1 with few modifications. First, the two series were decoupled such that the correlation of their error terms is zero in both state and observation covariance matrix. Second, each series not only depends on the lagged value of its own but also the lagged value of the second series (see $F$

matrix structure below). Third, the second series,$P_t$, (positive cases volume) has a covariate matrix $X_{t-1}$, which is a smoothed total test volume derived by smoothing other seasonal year data. For example, if we are analyzing 2016 data, we derived a smoothed total lab tests volume from seasonal year 2017 to 2023. Finally, we did not use ensemble technique and only one bivariate Kalman filter was used. The priors for the parameters, $F, W, \Sigma_s, \Sigma_o$ are informative with information borrowed from other seasonal years, Section 2.4. Note, $W$ is the coefficient for smoothed predictor $X_{t-1}$ to be estimated.

- *State Equation:*

$$S_t = FS_{t-1} + WX_{t-1} + E_{t-1},$$

where

$$S_t = \begin{pmatrix} \Delta T_t \\ \Delta P_t \end{pmatrix}, \qquad F = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix},$$
$$\Sigma_s = \begin{pmatrix} \sigma_T^2 & 0 \\ 0 & \sigma_P^2 \end{pmatrix}, \quad E_{t-1} \sim \text{MVN}(0, \Sigma_s);$$

- *Observation Equation:*

$$\Delta O_t = HS_t + V_t,$$

where

$$V_t \sim MVN(0, \Sigma_o),$$
$$\Sigma_o = \begin{pmatrix} \nu_T^2 & 0 \\ 0 & \nu_P^2 \end{pmatrix}, \qquad \Delta O_t = \begin{pmatrix} \Delta T_{o,t} \\ \Delta P_{o,t} \end{pmatrix}.$$

## 2.4 Priors for Bayesian Kalman Filter Model

In the Bayesian Kalman filter model for estimating weekly lab tests and positive cases volume, we utilized informative priors by incorporating relevant historical information to enhance the estimation process. We formed informative priors for key parameters the autoregressive matrix ($F$), the state ($\Sigma_s$) and observation ($\Sigma_o$) covariance matrix,

and the weight parameter ($W_1$) (only for lab tests model). These priors are designed to leverage insights from other seasonal influenza years, thereby enabling our model for more accurate and robust estimation. This is also one of the main reason why we decided to use Bayesian MCMC estimation for the Kalman filter. Below is a high level summary.

**Algorithm 2** Informative Prior Construction

**for** $Model$ in $totalLabTests$ and $positiveCases$ **do**:
  **for** $SeasinalYear$ in 2016 to 2023 **do**:
    1. Define a non-informative prior for model parameters ($F$, $\Sigma_s$, $\Sigma_o$, $W_1$).
    2. Using Algorithm 1, perform MCMC sampling using the appropriate $Model$ discussed in Section 2.
    3. For each model parameter, extract mean and standard deviation from the posterior distribution
  **end for**
**end for**

The approach in Algorithm 2 allowed us to capture the different influenza dynamics in the other seasons. For example, to create informative priors for 2016 positive cases model parameters ($F$, $\Sigma_s$ and $\Sigma_o$), we initially ran the final positive case bivariate Kalman filter model discussed in Section 2.3.2 independently for each seasonal year (2017 to 2023), assuming a non-informative beta prior for $F$ and a non-informative uniform prior for $\Sigma_s$ and $\Sigma_o$. We combined the resulting posterior distribution of each model parameters, using mean and standard deviation, for seasonal years (2017 to 2023) to create an informative beta and uniform priors for the 2016 influenza season. A similar approach was employed for deriving informative priors for the final lab tests model for each seasonal year in the final lab tests model, Section 2.3.1.

