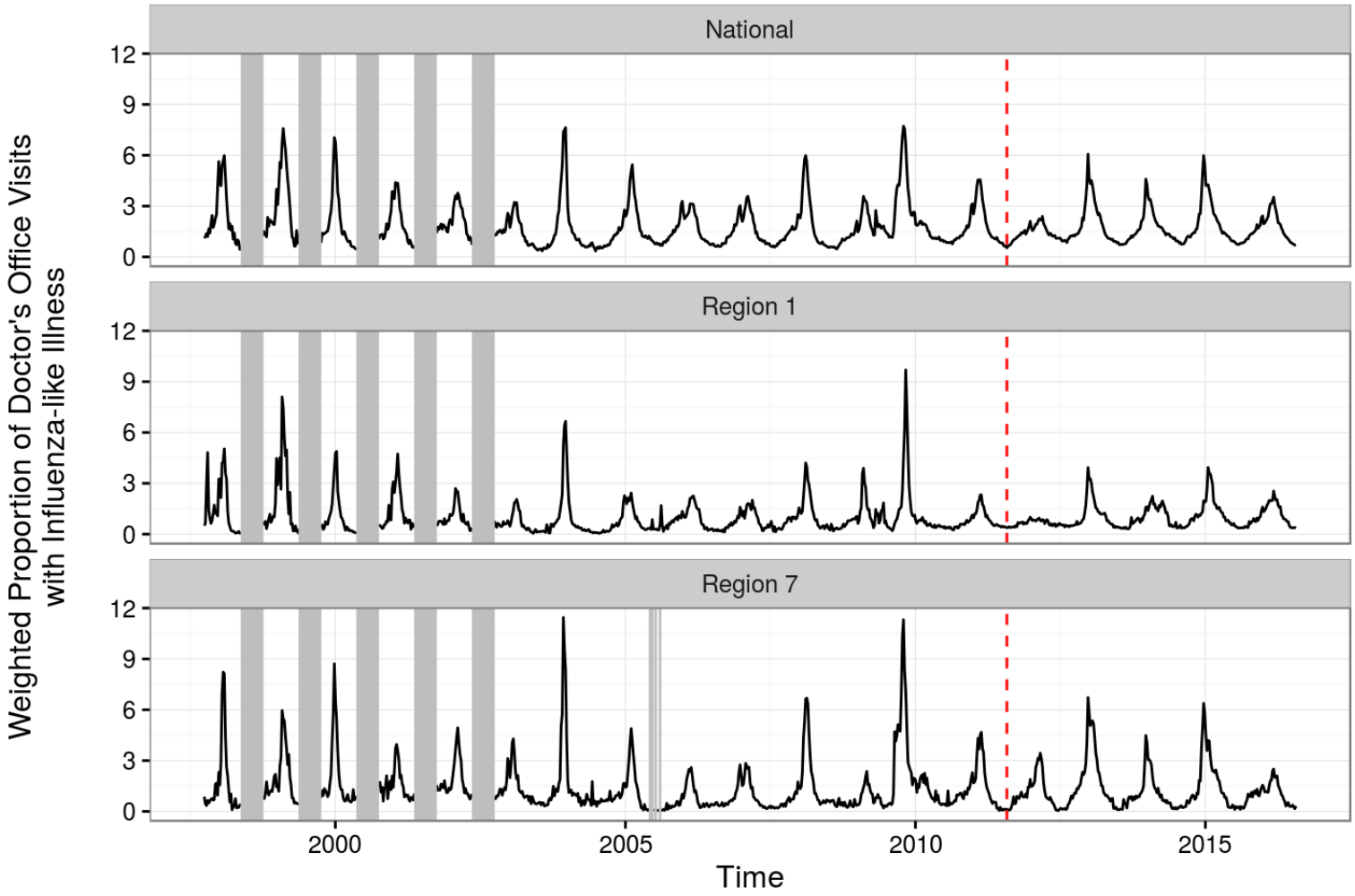
Ensemble Methods for Flu Prediction

Evan Ray, Nick Reich
Department of Biostatistics and Epidemiology
University of Massachusetts, Amherst
http://reichlab.io

