

Predicting Disease Outbreaks Using Ensemble Models

Jordan Aron, Will Thompson, Dean Gladish



Introduction

Reich Labs is an organization at the University of Massachusetts specializing in mathematical modeling of diseases—particularly Pandemic Influenza in the United States.

Their recent work has been focused on modeling disease in real time. Reich Labs has been working in tandem with several other agencies and universities to collect data produced from multiple different modeling techniques, and aggregate the results to create a visualization of the spread of Influenza on a weekly basis.

Methods

Ensemble modeling is the process of running multiple analytical models and then synthesizing the results into a single score. Ensemble models are especially useful when modeling noisy data, a common theme when dealing with public health prediction. Some models are most accurate at different points in the season, thus by varying the importance of each individual model in the ensemble, we can maximize accuracy. The following are the components of the ensemble model.

KDE

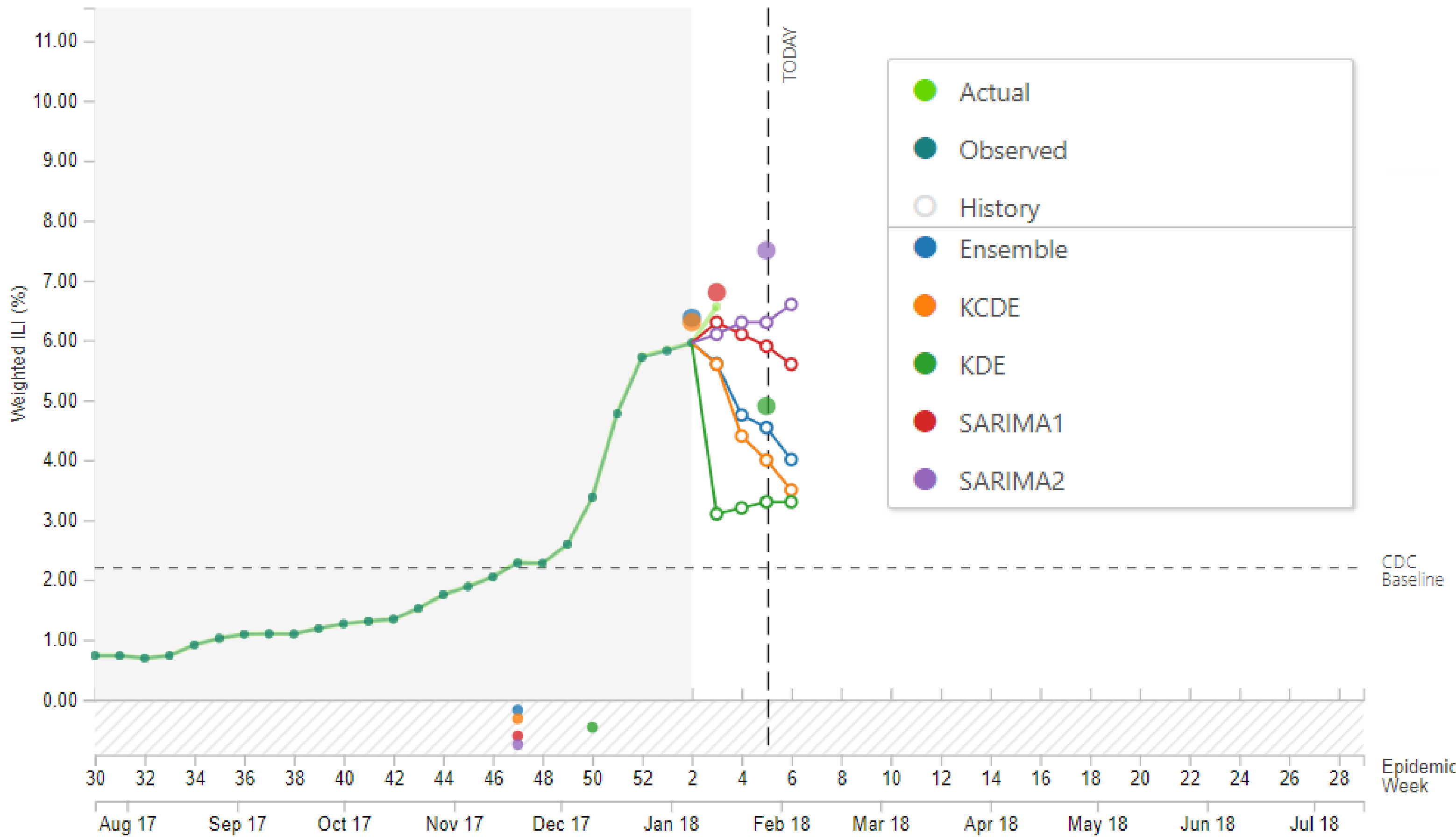
- Kernel Density Estimation is a non-parametric way of estimating a probability density function of a random variable
- It does not get updated as the season progresses

