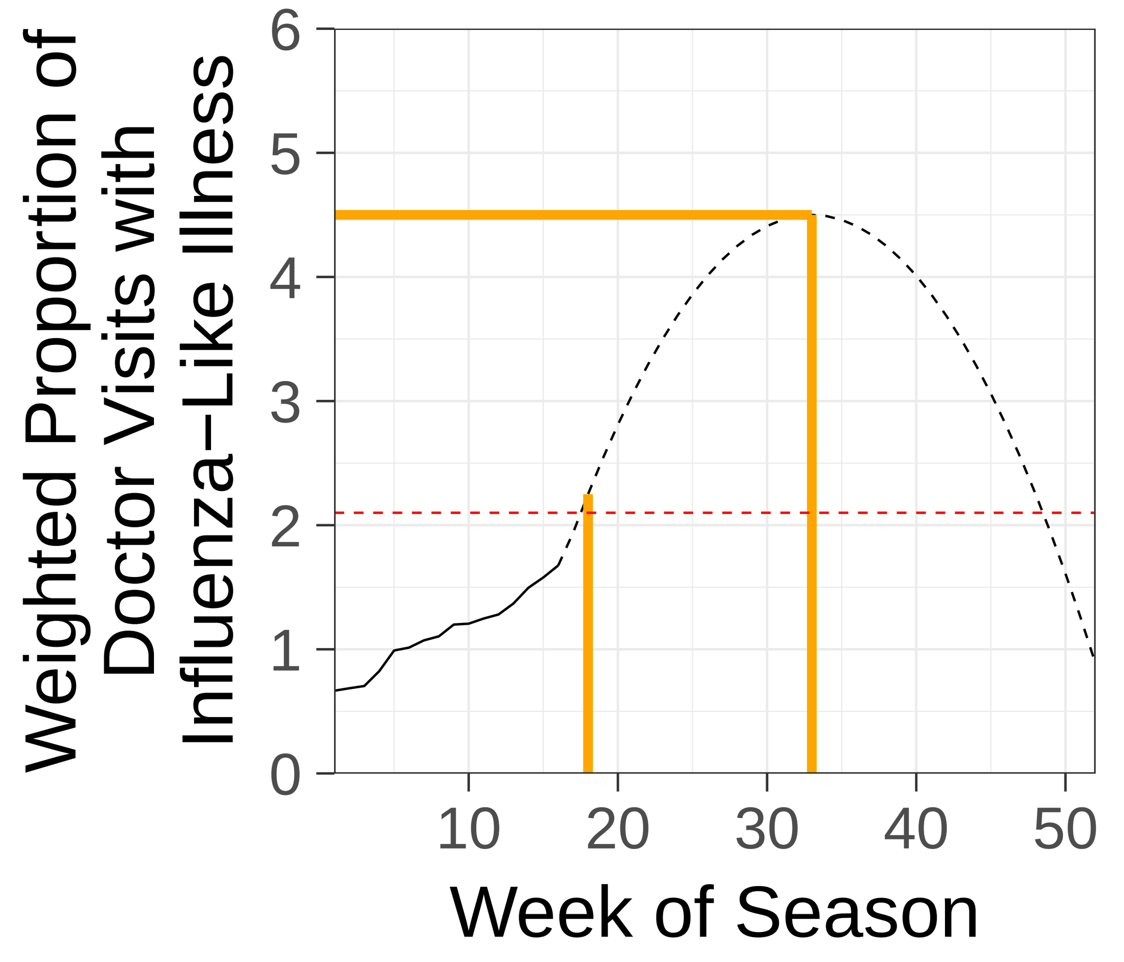
Model Averaging for Probabilistic Time Series Forecasts