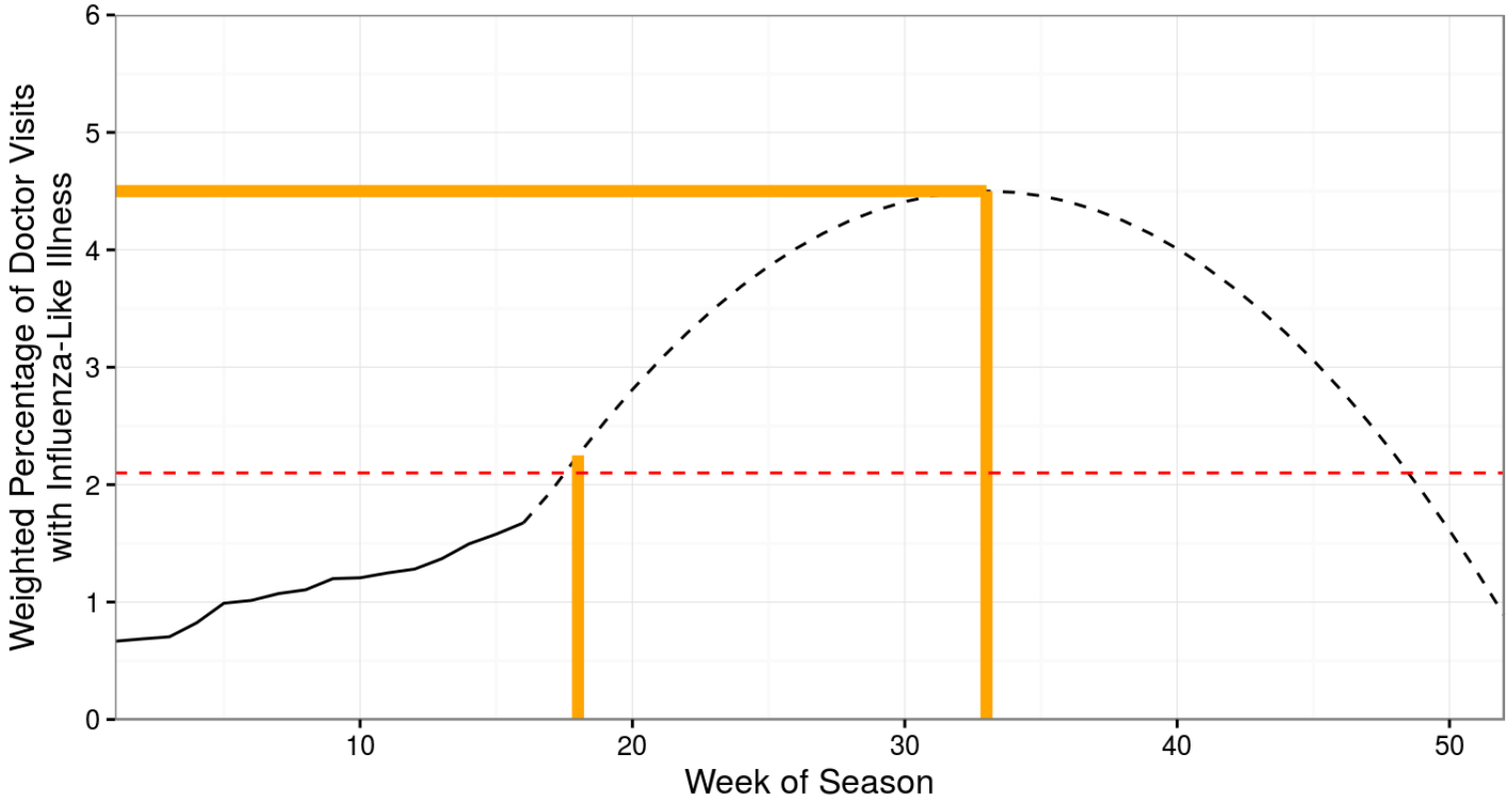
April 29, 2017

Flu Data


National
Region 1
Region 7
Weighted Proportion of Doctor's Office Visits with Influenza-like Illness
0
3
6
9
12
2000
2005
2010
2015
Time

Goal: Predictive Distributions for 3 Targets

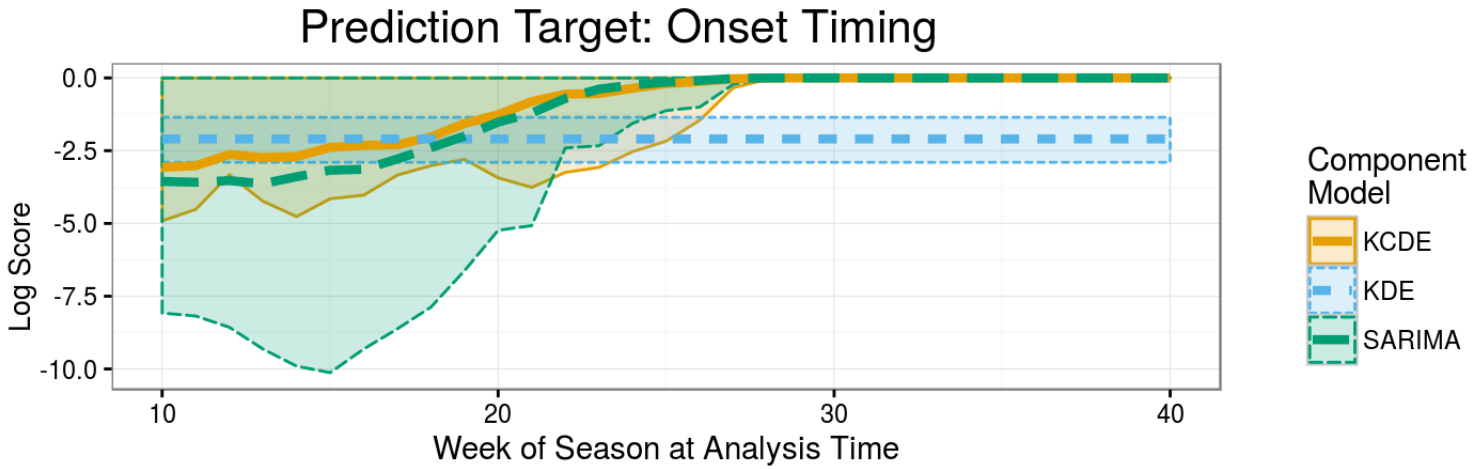
1. Peak Incidence
2. Peak Timing
3. Onset Timing