KCDE

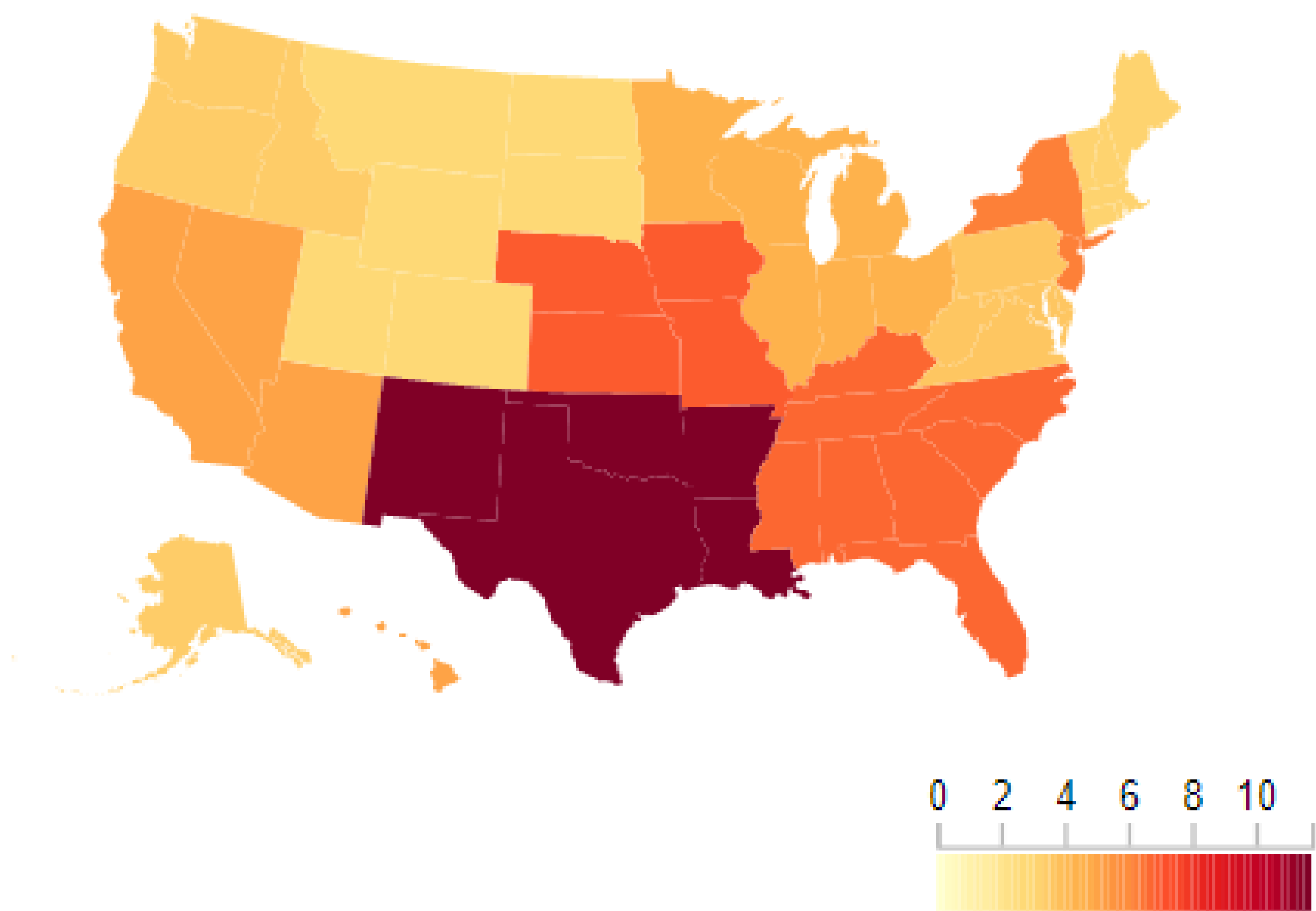
- Kernel Conditional Density Estimation acts similarly to KDE, however it incorporates past weeks values
- Developed specifically to predict infectious disease outbreaks