Evan L. Ray, Nicholas G. Reich

University of Massachusetts, Amherst

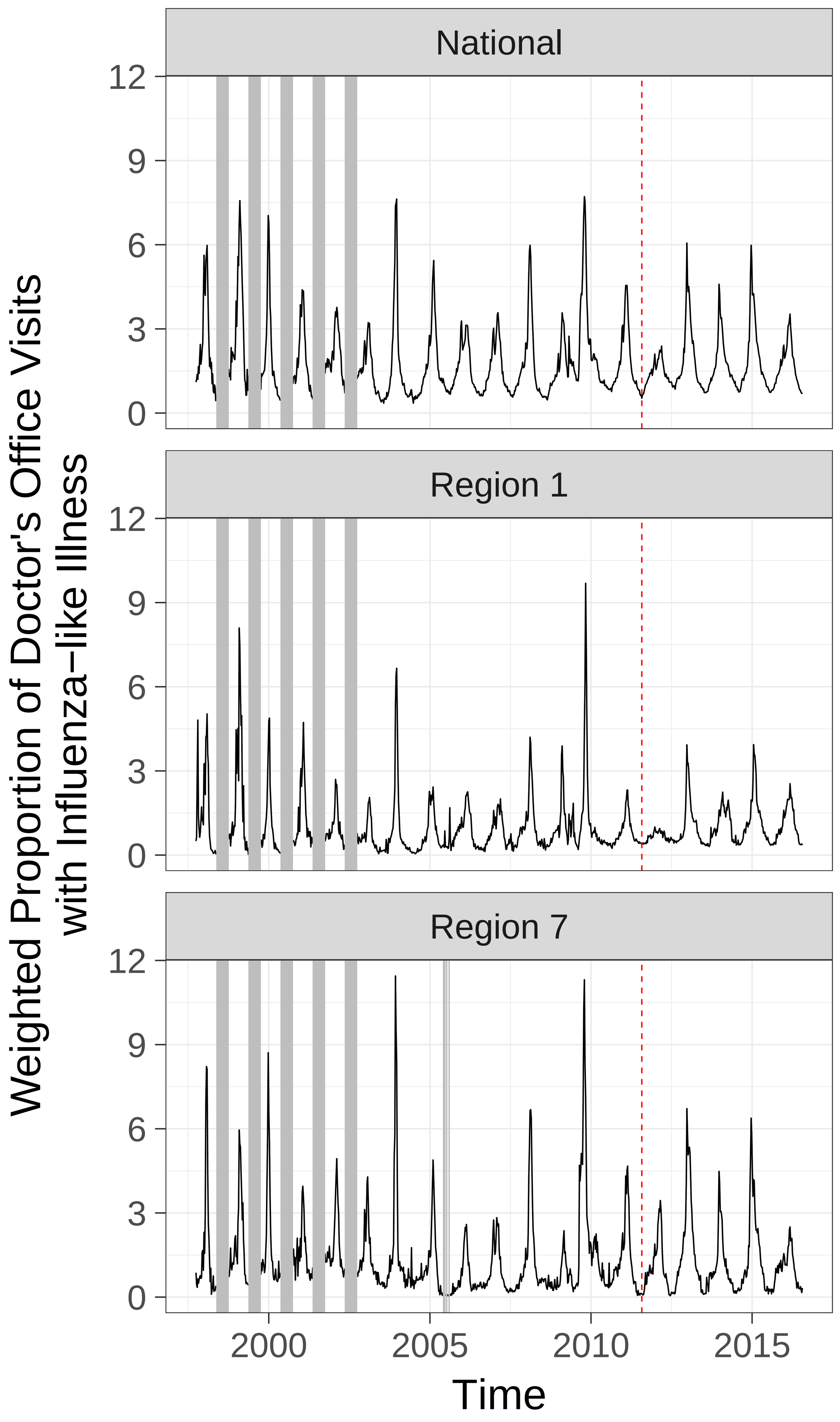
Introduction

- Predictions of infectious disease are helpful to public health decision makers.
- We focus on three quantities:
  1. timing of season onset
  2. timing of season peak week
  3. incidence in season peak week



- Predictions updated weekly.
- Relative performance of different models varies.
- Can we combine predictions from these models to improve performance?

Data



Component Models

- Our ensembles combine predictions from three component models:
  1. Kernel Density Estimation (KDE)
     - separate distribution estimates for each target based on observed values in training-phase seasons
     - predictions do not change over the season
  2. Kernel Conditional Density Estimation (KCDE) with Copulas
     - KCDE: separate predictive distributions for flu incidence in each future week of the season given recent observations of wILI and the current week of the season.
     - Copula: model dependence among incidence in different weeks
     - Get predictive distributions for onset timing, peak timing, and peak incidence as integrals of joint distribution for incidence in all remaining weeks
  3. Seasonal Auto-Regressive Integrated Moving Average (SARIMA)
     - Log-transform wILI, first-order seasonal differencing
     - Integrate to obtain predictive distributions for onset timing, peak timing, and peak incidence as with KCDE

This work was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health (grants R21AI115173, R01AI102939, and R35GM119582).