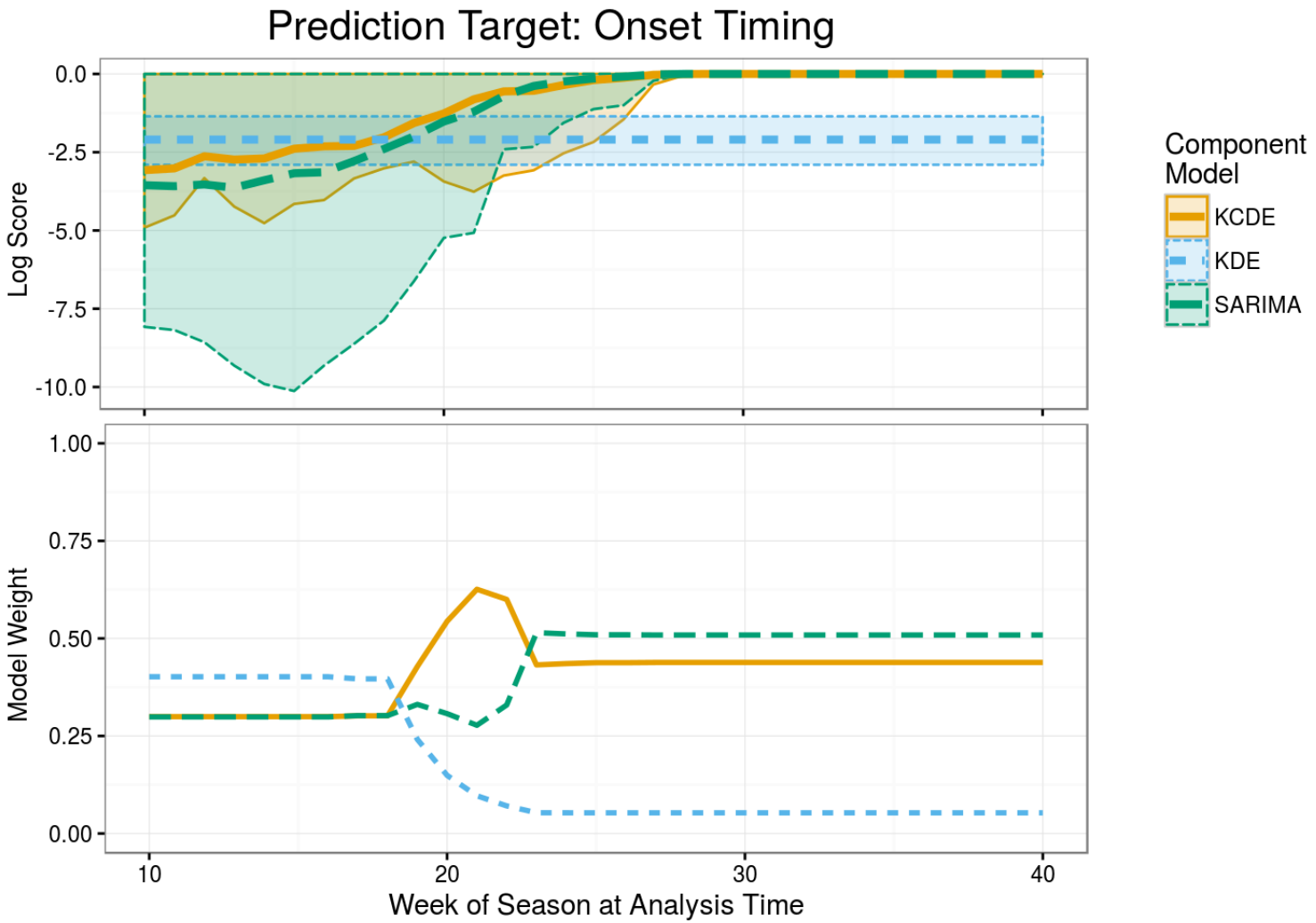
Base Models

- We've Developed Several Models for This Task
  1. **KDE**: Kernel Density Estimation
     - Directly model distribution of season peak incidence/timing and onset timing
     - Distributions are not updated over the course of the season!
  2. **KCDE**: Kernel Conditional Density Estimation + Copulas (Ray et al., under review)
     - Get a joint distribution for incidence in all remaining weeks conditional on recent incidence and time of year
     - Integrate to get distributions for peak incidence/timing and onset timing
  3. **SARIMA**: Seasonal Auto-Regressive Integrated Moving Average model
     - Similar in spirit to **KCDE**, but with more parametric assumptions
- Relative performance varies
  - across different seasons
  - with the week when we're making predictions
  - with model uncertainty
  - with recently observed incidence

