

# Supplement to Prediction of Infectious Disease Epidemics via Feature-Weighted Density Ensembles

*Evan L Ray, Nicholas G Reich*

*March 2017*

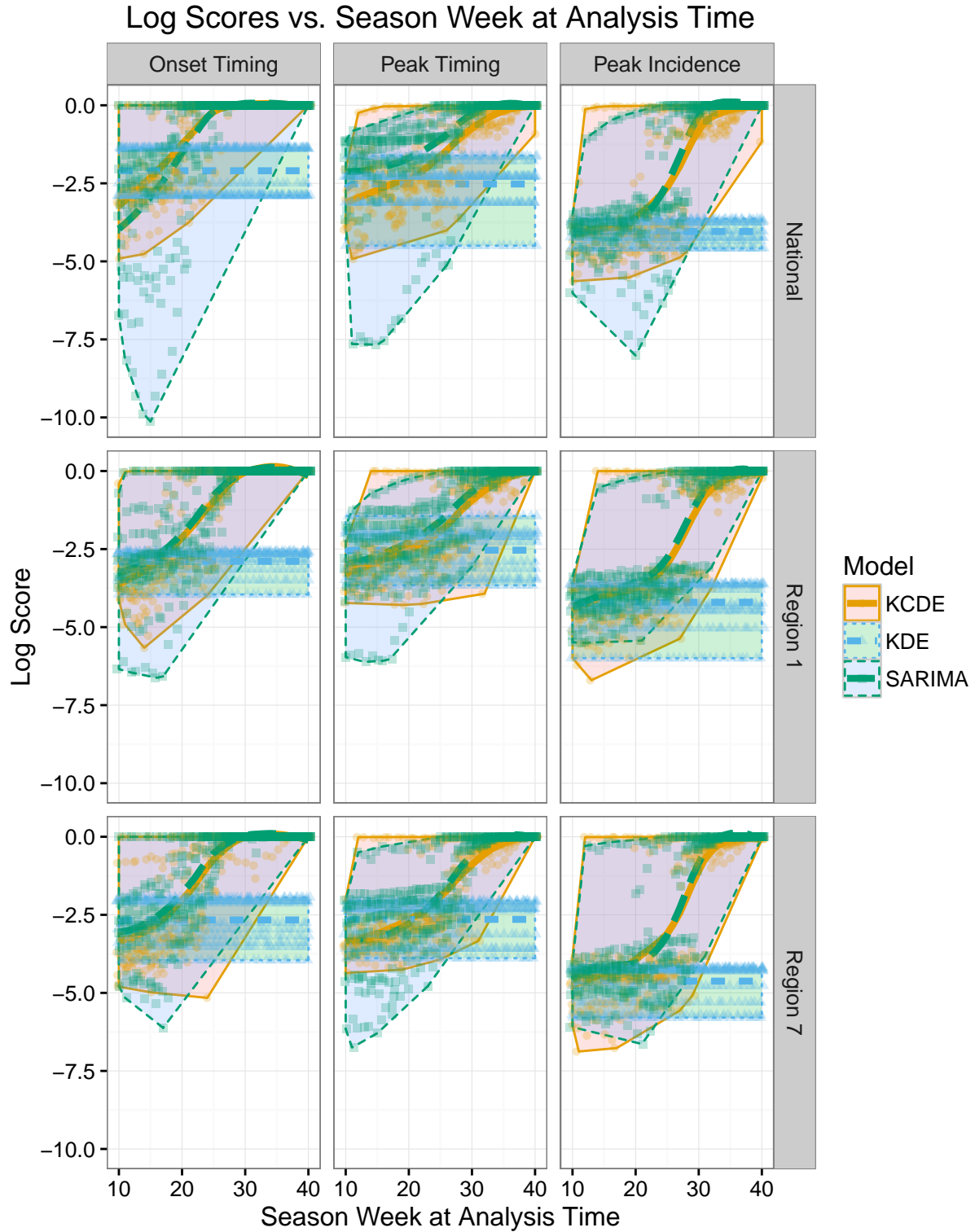
In this supplement, we include additional figures and results.