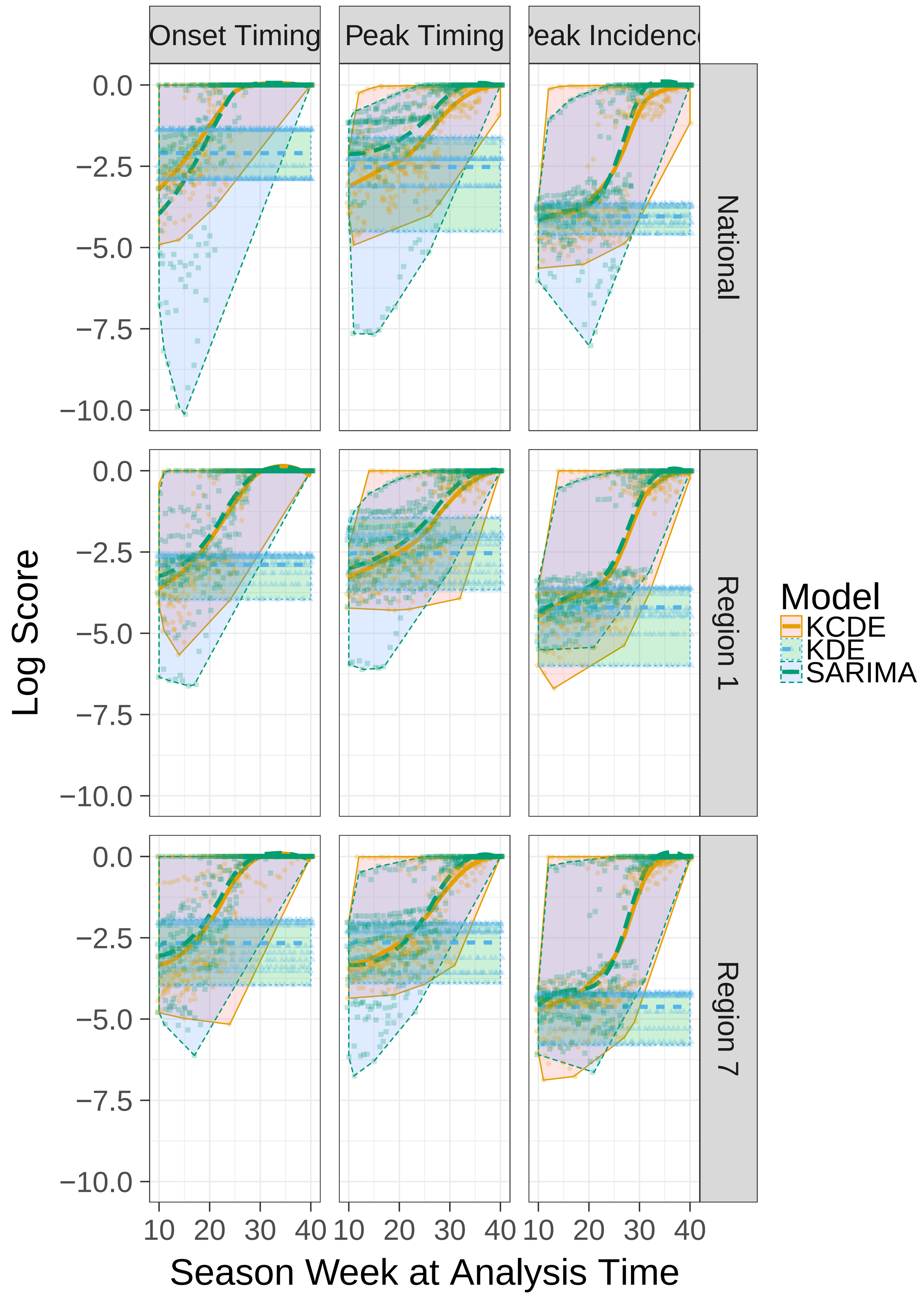
Model Averaging for Probabilistic Time Series Forecasts