Relative Performance Varies with the Week



Relative Performance Varies with the Week



Stacking: Weighted Sum of Predictive Distributions

- We want a predictive distribution for $y|\mathbf{x}$.
  - y = e.g. season onset timing, peak timing, or peak incidence
  - $\mathbf{x}$ = time of year, recent incidence, weather, ...
- We have $M = 3$ predictive distributions $f_m(y|\mathbf{x})$ from different models
- Combine with covariate-dependent weights $\pi_m(\mathbf{x})$:

$$f(y|\mathbf{x}) = \sum_{m=1}^{M} \pi_m(\mathbf{x}) f_m(y|\mathbf{x})$$

- We require $\pi_m(\mathbf{x}) \geq 0$ and $\sum_{m=1}^{M} \pi_m(\mathbf{x}) = 1$ for each $\mathbf{x}$. One approach:

$$\pi_m(\mathbf{x}) = \frac{\exp\{\rho_m(\mathbf{x})\}}{\sum_{m'=1}^{M} \exp\{\rho_{m'}(\mathbf{x})\}}$$

- We estimate the functions $\rho_m(\mathbf{x})$ via Gradient Tree Boosting (GTB)

We'll Compare Several Variations

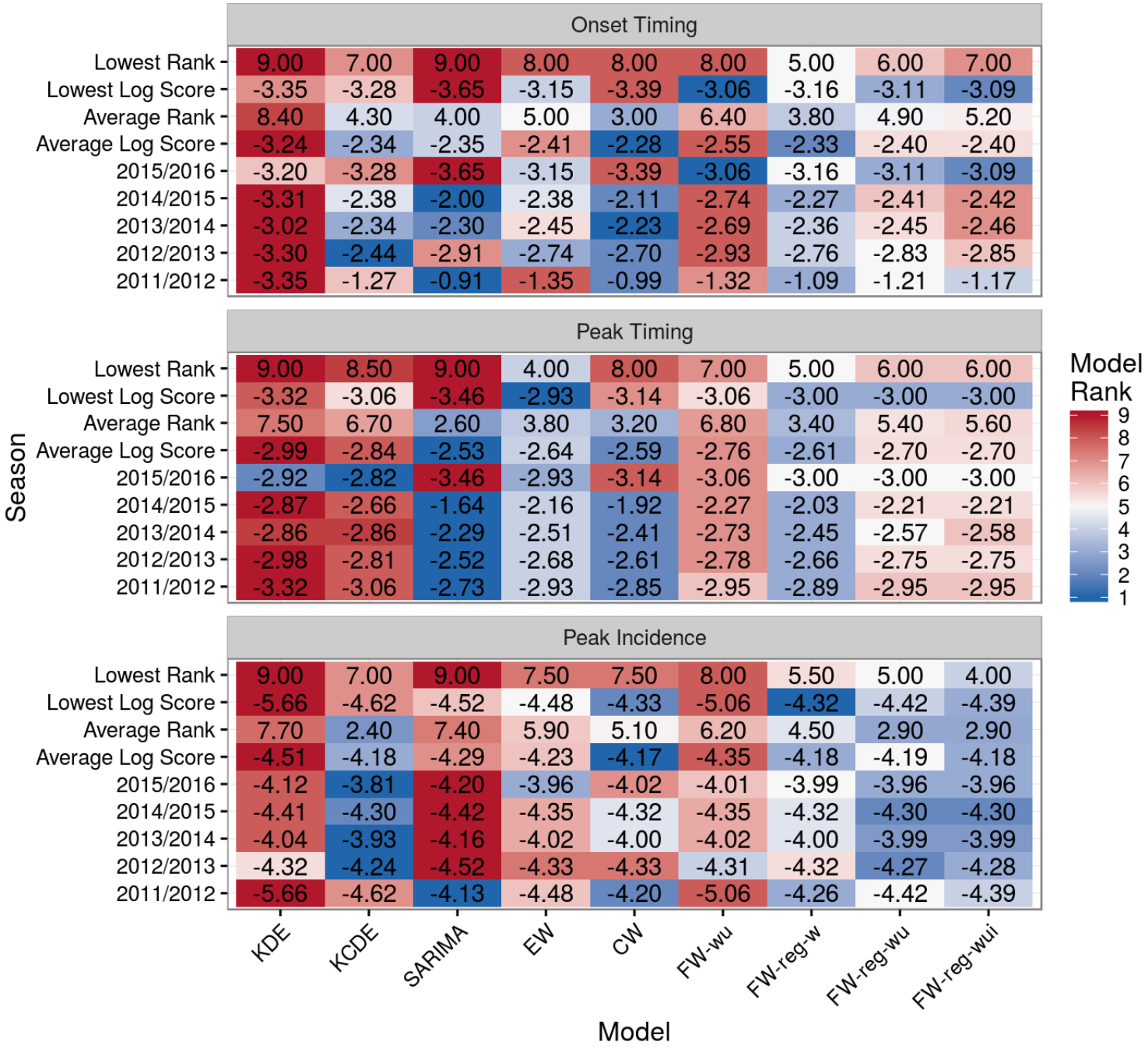
1. Equal Weights (**EW**): $\pi_m(x) = 1/M$.
2. Constant Weights (**CW**): $\pi_m(x) = \pi_m$.
3. Feature-weighted (**FW**): $\pi_m(x)$ depends on features including week of the season and model uncertainty for the KCDE and SARIMA models.
4. Feature-weighted with regularization: $\pi_m(x)$ depends on features, but with regularization:
    1. (**FW-reg-w**) week of the season;
    2. (**FW-reg-wu**) week of the season and model uncertainty for the KCDE and SARIMA models;
    3. (**FW-reg-wui**) week of the season, model uncertainty for the KCDE and SARIMA models, and incidence in the most recent week.

