

# Manuscript Unformatted

*Craig McGowan*

*November 16, 2018*

## Abstract

## Intro

- Flu is bad - here's why
- Why forecast flu
- Summary of other methodologies
- CDC challenge
- What this paper does

## Methods

### Forecast structure

Weekly forecasts were structured according to the template used by the FluSight challenge (Biggerstaff et al 2018, McGowan et al 2019). Within each week, forecasts were generated for each of seven targets in eleven locations. Forecast locations included the United States as a whole and each of the 10 Health and Human Services Regions, which each encompasses several US states. Forecast targets could be generally grouped into seasonal targets and short-term targets. Seasonal targets included season peak percentage, defined as the maximum wILI% value, season peak week, defined as the week when the maximum wILI% occurred, and season onset, defined as the first week where the wILI% value matched or exceeded the location-specific baseline value and remained at or above for three consecutive weeks. Short-term targets included forecasts of wILI% values 1, 2, 3, and 4 weeks ahead of the most recently published data. Possible outcomes for each target are grouped into bins, and the forecasts assign a probability of occurrence to each possible outcome for each target. (Table 1)

### Model definition and terminology

Let  $y_{l,t}$  be the observed wILI% value for location  $l$  at MMWR week  $t$ . Models for all locations were of the general form

$$g_l(y_{l,t}) = a_l + \sum_{k=1}^{K_l} \left[ \alpha_{k1} \sin\left(\frac{2\pi kt}{52}\right) + \alpha_{k2} \cos\left(\frac{2\pi kt}{52}\right) \right] + \beta_l \mathbf{X}_{l,t} + N_{l,t} \quad (1)$$

where  $g_l()$  is a location-specific link function,  $K_l$  is a location-specific number of Fourier terms,  $\mathbf{X}_{l,t}$  is a matrix of observed covariates for location  $l$ , and  $N_l$  is an ARIMA error process. Separate models were fit for each forecast location, with the link function, number of Fourier terms, ARIMA error structure, and included covariates allowed to vary between locations.

Model component	Options evaluated in cross-validation
Link function $g_l()$	1. Identity 2. Natural logarithm 3. Box-Cox transformation
Number of Fourier terms $K_l$	1, 2, ..., 12
ARIMA process $N_l$	AR term: 0, 1, 2, 3 I term: 0, 1 MA term: 0, 1, 2, 3 All combinations of:
Covariates $\mathbf{X}_l$	1. National or regional Google Trends 2. Cumulative H1 and H3 subtype prevalence 3. ILI backfill

## Model fitting

Models for each location were fit using a sequential cross-validation process in R using tools from the `forecast` package. The link function  $g_l()$  was chosen first, followed by the number of Fourier terms  $K_l$ , the ARIMA error structure  $N_l$ , and covariates  $\mathbf{X}_l$ . Possible choices for  $g_l()$ ,  $K_l$ ,  $N_l$ , and  $\mathbf{X}_l$  are summarized in Table ???. Training data consisted of the 2004/2005 through 2013-2014 influenza seasons, excluding the 2008/2009 and 2009/2010 seasons where typical seasonal patterns were disrupted by the H1N1 pandemic. For models of later test seasons, data from test seasons that had already occurred were included in the training data (e.g. 2015/2016 for models of 2016/2017).

To determine  $g_l()$  for a particular test season, we used data from 2010/2011 season through the prior to that test season as cross-validation seasons. For each CV season, we fit a location-specific SARIMA model using the `auto.arima` function in the `forecast` package for each possible link function  $g_l()$  to the training data prior to the CV season. We generated weekly probabilistic forecasts for the CV season by sampling from the predicted trajectory using bootstrapped errors. Forecasts were scored using the scoring rules for the FluSight challenge (see ‘Scoring Rules’ below) and averaged across all CV seasons. The link function with the highest average score across the CV seasons was selected for each location. Values and/or structures for  $K_l$ ,  $N_l$ , and  $X_l$  were determined sequentially in an analogous fashion, incorporating decisions made for prior model components (e.g. CV models to determine  $N_l$  incorporated previously calculated values for  $g_l()$  and  $K_l$ ).