Evan L. Ray, Nicholas G. Reich

University of Massachusetts, Amherst

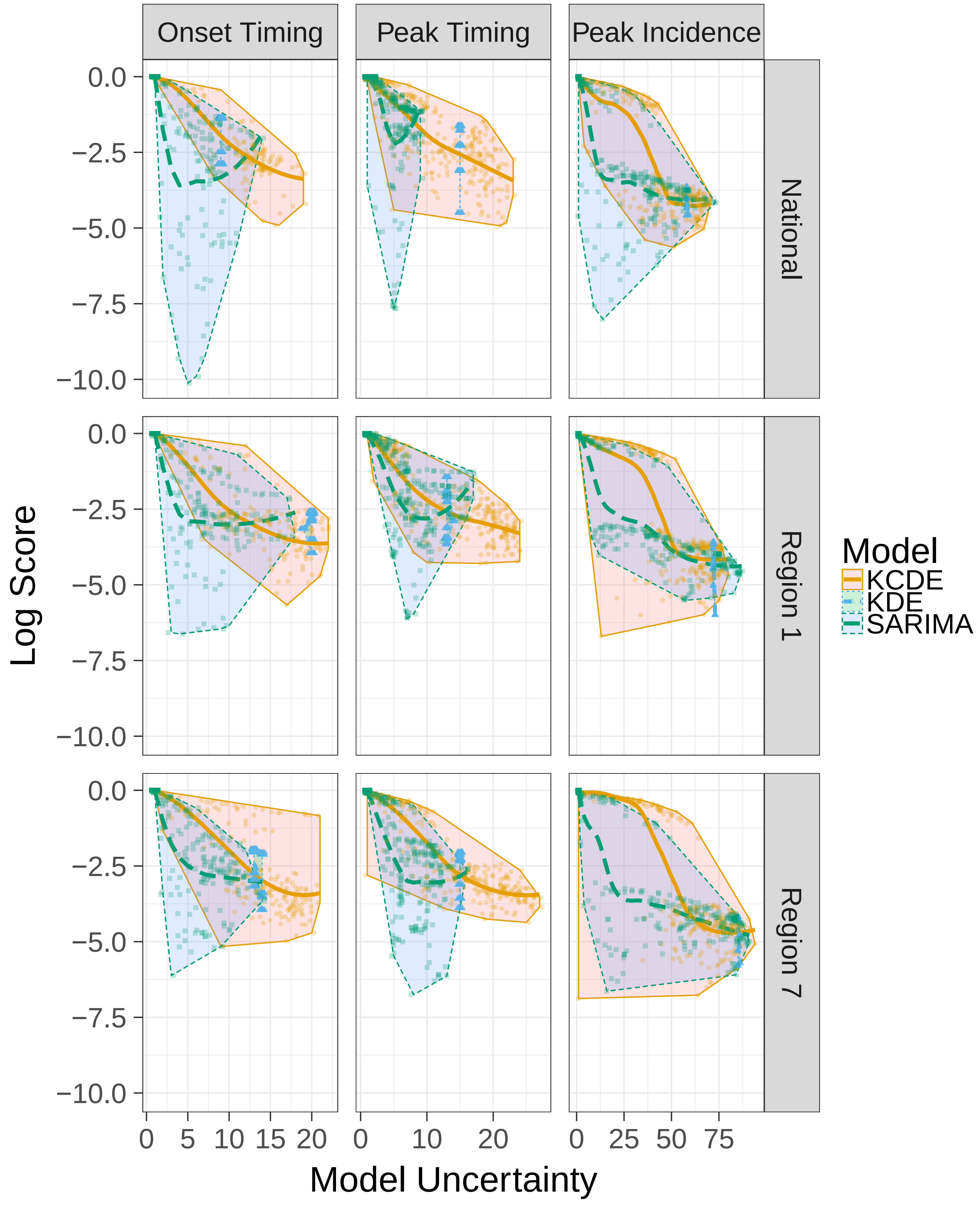
Performance Varies with Week

Log Scores vs.
Season Week at Analysis Time



Performance Varies with Uncertainty

Log Scores vs. Model Uncertainty



Ensemble Models

- The ensemble models we consider are *weighted averages* of the component models:

$$f(y|\mathbf{x}) = \Sigma_{m=1}^{M} \pi_m(\mathbf{x}) f_m(y|\mathbf{x}), \text{ where}$$

  - y is a possible value for one of the prediction targets
  - $\mathbf{x}$ is a vector of covariates such as recent incidence, time of year at which we are making the predictions, ...
  - $f_m(\cdot)$ are predictive distributions from $M = 3$ component models
  - $\pi_m(\mathbf{x})$ are model weights with $\Sigma_{m=1}^{M} \pi_m(\mathbf{x}) = 1 \, \forall \mathbf{x}$
- We consider six variations:

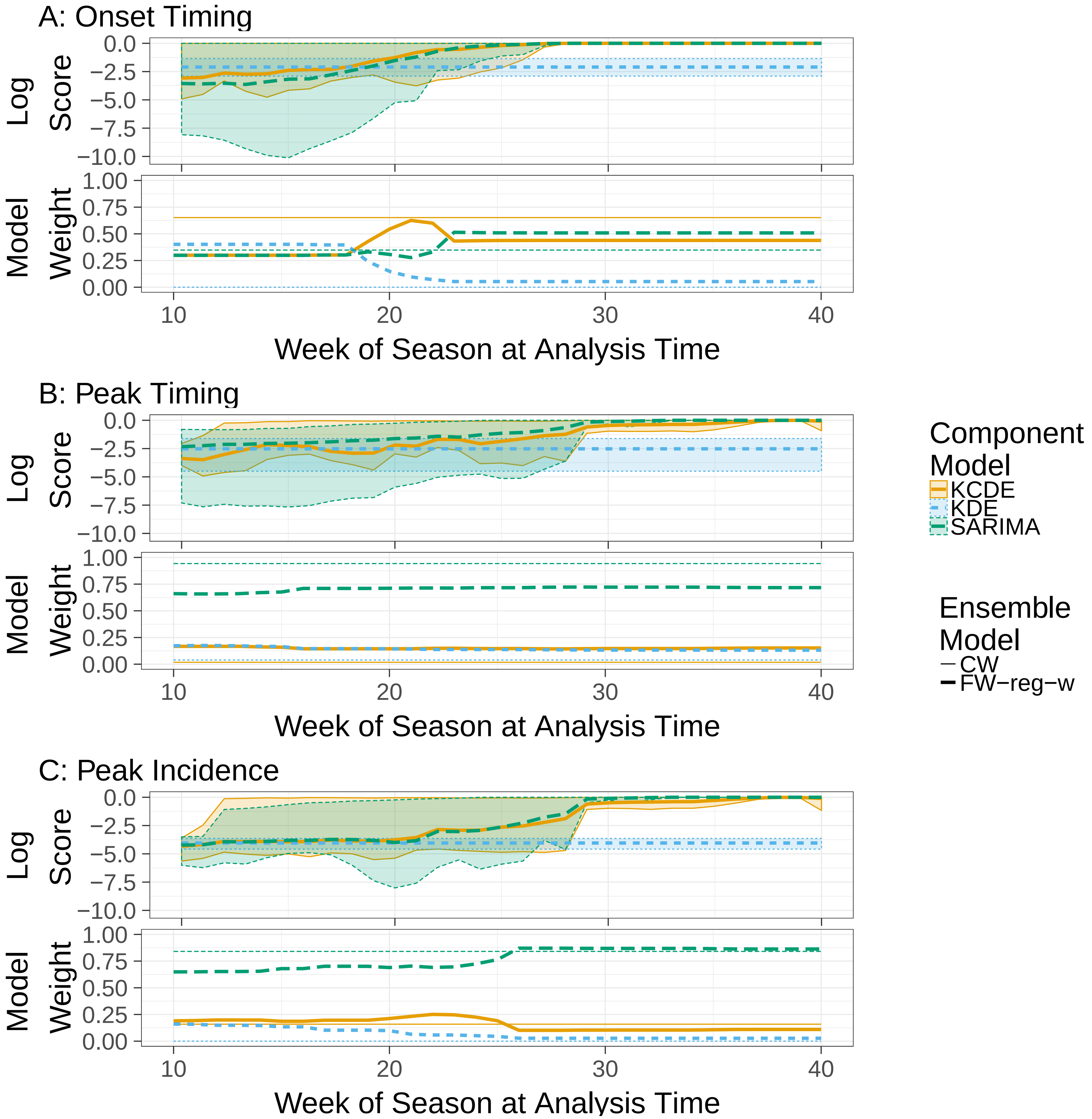
	Component Model Weights Vary with...					
Model	Region	Prediction Target	Week of Season	SARIMA Uncertainty	KCDE Uncertainty	Current wILI
EW						
CW	X	X				
FW	X	X	X	X	X	
FW-reg-w	X	X	X			
FW-reg-wu	X	X	X	X	X	
FW-reg-wui	X	X	X	X	X	X