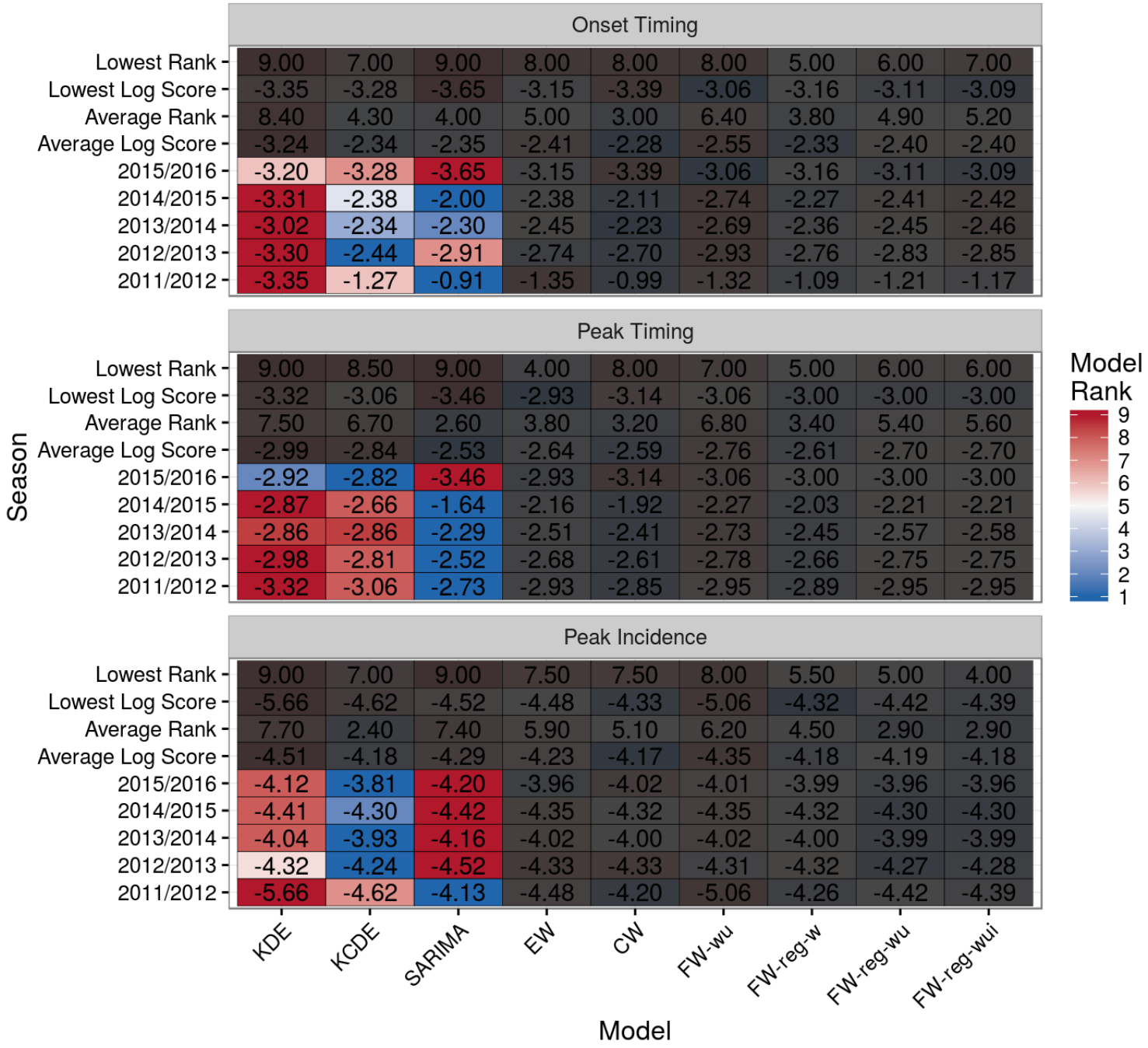
Evaluation

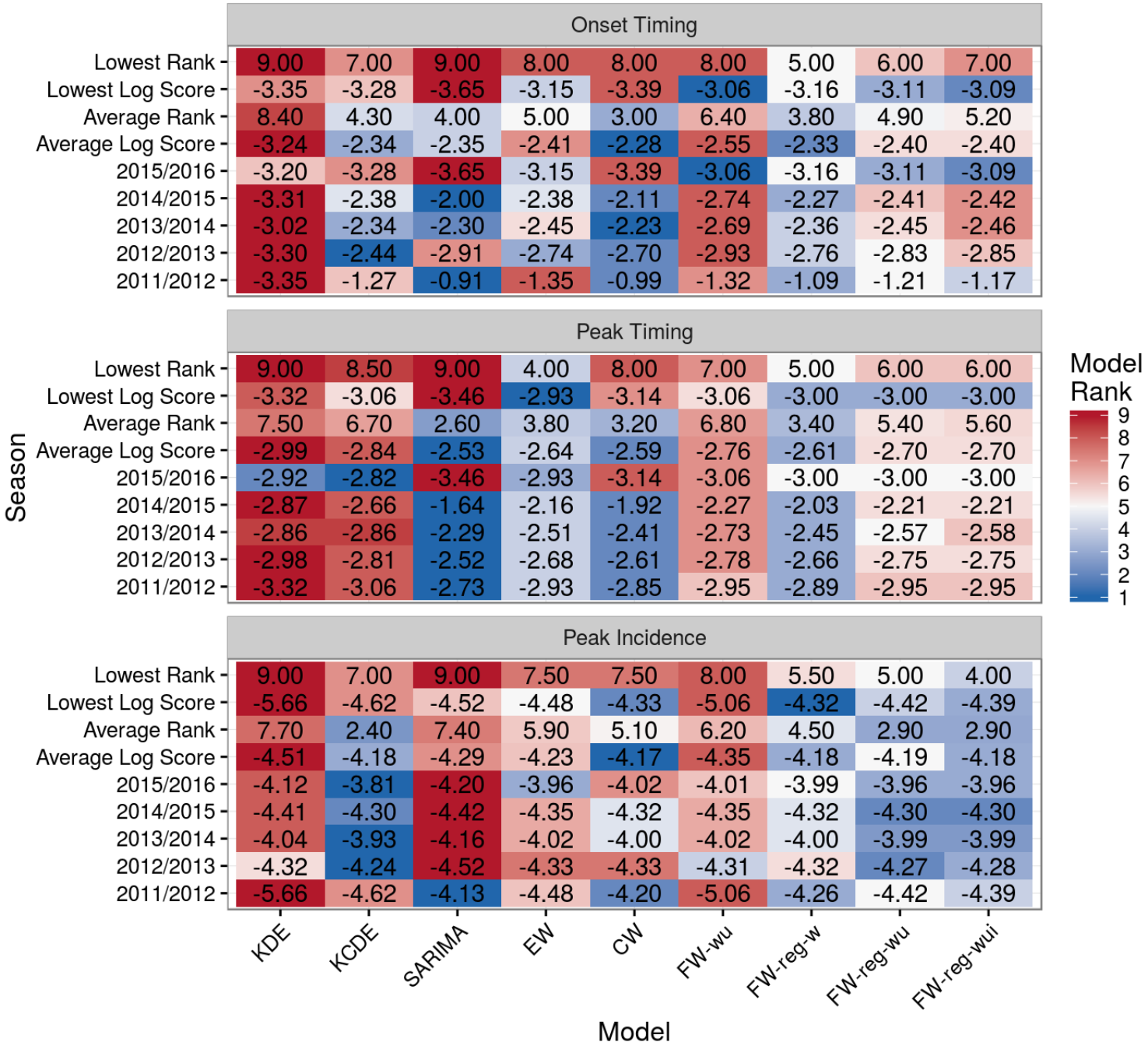
"Public health actions informed by forecasts that later prove to be inaccurate can have negative consequences, including the loss of credibility, wasted and misdirected resources, and, in the worst case, increases in morbidity or mortality."
– Biggerstaff et al. BMC Infectious Diseases 2016; 16(1):357.