## Component Model Log Scores and Weighting Features

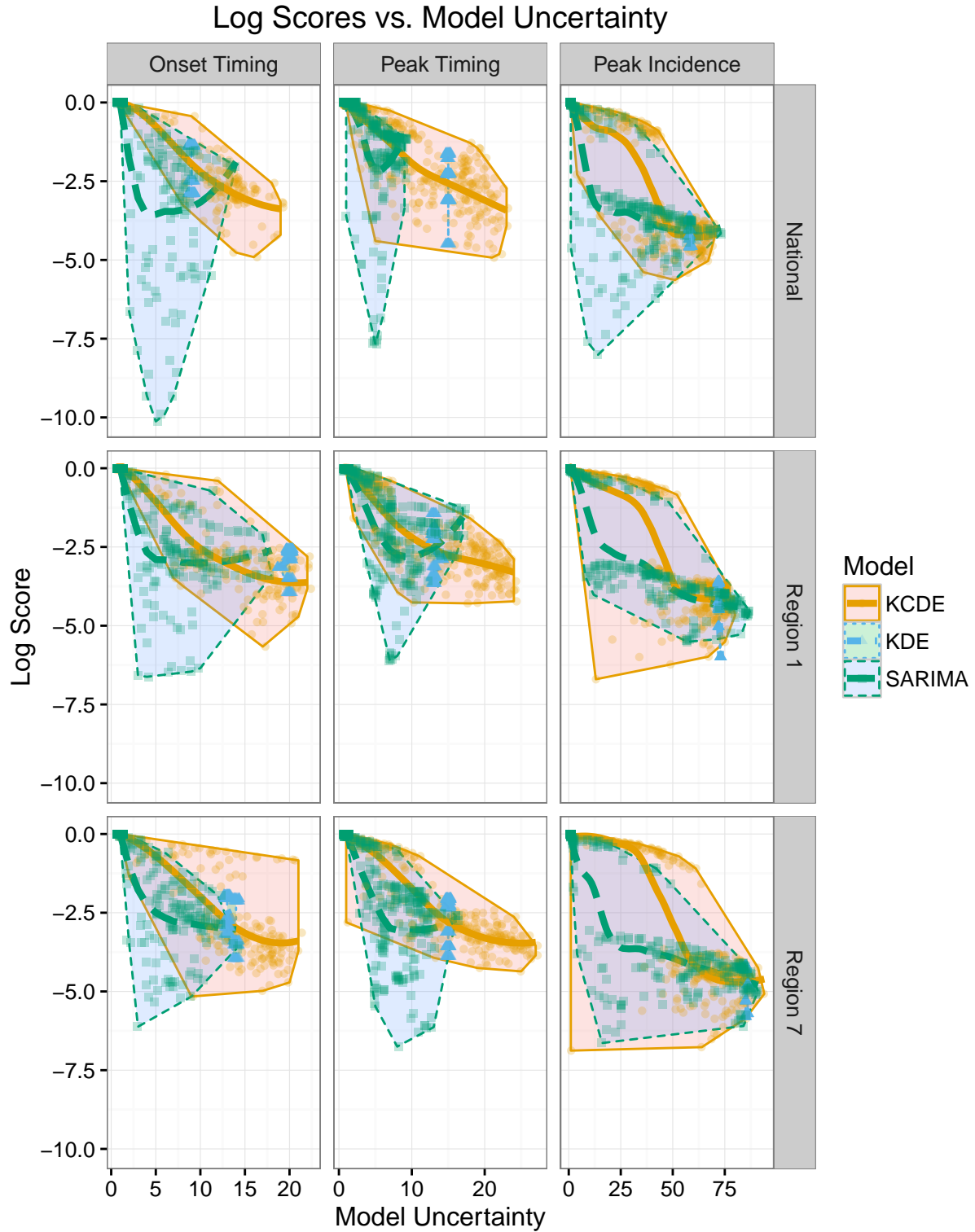
Supplemental Figs 1, 2, and 3 illustrate the relationship between log scores and weighting features for predictions from the three component models made during the training phase in weeks before the season onset (for predictions of onset timing) or the season peak (for predictions of peak timing or peak incidence). Supplemental Fig 4 represents how the model weights vary as a function of the season week and model uncertainty features for the **FW-reg-wu** model for the example of predicting peak season incidence at the national level.