- Weighting functions estimated via gradient tree boosting

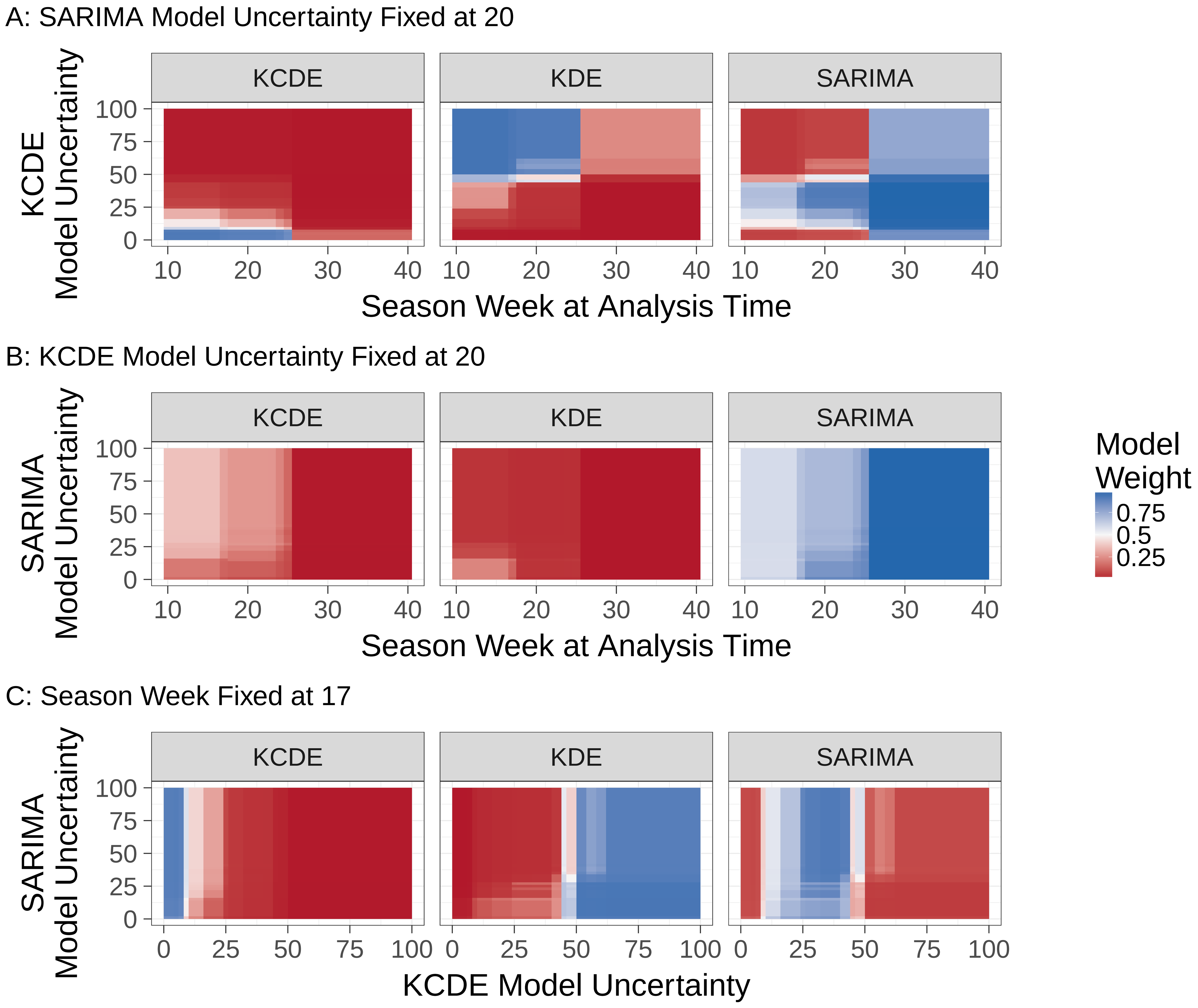
Model Averaging for Probabilistic Time Series Forecasts