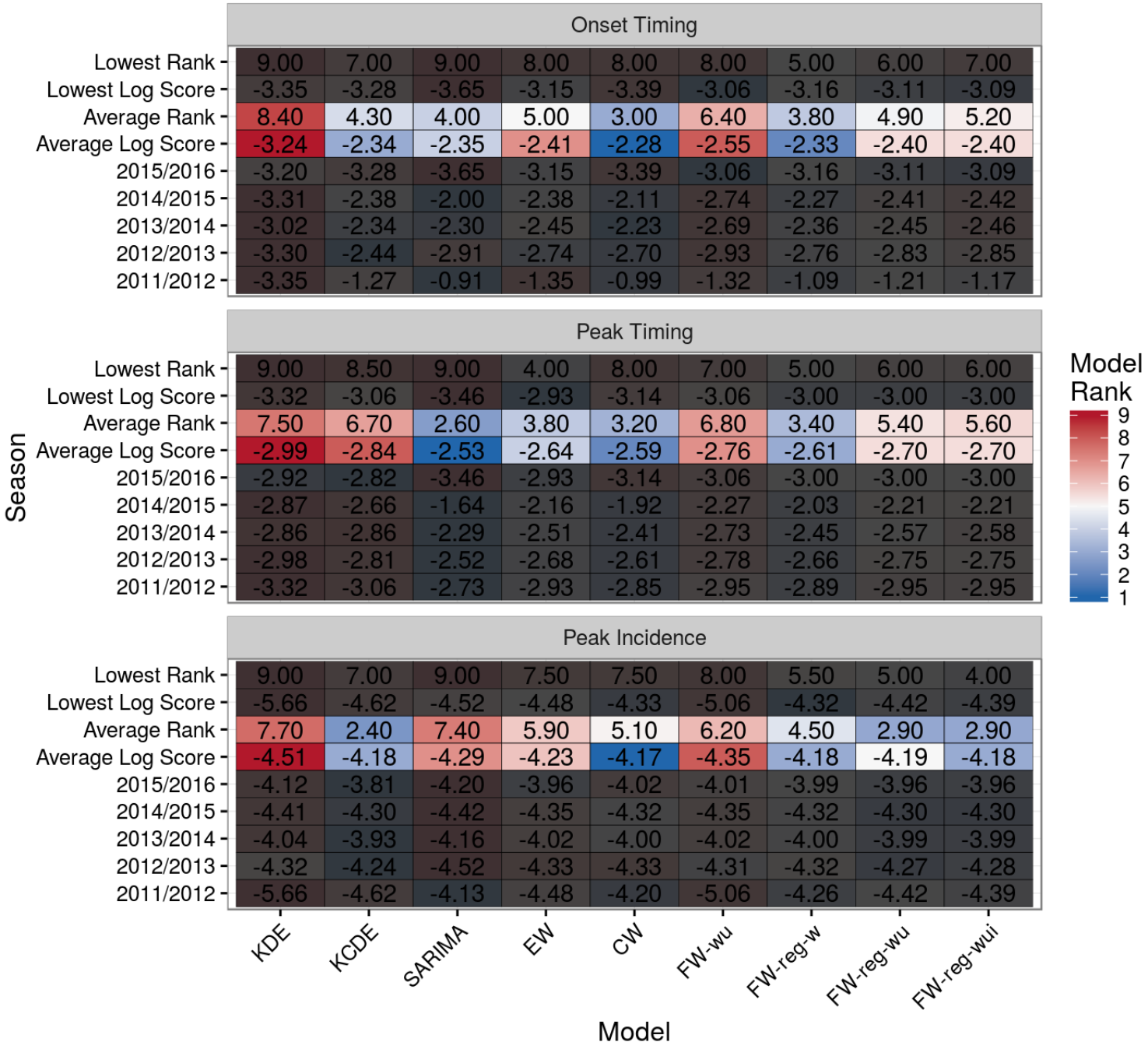
We're looking for two things:

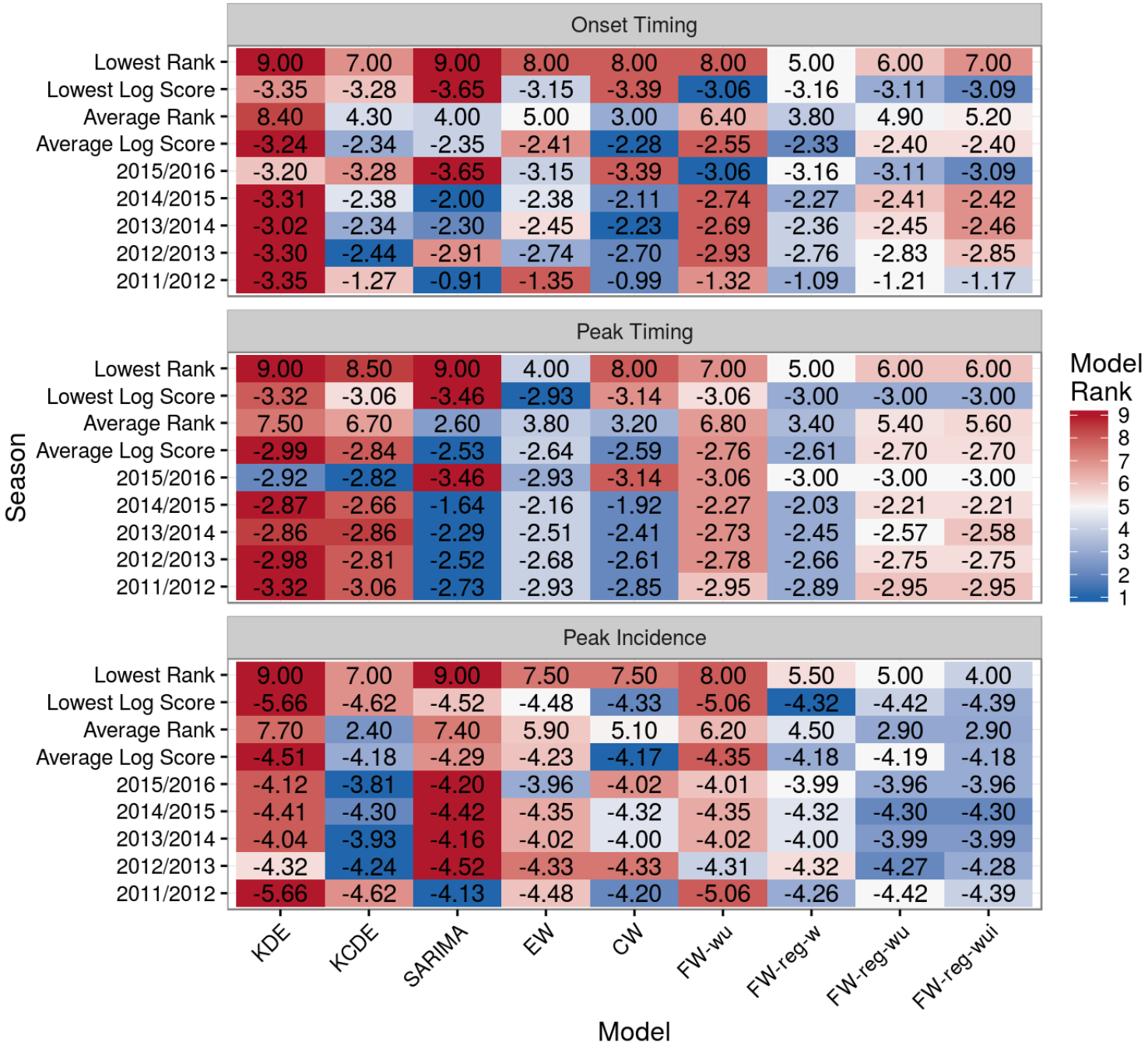
1. Good overall performance
2. Consistency