## Estimated Mean Model Performance

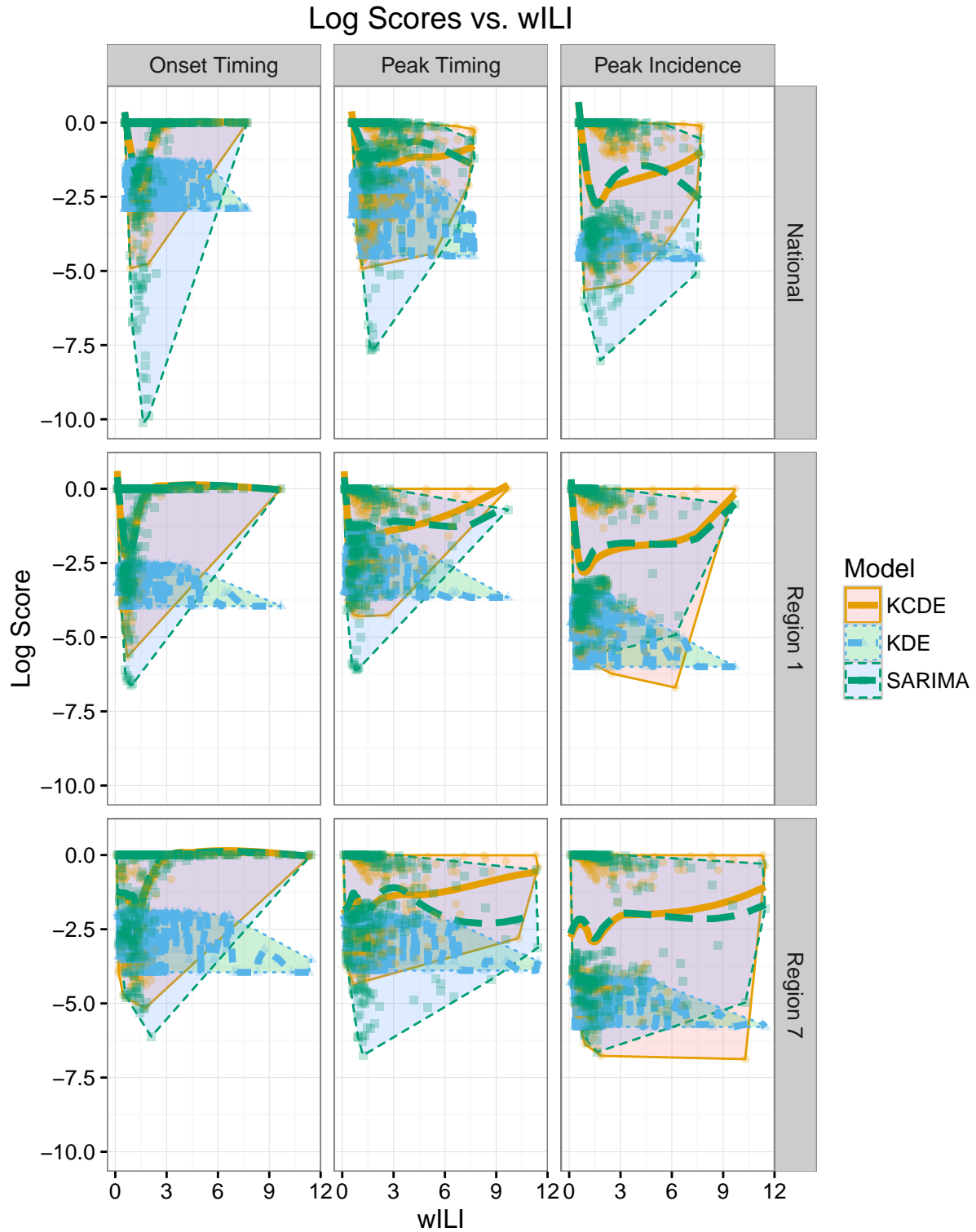
Supplemental Fig 5 displays estimated mean log scores for each model and ensemble method in predictions made before the season onset or season peak. Supplemental Fig 6, 7, and 8 show estimates for the difference in mean log scores between the model with the highest estimated mean log score for each target and each other model. The point and 95% interval estimates are obtained from a mixed effects model with a separate fixed effect mean for the interaction of model and prediction target; random effects for each combination of region, season, model, and prediction target; and lag 1 autocorrelation nested within each combination of region, season, model, and prediction target. For predictions of onset timing, the only difference in model performance that is statistically significant is the difference between **CW** and **KDE**. For predictions of peak timing, the only difference that is statistically significant is between **SARIMA** and **KDE**. For predictions of peak incidence, none of the differences between the models are statistically significant.