### 2.4.1 AR Parameters

For the AR parameters, we first ran uninformative beta priors and the hyperparameters were set by borrowing information from other seasonal influenza years. By leveraging the patterns and dynamics observed in the other influenza seasons, we enhanced the model's ability to capture the temporal dependencies inherent in the influenza process. Here, we chose a beta prior because often the case that in a influenza cycle the incidences of cases are often positively correlated over time.

$$F = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}, \quad \phi_{ij} \sim \text{Beta}(a_{ij}, b_{ij})$$

### 2.4.2 State and Observation Covariance Matrix

Informative priors were incorporated for the state and observation covariance matrix, which govern the uncertainty in the state dynamics and the relationship between the observed data and the underlying state variables. We first used an uninformative uniform prior and then estimated the hyperparameters using posterior data from other seasonal years. By using the covariance patterns observed in these years, we ensure that the model appropriately accounts for the variability and correlation present in influenza volume data. For instance, consider state and observation covariance matrix below, all the hyper parameter values were formed from other seasonal years.

$$\Sigma_s = \begin{pmatrix} \sigma_T^2 & \rho_s \sigma_T \sigma_P \\ \rho_s \sigma_T \sigma_P & \sigma_P^2 \end{pmatrix}, \quad \Sigma_o = \begin{pmatrix} \nu_T^2 & \rho_o \nu_T \nu_P \\ \rho_o \nu_T \nu_P & \nu_P^2 \end{pmatrix},$$

$$\sigma_T \sim \text{Unif}(T_{s.a}, T_{s.b}), \qquad \sigma_P \sim \text{Unif}(P_{s.a}, P_{s.b}),$$

$$\rho_s \sim \text{Unif}(\rho_{s.a}, \rho_{s.b}), \qquad \nu_T \sim \text{Unif}(T_{o.a}, T_{o.b}),$$

$$\nu_P \sim \text{Unif}(P_{o.a}, P_{o.b}), \qquad \rho_o \sim \text{Unif}(\rho_{o.a}, \rho_{o.b}).$$

### 2.4.3 Ensemble Weight Parameter

Finally, in the ensemble model for total lab cases, the weight parameter, $W_1$, is informative beta prior and the information is borrowed from other influenza years using the process described in Algorithm 2.

$$W_1 \sim \text{Beta}(W_a, W_b).$$

# Chapter 3: Results

## 3.1 Data Summary

The dataset encompasses weekly counts of lab test and influenza cases recorded across multiple seasonal years, spanning from September 2016 to August 2023, extracted from a Hospital EMR system. Notably, the dataset includes observations from both traditional influenza seasons (2016 to early 2020) and COVID years (early 2020 to 2023), during which the testing and incidence for COVID-19 significantly impacted public health surveillance efforts and confounded the lab test and influenza case volume.

Figure 3.1 plots the incidence of weekly lab test and positive cases across different seasonal years. Figures 3.2 and 3.3 depict a box plot of lab tests and influenza case counts by seasonal year, providing a visual representation of the variability and distribution of flu activity across different years. The inclusion of COVID years in the dataset underscores the dynamic nature of public health surveillance and the challenges posed by overlapping epidemics and introduces confounding factors, such as increased COVID test volume and heightened awareness of respiratory illness. We can clearly see that volume of lab tests is significantly higher during the COVID years highlighting this confounding. This was also one of the reason why we modeled each

influenza season independently but borrowed information from one another through informative priors.