Evan L. Ray, Nicholas G. Reich
University of Massachusetts, Amherst

Weights as a Function of Week



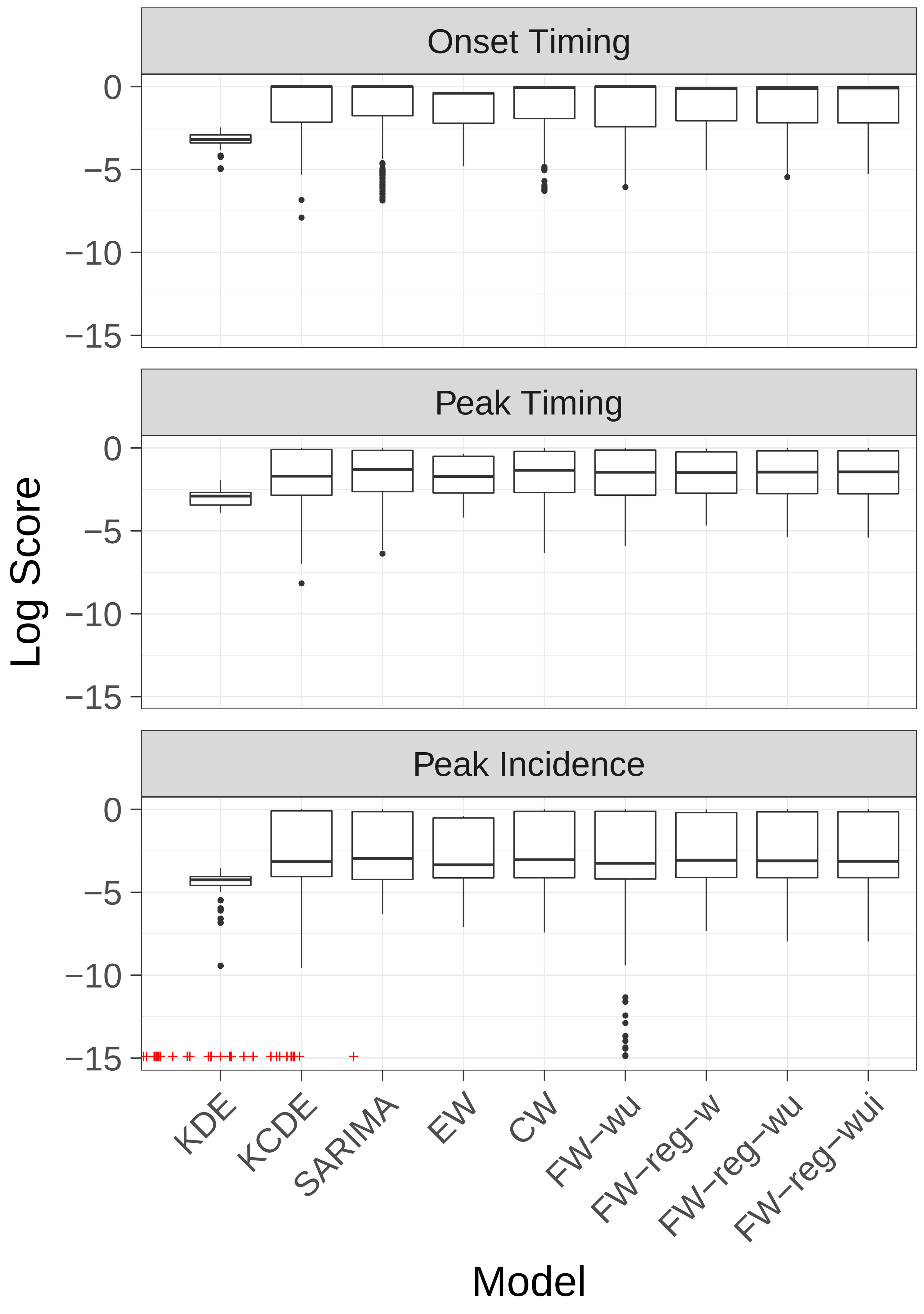
Weights as a Function of Week & Uncertainty



