Brief review

Shaman, J., Karspeck, A., Yang, W., Tamerius, J. and Lipsitch, M., 2013. Real-time influenza forecasts during the 2012–2013 season. *Nature communications*, *4*, p.2837.

The timing of seasonal influenza epidemic in temporal regions contains high variances. The public health system in the US has not yet adopt a robust forecast system to predict the peak timing of seasonal influenza accurately and in a timely fashion. Due to the inaccurate forecasts, it is difficult for regional health practitioners to improve the epidemic preparedness and responses. To address these issues, Lipstich et al describes an ensemble model (SIRS-EAKF) to predict the peak week of seasonal influenza. Two foundational concepts of influenza modeling have been utilized: SIRS model and the incorporation of absolute humidity (AH). The model was calibrated based on an early finding, the relationship between ensemble spread and the accuracy. This calibration process is similar to the hyperparameter tuning of machine learning models. In this study, basic reproductive number (R_0) was treated as a hyperparameter.

It is worth to point out that this study is one of the earliest to adopt the methodology of machine learning, by using retrospective data for model training and real-time forecasts (CDC and GFT) for data assimilation (model retraining). In addition, the metric for seasonal influenza incidence estimation (GFT estimation multiplied by CDC prevalence under bayes rule) is an innovation approach. However, the accuracy of the forecasts is still questionable. It is speculative to compare this model with a relative primitive resampling method rather than more advanced methods, only to shed light on the slightly more robust performance. Another limitation that the author recognized is that the performance of the forecasts is not consistent among all the cities.

Data sparsity and incompleteness are two major issue in epidemiological data. Kinsa's technology will provide a more comprehensive dataset to help to reduce data incompleteness and selection bias. GPS information will enhance the granularity of prediction. Furthermore, the real-time datastream produced by the device can help increase the model retraining efficiency.