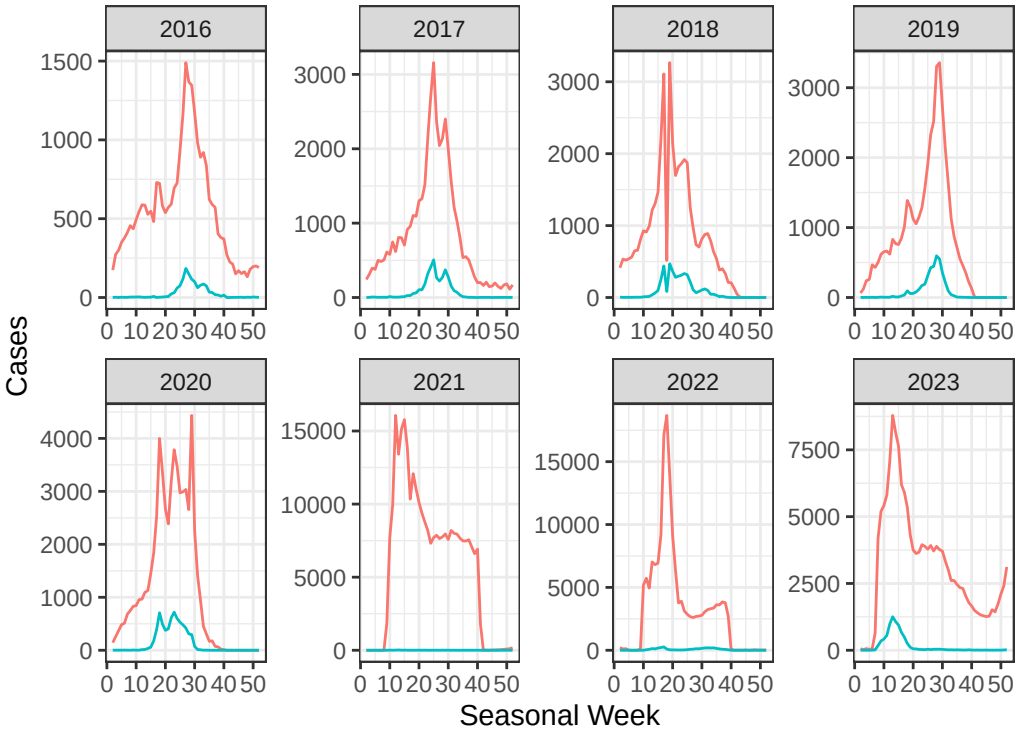


Figure 3.1: Weekly incidence of lab tests and positive cases by seasonal years. Total lab test volume (in red) and positive case volume (in blue).

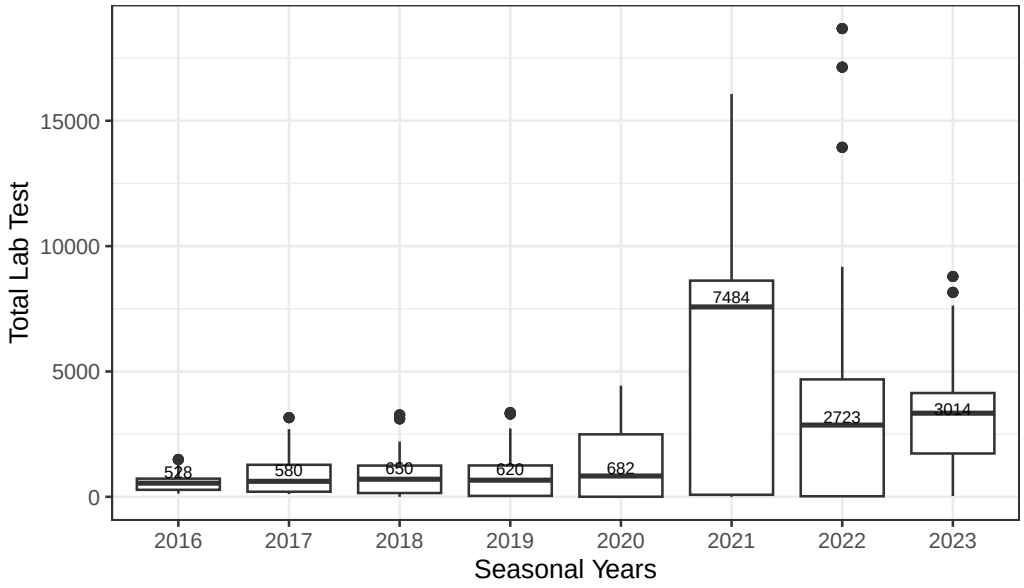


Figure 3.2: Box plot of total lab tests by influenza seasonal year.