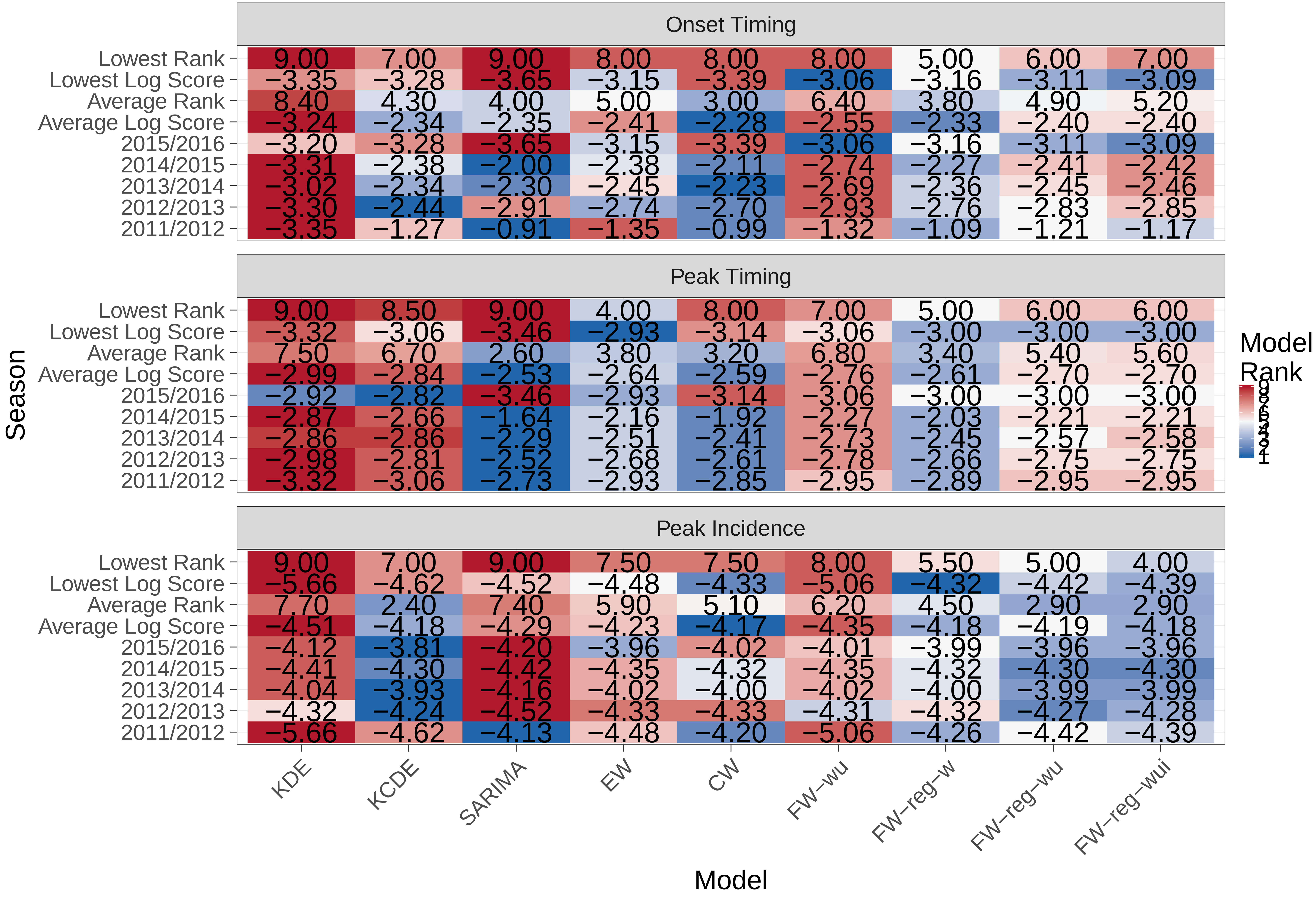
Model Averaging for Probabilistic Time Series Forecasts