Supplemental Figure 1: Log scores achieved by each component model in each week of the season, summarizing across all seasons in both the training phase when all three component models produced predictions. The thick line is a smoothed estimate of mean log score at each week in the season; the shaded region indicates the convex hull of log scores achieved by each model; and the actual log scores achieved in each week are indicated with points.

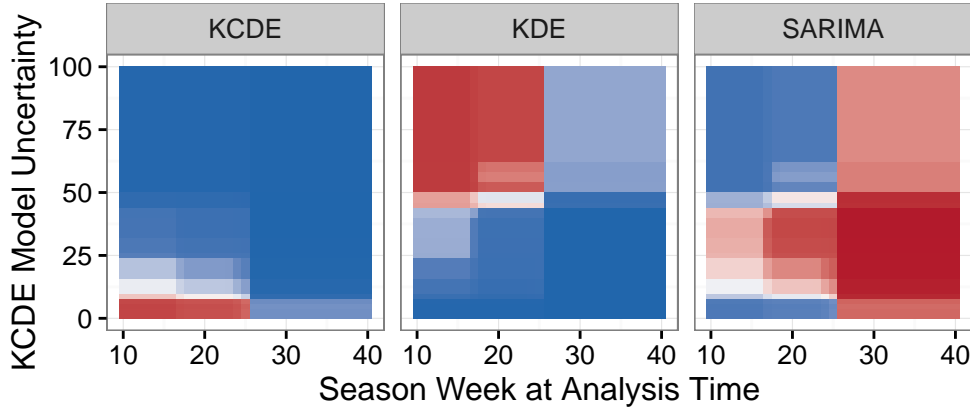


Supplemental Figure 2: Log scores achieved by each component model vs. model uncertainty as measured by the number of bins required to cover 90% of the predictive distribution. The plot summarizes results across all seasons in the training phase when all three component models produced predictions. The thick line is a smoothed estimate of mean log score at each week in the season; the shaded region indicates the convex hull of log scores achieved by each model; and the actual log scores achieved in each week are indicated with points.

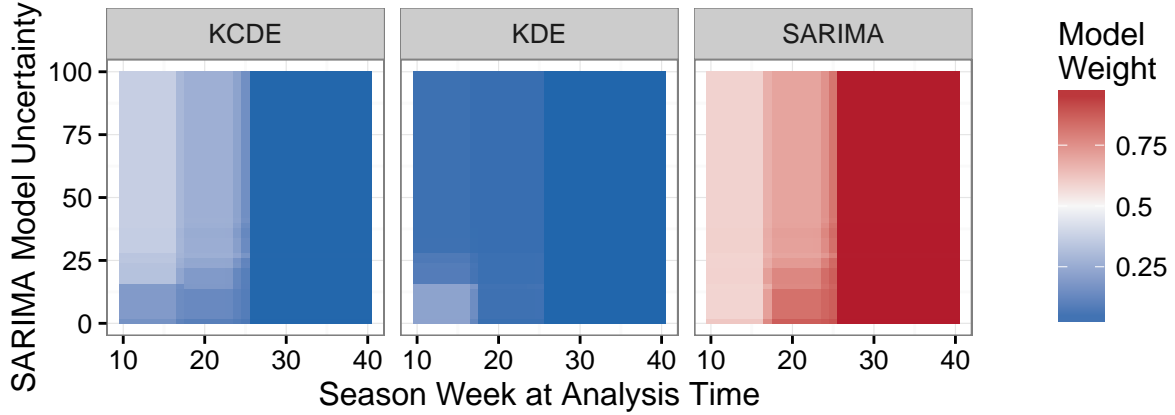


Supplemental Figure 3: Log scores achieved by each component model vs{.} wILI in the week of the season when predictions were made. The plot summarizes results across all seasons in the training phase when all three component models produced predictions. The thick line is a smoothed estimate of mean log score at each week in the season; the shaded region indicates the convex hull of log scores achieved by each model; and the actual log scores achieved in each week are indicated with points.