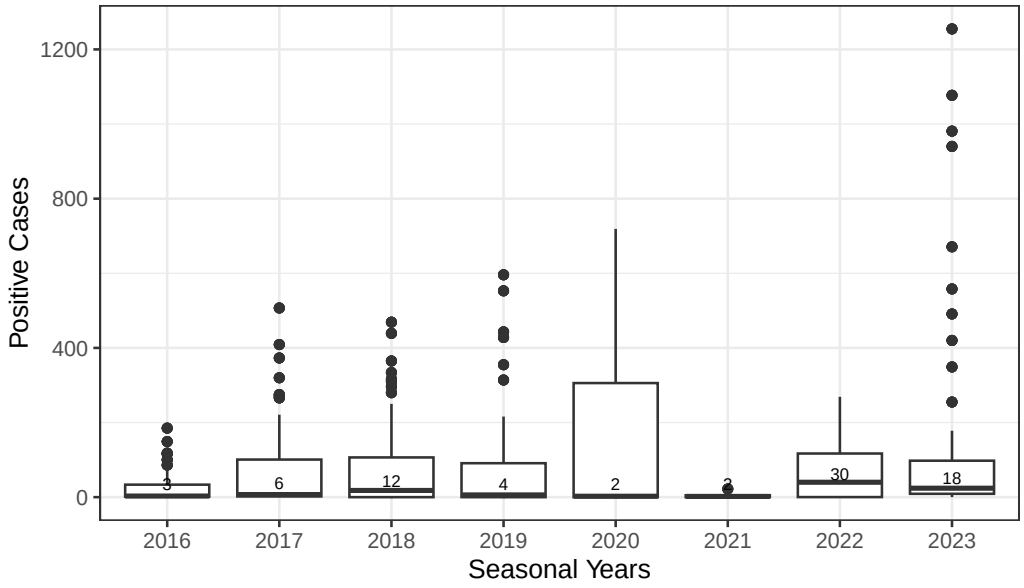


Figure 3.3: Box plot of positive cases by influenza seasonal year.

## 3.2 Predicting influenza lab tests and positive cases using a rolling 4 weeks out forecast

The predictive performance of the model was evaluated using a rolling forecasting framework, where forecasts were generated for the next weeks 4 out at each forecasting horizon, weeks 4 to 52 of a seasonal year. The predicted volume, along with their corresponding 95% credible intervals, were compared against the actual cases for each forecasting horizon. Figures 3.4-3.7 illustrate the comparison.

The rolling forecast methodology involved generating forecasts 1 to 4 weeks ahead using a rolling window (horizon) that started at week 4 and extended up to week 52 of the seasonal year. At each iteration of the rolling window, the model was trained on historical flu data up to the forecast horizon, and predictions (and credible intervals) were made for the next 4 weeks from the forecasting horizon.

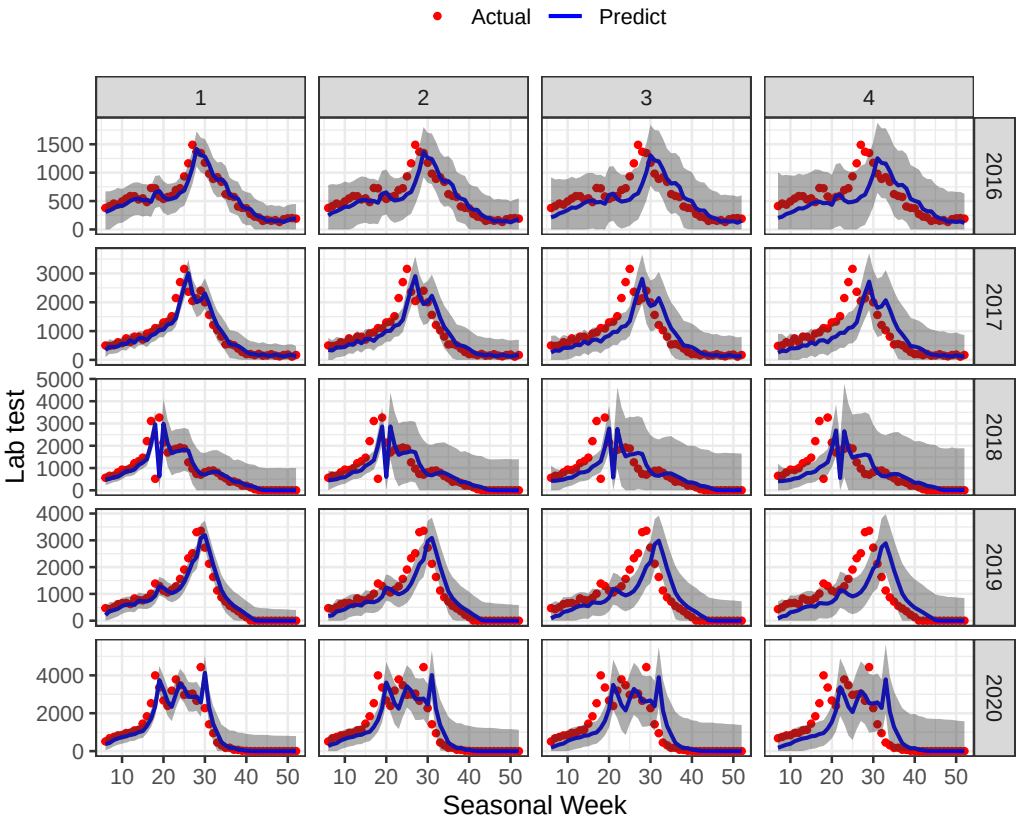


Figure 3.4: Total lab test actuals vs. prediction 1 to 4 weeks out (non COVID years).