We considered several potential covariates for inclusion in the test model for each location.

## Development of ‘prospective’ forecasts

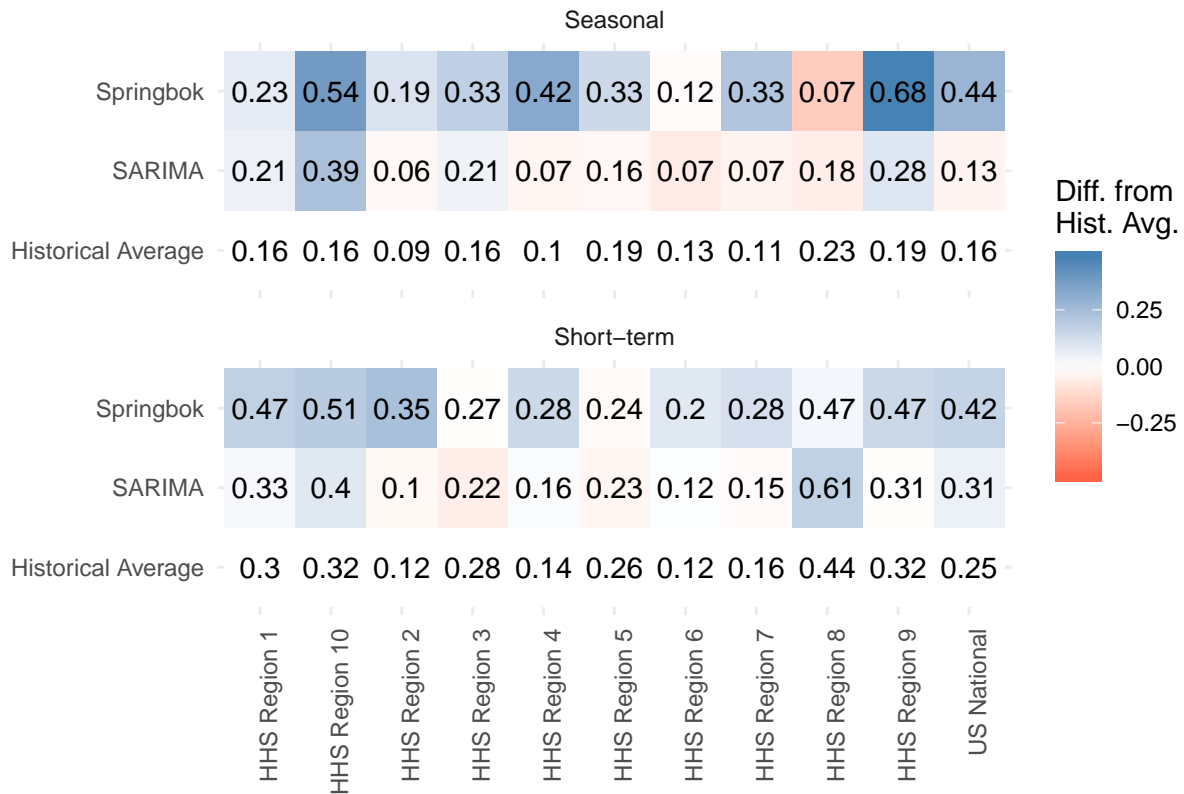
### Scoring rules

## Results

- Summary results from each test season

season	Historical Average	SARIMA	Springbok
2014/2015	0.2296	0.2591	0.3182
2015/2016	0.2225	0.1284	NA
2016/2017	0.214	0.2397	NA
2017/2018	0.118	0.1531	NA

- Comparisons across seasons and locations and targets



- Compare to seasonal ARIMA forecast - illustrate improvement over baseline model
- Check Nick's papers for potential comparisons and graphs that could be made
- Comparisons to CDC results using relevant scoring rules

## Discussion

- New modeling approach compared to previous literature
- Strong performance compared to existing models
- Submitted as true prospective model during 18/19 season and can evaluate further then (slash maybe wait until season is over)
- Strengths - fairly simple mathematical model, easy to implement with existing software
- Limitations - current analysis assumes Google Trend data are