SARIMA

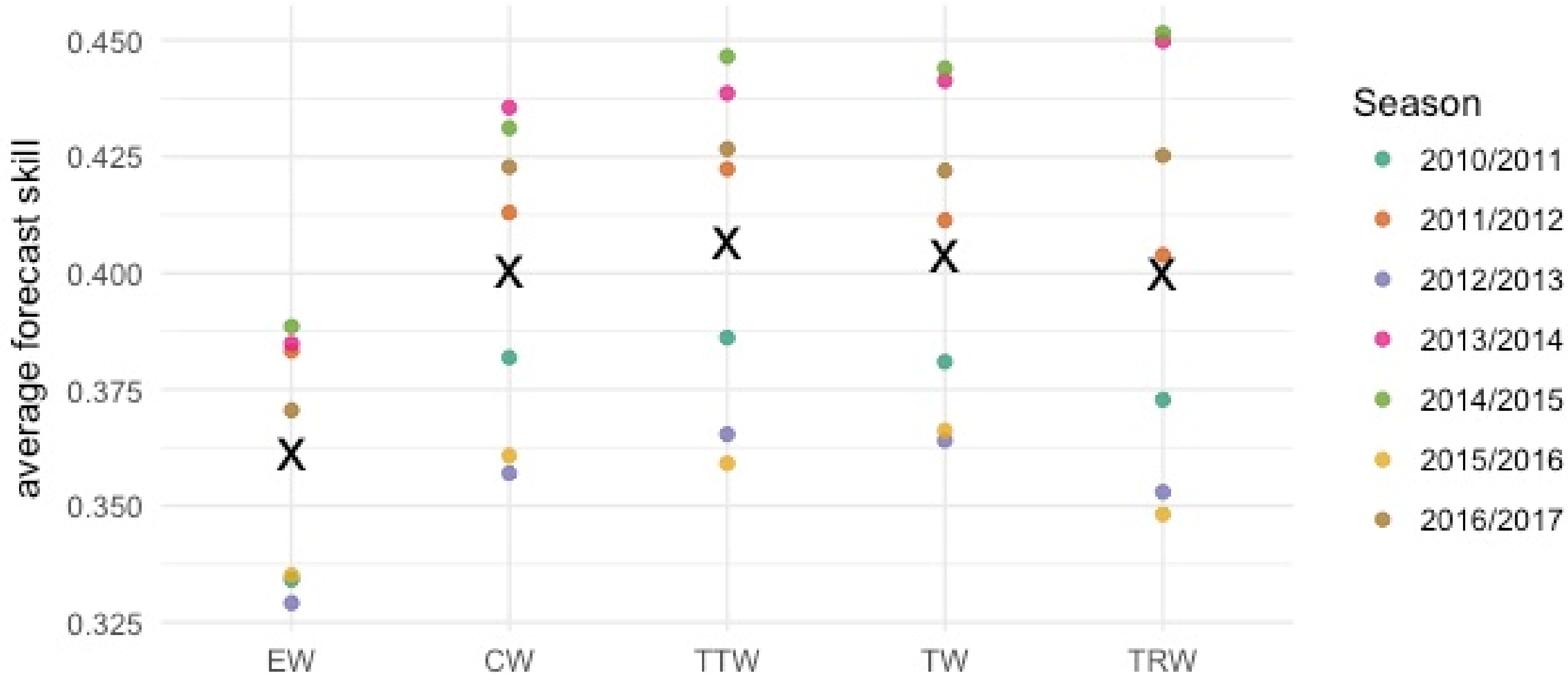
- This method is comprised of four parts: seasonal (S), auto-regressive (AR), integrated (I) and moving-average (MA).
- Two SARIMA models were included in the ensemble, one with an auto regressive order of 1 and the other with an order of 2.



This figure depicts the flusight app that the Reich lab has created. It gives the current wILI as well as predictions for each individual model and the ensemble. The colored circles on the graph represents the predicted peak incidence and the circles on the bottom give the predicted season onset. This data is for January 25th, 2018.



This is a map of the current flu outbreak in the United States. The scale given refers to the wILI of each state. This is the corresponding map to the above figure. Again this data is from January 25th, 2018.



This figure depicts the forecast efficacy of different weighting systems. From left to right, each model demonstrates increasing numbers of weights.

Results

We had many results. We were introduced to model stacking, predictive density networks, regression analysis, and other mathematical methods of inference. In addition to discussing the limitations of induction, we gained valuable knowledge of graduate-level math work.

Daily work and Learning Experience

While at Reich Labs, we participated in weekly lab meetings, attended several lectures and seminars, and our supervisor introduced us to other Biostatisticians and graduate students. Together, we learned more and more about the field of Epidemiology and modeling, but also about the lives of Biostatistics and the daily work and routine for Ph. D students. The exposure to the full picture of being a researcher in an academic setting was invaluable. This experience helped all of us consider new fields in mathematics, as well as the what to expect from graduate school in a mathematical field.

Acknowledgments

This work was made possible by Brian, Ashley, and the Career Center. We would also like to thank all the grad students and the Reich Lab for telling us about how to conduct research.