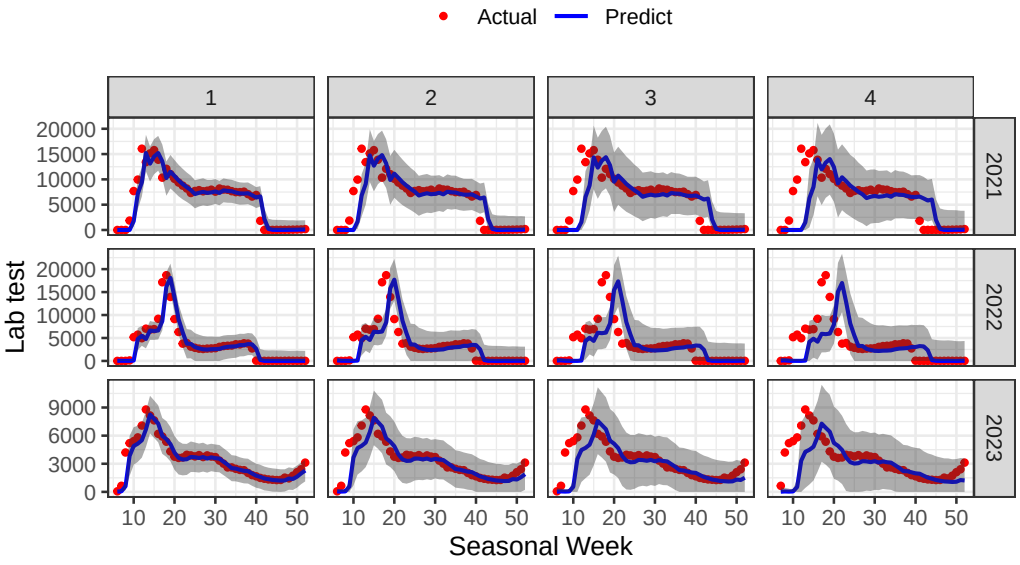


Figure 3.5: Total lab test actuals vs. prediction 1 to 4 weeks out (COVID years).