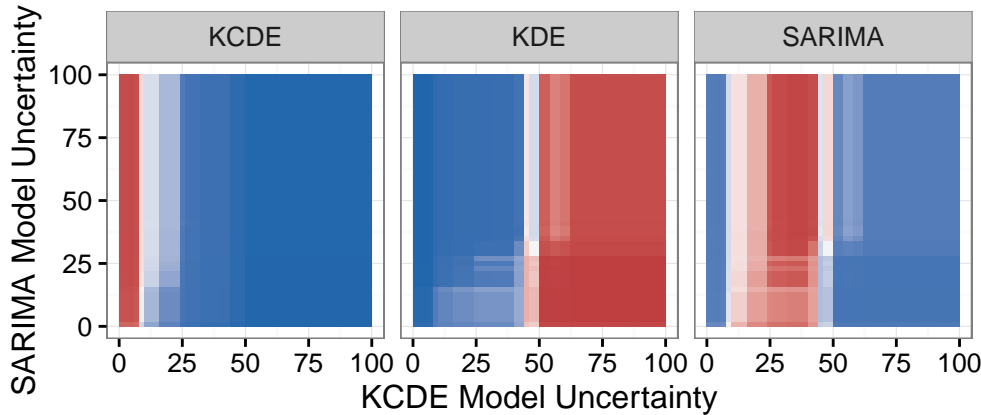
A: Component Model Weights Varying Season Week and KCDE Model Uncertainty  
SARIMA Model Uncertainty Fixed at 20



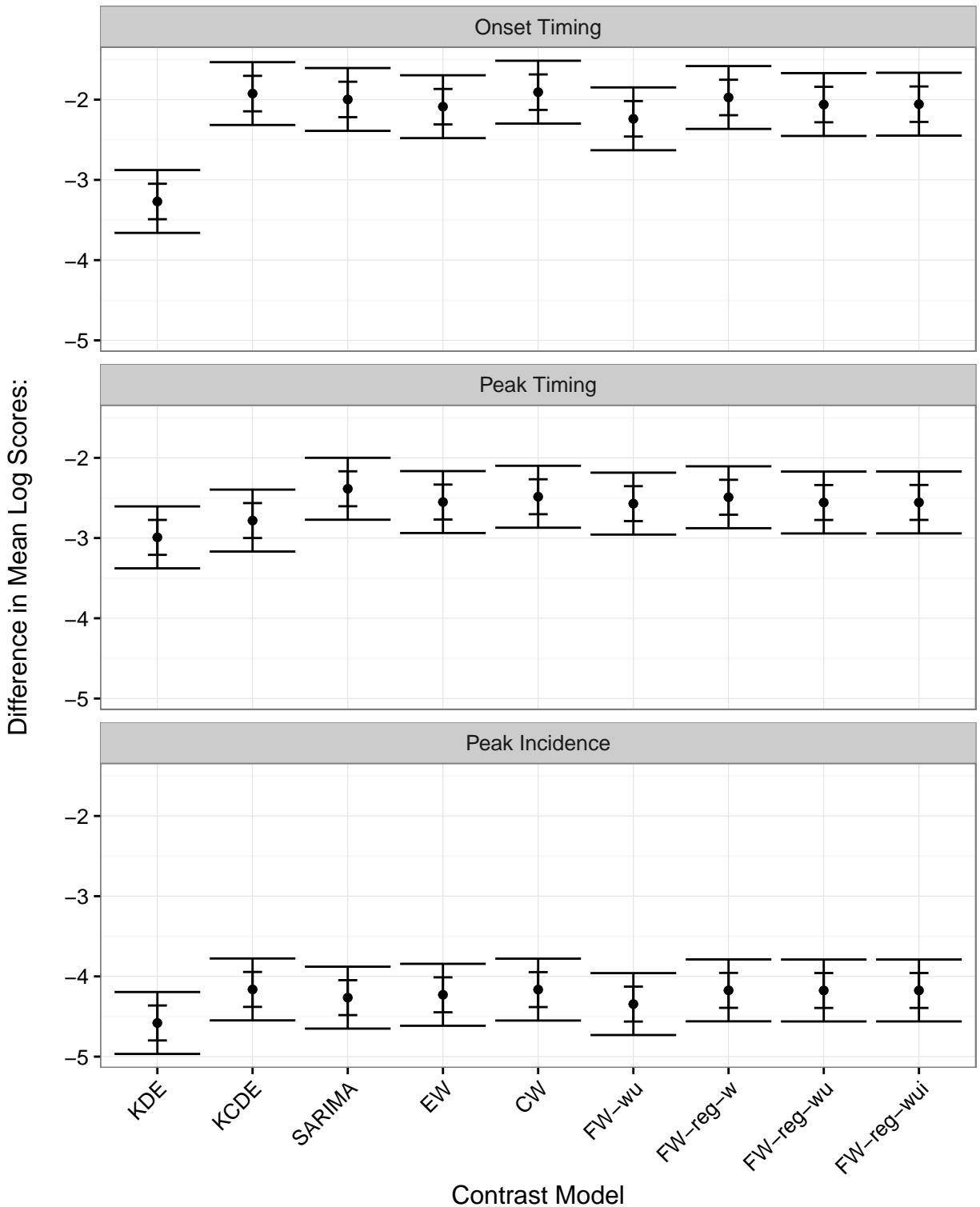
B: Component Model Weights Varying Season Week and SARIMA Model Uncertainty  
KCDE Model Uncertainty Fixed at 20