Evan L. Ray, Nicholas G. Reich
University of Massachusetts, Amherst

Aggregated Results – All Regions and Test Phase Seasons



Results By Test Phase Season

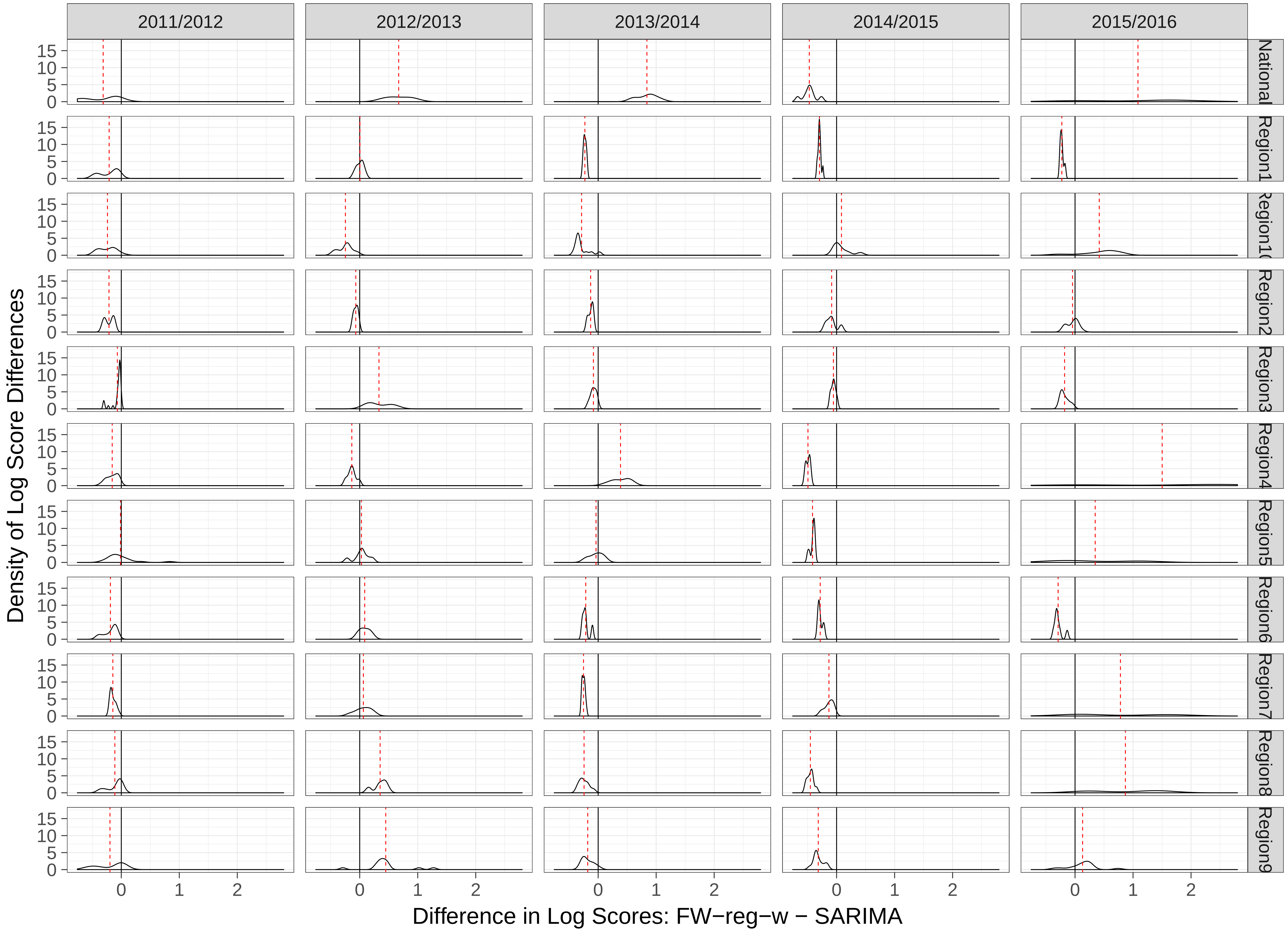


Model Averaging for Probabilistic Time Series Forecasts

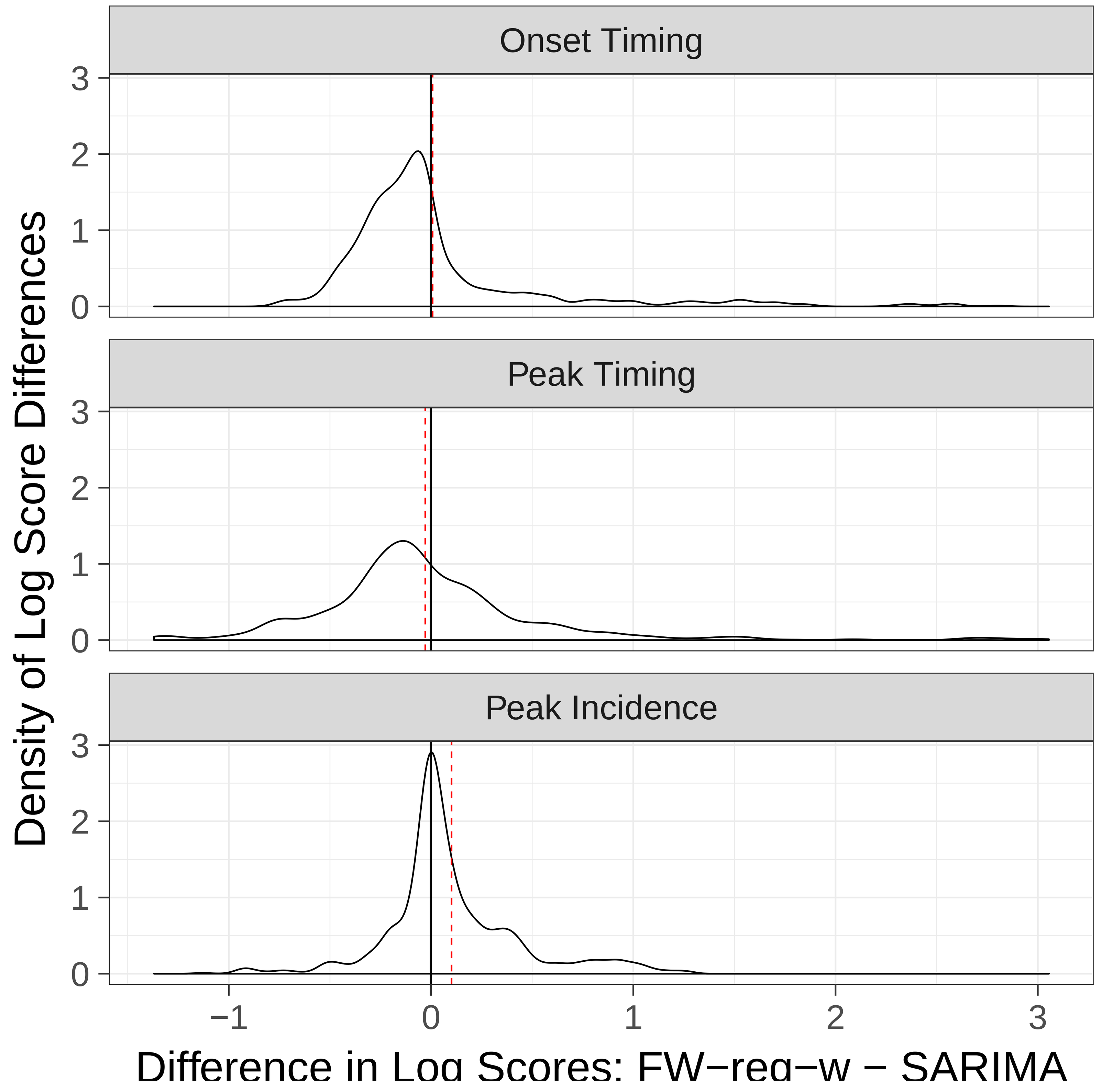
Evan L. Ray, Nicholas G. Reich

University of Massachusetts, Amherst

Log Score Differences for Onset Timing via SARIMA and FW-reg-w



Log Score Differences for SARIMA and FW-reg-w



Conclusions

- Ensemble methods had similar performance as the best of the component models in aggregate
- Ensemble methods had more stable performance across region-seasons than the component models
- In future work, would benefit from using a more diverse set of component models.