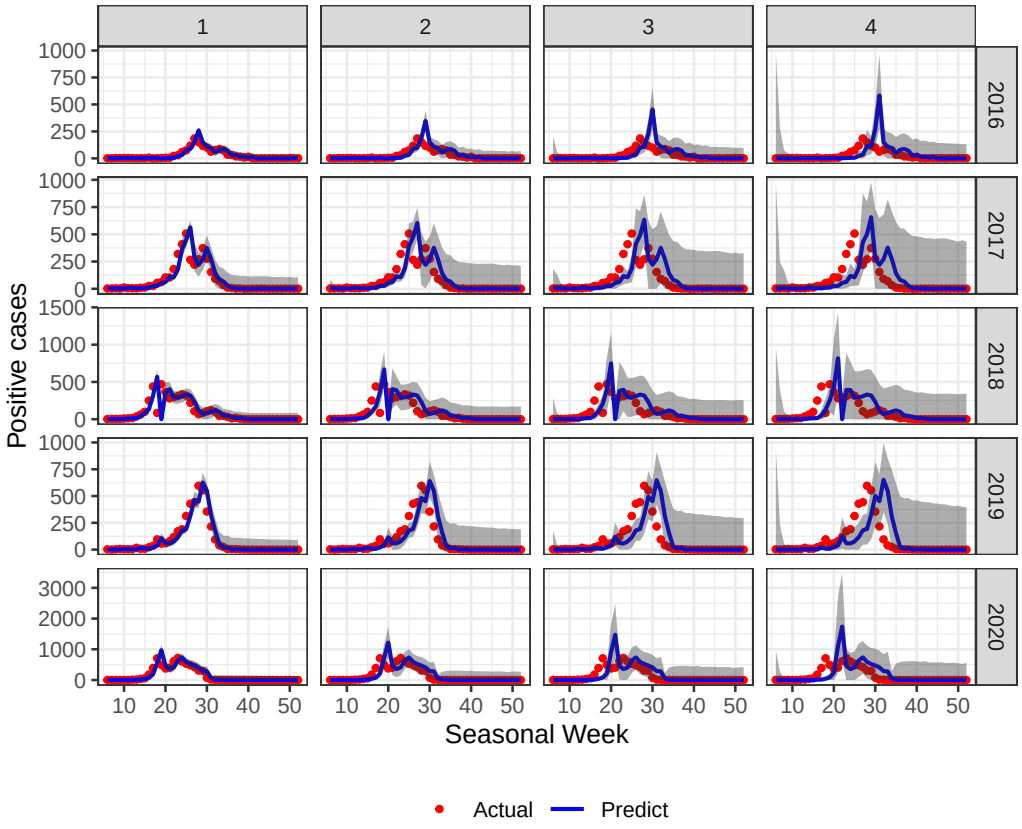


Figure 3.6: Positive cases actuals vs. prediction 1 to 4 weeks out (non COVID years)

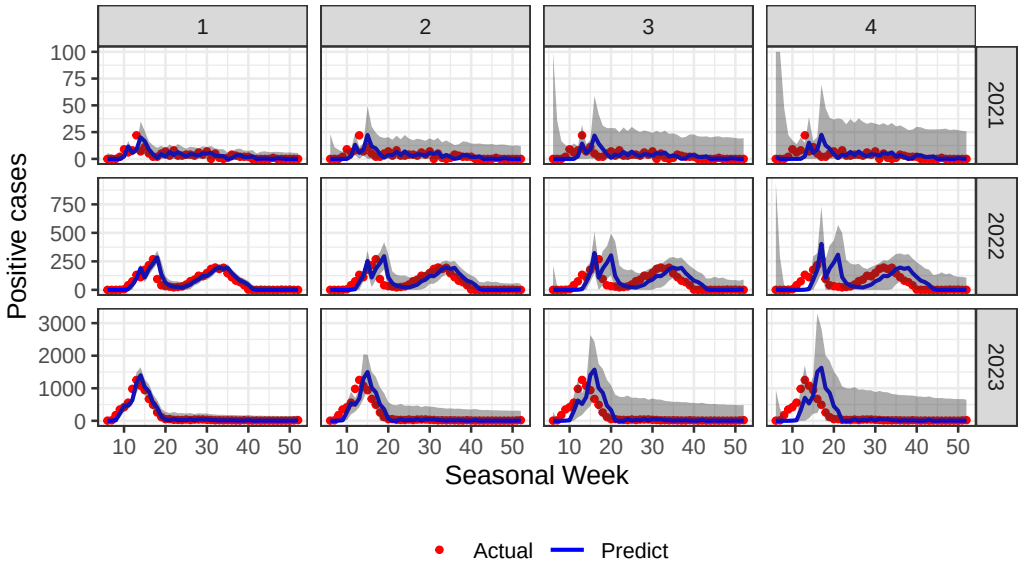


Figure 3.7: Positive cases actuals vs. prediction 1 to 4 weeks out (COVID years).

## 3.3 Assessment of Predictive Performance

The accuracy and reliability of the predictions were assessed using mean absolute deviation (MAD), and percent of observations contained with in the 95% credible intervals (coverage probability) for each of the predictions made 1 to 4 weeks out. Figure 3.8 illustrates the performance of the models.