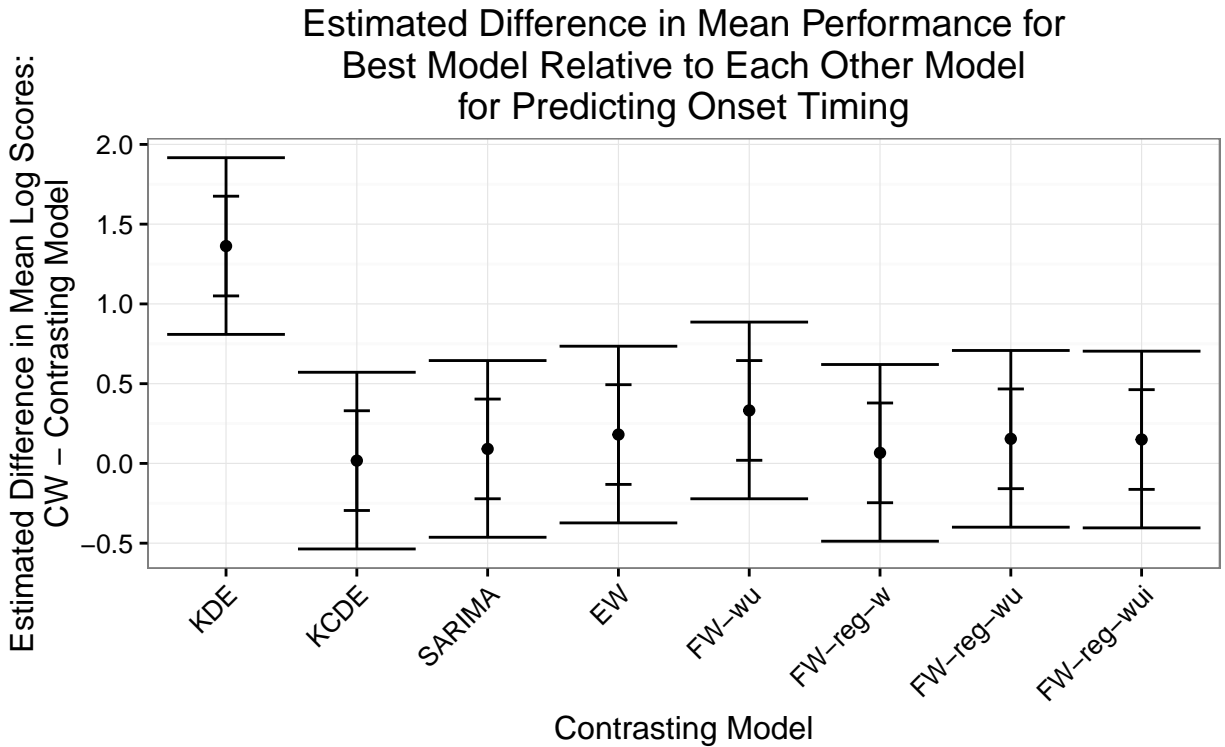
C: Component Model Weights Varying KCDE and SARIMA Model Uncertainty  
Season Week Fixed at 17