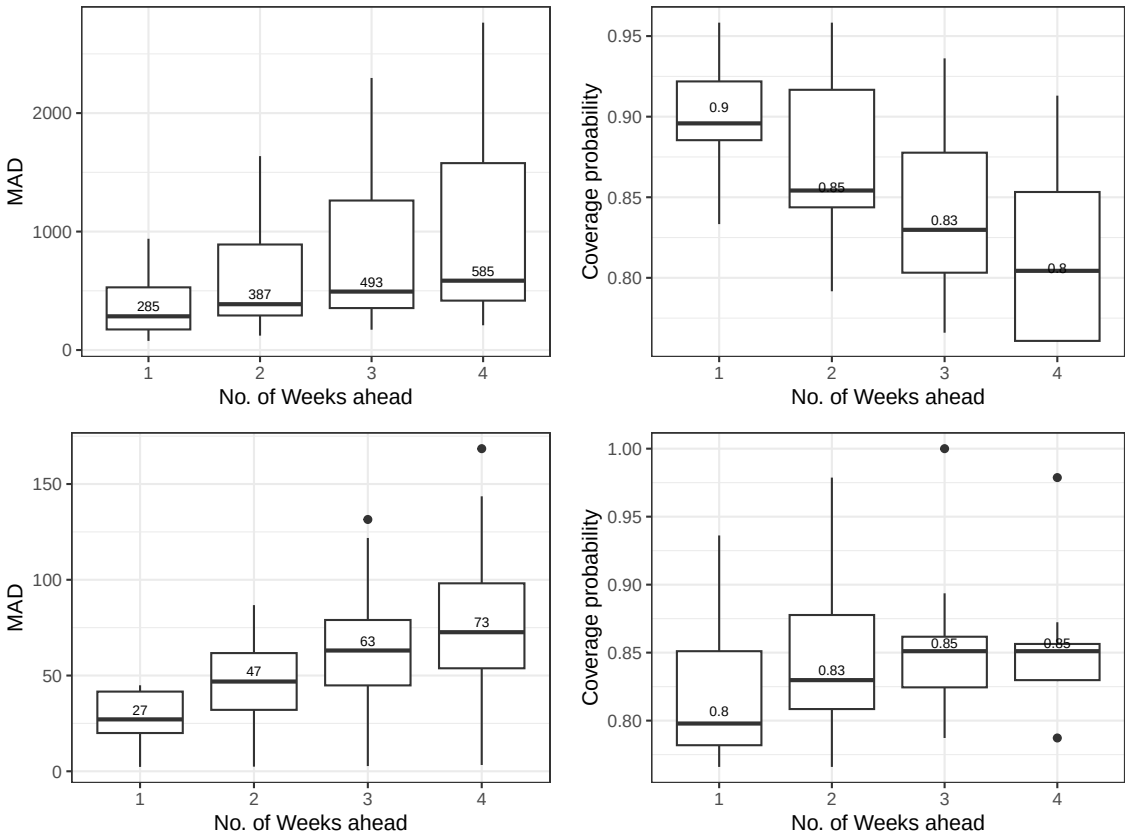


Figure 3.8: Model performance. Mean absolute deviation (MAD) and coverage probability for total lab test (top) and for positive cases (bottom).

## 3.4 Summary and Results

For predictions made for 1 and 2 weeks ahead, the model demonstrated robust performance, with MAD and coverage probabilities within acceptable values, indicating accurate predictions for both lab tests and positive cases. Week 3 predictions also exhibited satisfactory performance, albeit with a slightly higher MAD value and lower coverage probability. However, the model encountered challenges in predicting volume 4 weeks ahead. Week 4 predictions showed slightly elevated MAD values and

lower coverage probabilities compared to earlier forecasting horizons underpinning the difficulty in predicting the lab test and positive case volume four weeks out.

Note that we implemented two distinct bivariate models to forecast weekly influenza lab tests and positive cases, with each model focusing on optimizing performance for its respective outcome. The lab tests model excelled in predicting lab test volume even though both lab tests and positive cases were modeled together, while the positive case model demonstrated better accuracy in predicting positive cases.

# Chapter 4: Discussion

In summary, the prediction model demonstrated satisfactory performance for forecasting horizons of one to four weeks out. These forecast can be made and updated weekly with real time hospital data and can be modified to produce for a specific hospital location or catchment area. The model's accuracy was slightly diminished for predictions made four weeks ahead, highlighting the inherent challenges with long-term forecasting and reflecting heterogeneities in influenza transmission dynamics during a seasonal year [23]. We experimented different Kalman filters with various ARIMA parameters (p,d,q) including a model using the KFAS package [9]. Our selection of the final model was based on minimizing the MAD and optimizing coverage probability. Through an iterative process, we identified the final models, for total lab test and positive case volume, that balanced accuracy and reliability in forecasting weekly influenza lab tests and positive cases.

While our study employed a linear Gaussian Kalman filter for forecasting weekly influenza lab test and positive cases, it is important to acknowledge that these outcomes are better modeled as count processes, such as Poisson distributions. Although the version of the Kalman filter discussed in this paper is optimal only for linear Gaussian processes, we opted to remain within the Gaussian realm for our modeling approach

because it can be a reasonable approximation for large count data. An avenue for improvement is to adopt a count process framework and refining the Kalman filter methodology [20] to better accommodate the non linearity and count nature of influenza. These findings underscore the need for further research and refinement of modeling approaches to improve the accuracy and reliability of influenza forecasting.