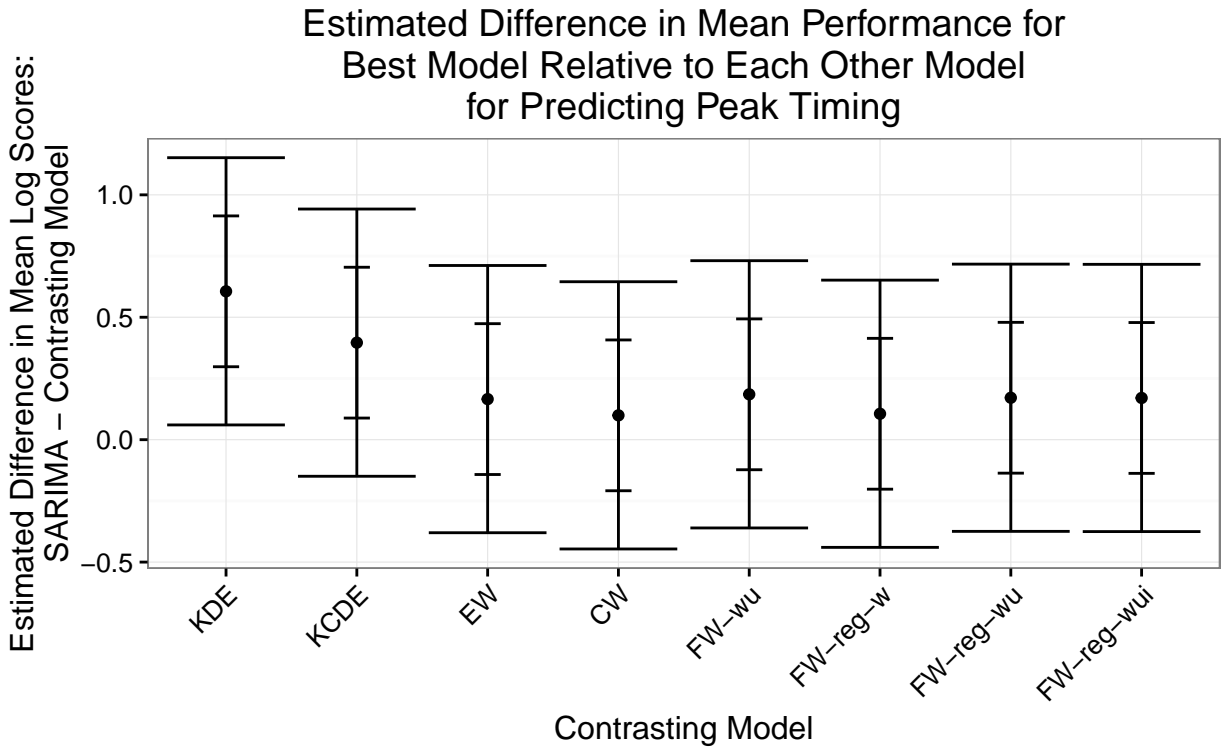
Supplemental Figure 4: Weights assigned to each component model by the **FW-reg-wu** model for prediction of season peak incidence at the national level. There are three weighting functions (one for each component model), and these are functions of three features: the week of the season at the time when the predictions are made, KCDE model uncertainty, and SARIMA model uncertainty. Each panel (A, B, and C) shows how the component model weights vary with two of these features, holding the third fixed at a “typical” value.



Supplemental Figure 5: Point estimates and confidence intervals for mean log score for each model in weeks before the target (onset or peak) occurred. Estimates are obtained from a mixed effects model with a separate fixed effect mean for the interaction of model and prediction target; random effects for each combination of region, season, model, and prediction target; and lag 1 autocorrelation nested within each combination of region, season, model, and prediction target. The wider confidence interval bounds are simultaneous confidence intervals with an approximate familywise 95% coverage rate for all intervals. The inner confidence intervals are calculated separately, with approximate coverage rates of 95%. Log scores of -Infinity were truncated at -10 before fitting this model.