# Bibliography


[1] Marie Auger-Méthé, Ken Newman, Diana Cole, Fanny Empacher, Rowenna Gryba, Aaron A King, Vianey Leos-Barajas, Joanna Mills Flemming, Anders Nielsen, Giovanni Petris, et al. A guide to state–space modeling of ecological time series. *Ecological Monographs*, 91(4):e01470, 2021.

[2] Alex Becker. Kalman Filter Summary. https://www.kalmanfilter.net/multiSummary.html, 2024. Accessed: February 10, 2024.

[3] Arbeitskreis Blut et al. Influenza virus. *Transfusion Medicine and Hemotherapy*, 36(1):32, 2009.

[4] CDC and National Center for Immunization and Respiratory Diseases (NCIRD). How Flu Viruses Can Change. https://www.cdc.gov/flu/about/viruses/change.htm#:~:text=One%20way%20flu%20viruses%20change,)%20and%20NA%20(neuraminidase), 2024. Accessed: 2024, March 25.

[5] Centers for Disease Control. Disease Burden of Flu. https://www.cdc.gov/flu/about/burden/index.html, 2024. Accessed: 2024, February 8.

[6] Leah S Fischer, Scott Santibanez, Richard J Hatchett, Daniel B Jernigan, Lauren Ancel Meyers, Phoebe G Thorpe, and Martin I Meltzer. Cdc grand rounds: modeling and public health decision-making. *Morbidity and Mortality Weekly Report*, 65(48):1374–1377, 2016.

[7] Patrick J Gavin and Richard B Thomson Jr. Review of rapid diagnostic tests for influenza. *Clinical and Applied Immunology Reviews*, 4(3):151–172, 2004.

[8] John W Glasser, Nathaniel Hupert, Mary M McCauley, and Richard Hatchett. Modeling and public health emergency responses: lessons from sars. *Epidemics*, 3(1):32–37, 2011.

[9] Jouni Helske. KFAS: Exponential family state space models in r. *arXiv preprint arXiv:1612.01907*, 2016.

[10] Simon Jackman. Estimation and inference via bayesian simulation: An introduction to markov chain monte carlo. *American journal of political science*, pages 375–404, 2000.

[11] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1), 1960.

[12] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[13] Wasiur R KhudaBukhsh, Caleb Deen Bastian, Matthew Wascher, Colin Klaus, Saumya Yashmohini Sahai, Mark H Weir, Eben Kenah, Elisabeth Root, Joseph H Tien, and Grzegorz A Rempała. Projecting covid-19 cases and hospital burden in ohio. *Journal of Theoretical Biology*, 561:111404, 2023.

[14] Wasiur R KhudaBukhsh, Boseung Choi, Eben Kenah, and Grzegorz A Rempała. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Interface focus*, 10(1):20190048, 2020.

[15] Chelsea S Lutz, Mimi P Huynh, Monica Schroeder, Sophia Anyatonwu, F Scott Dahlgren, Gregory Danyluk, Danielle Fernandez, Sharon K Greene, Nodar Kipshidze, Leann Liu, et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health*, 19(1):1–12, 2019.

[16] Hamed Masnadi-Shirazi, Alireza Masnadi-Shirazi, and Mohammad-Amir Dastgheib. A step by step mathematical derivation and tutorial on kalman filters. *arXiv preprint arXiv:1910.03558*, 2019.

[17] Martin I Meltzer. Modeling in real time during the ebola response. *MMWR supplements*, 65, 2016.

[18] Yan Pei, Swarnendu Biswas, Donald S Fussell, and Keshav Pingali. An elementary introduction to kalman filtering. *Communications of the ACM*, 62(11):122–133, 2019.

[19] Hein Putter, Jelle Goeman, and Jacco Wallinga. Stochastic epidemic models and their link with methods from survival analysis. *medRxiv*, 2024.

[20] Maria Isabel Ribeiro. Kalman and extended kalman filters: Concept, derivation and properties. *Institute for Systems and Robotics*, 43(46):3736–3741, 2004.

[21] Carlos Rodrigo and Maria Méndez. Clinical and laboratory diagnosis of influenza. *Human Vaccines & Immunotherapeutics*, 8(1):29–33, 2012.

[22] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.

[23] Cécile Viboud and Alessandro Vespignani. The future of influenza forecasts. *Proceedings of the National Academy of Sciences*, 116(8):2802–2804, 2019.