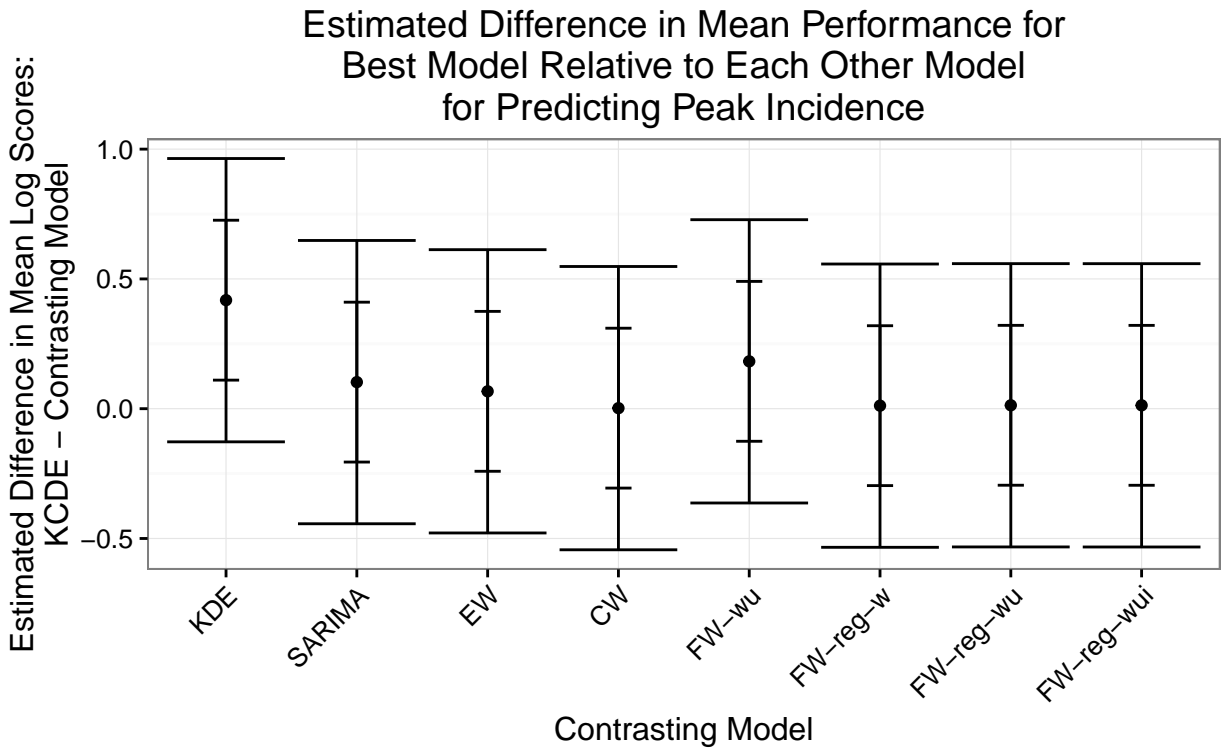


Supplemental Figure 6: Point estimates and confidence intervals for the difference in mean log score between the best model and each other model for predictions of season onset timing made before the season onset occurred. Estimates are obtained from a mixed effects model with a separate fixed effect mean for the interaction of model and prediction target; random effects for each combination of region, season, model, and prediction target; and lag 1 autocorrelation nested within each combination of region, season, model, and prediction target. The wider confidence interval bounds are simultaneous confidence intervals with an approximate familywise 95% coverage rate for all individual model means and all pairwise model contrasts for all prediction targets. The inner confidence intervals are calculated separately, with approximate coverage rates of 95%. Log scores of -Infinity were truncated at -10 before fitting this model.



Supplemental Figure 7: Point estimates and confidence intervals for the difference in mean log score between the best model and each other model for predictions of peak timing made before the season onset occurred. Estimates are obtained from a mixed effects model with a separate fixed effect mean for the interaction of model and prediction target; random effects for each combination of region, season, model, and prediction target; and lag 1 autocorrelation nested within each combination of region, season, model, and prediction target. The wider confidence interval bounds are simultaneous confidence intervals with an approximate familywise 95% coverage rate for all individual model means and all pairwise model contrasts for all prediction targets. The inner confidence intervals are calculated separately, with approximate coverage rates of 95%. Log scores of -Infinity were truncated at -10 before fitting this model.





Supplemental Figure 8: Point estimates and confidence intervals for the difference in mean log score between the best model and each other model for predictions of peak incidence made before the season peak occurred. The wider confidence interval bounds are simultaneous confidence intervals with an approximate familywise 95% coverage rate for all individual model means and all pairwise model contrasts for all prediction targets. The inner confidence intervals are calculated separately, with approximate coverage rates of 95%. Log scores of -Infinity were truncated at -10 before fitting this model.