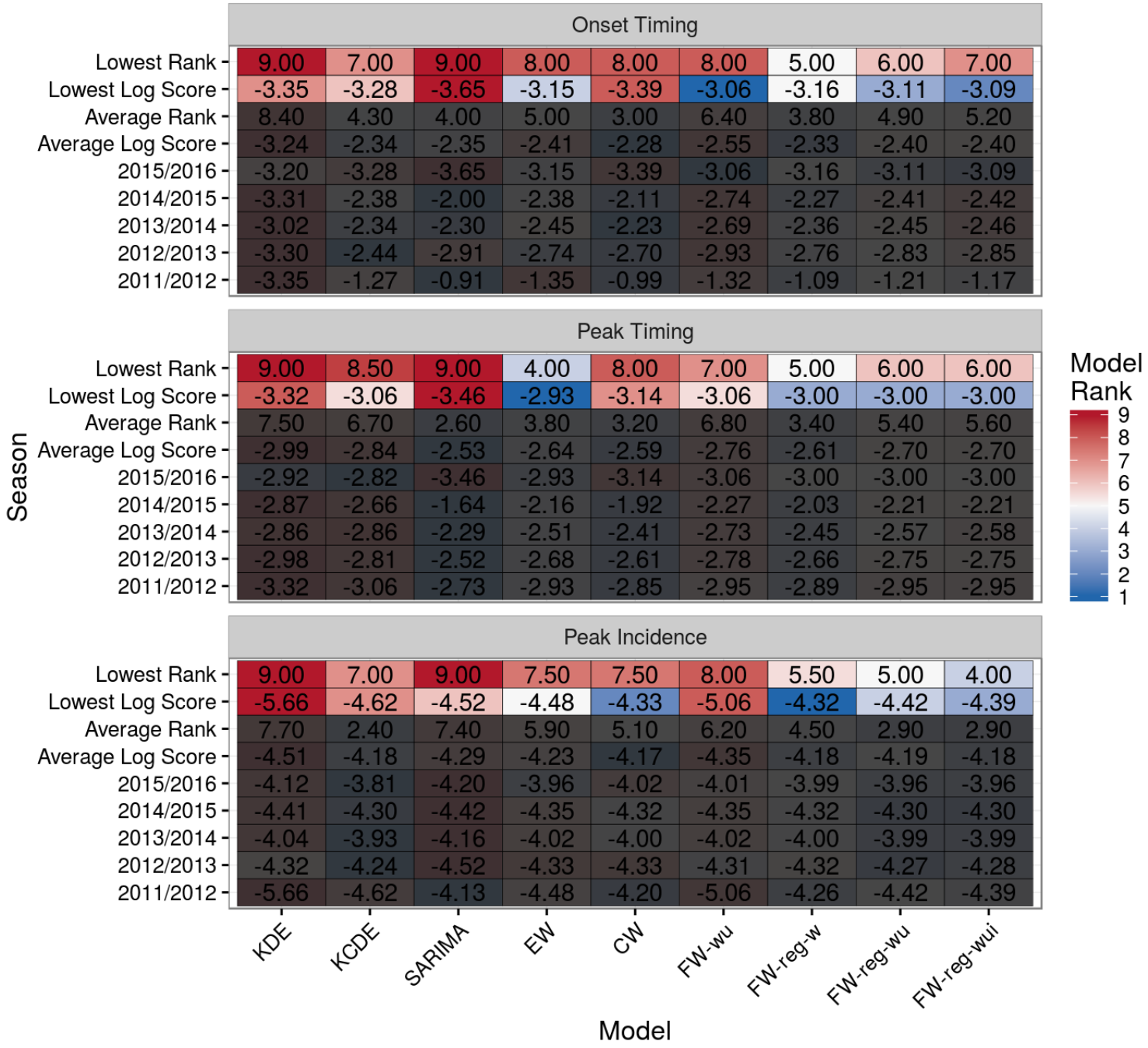

Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE KCDE SARIMA EW CW FW-wu FW-reg-w FW-reg-wu FW-reg-wui
Model
Model Rank
9 8 7 6 5 4 3 2 1

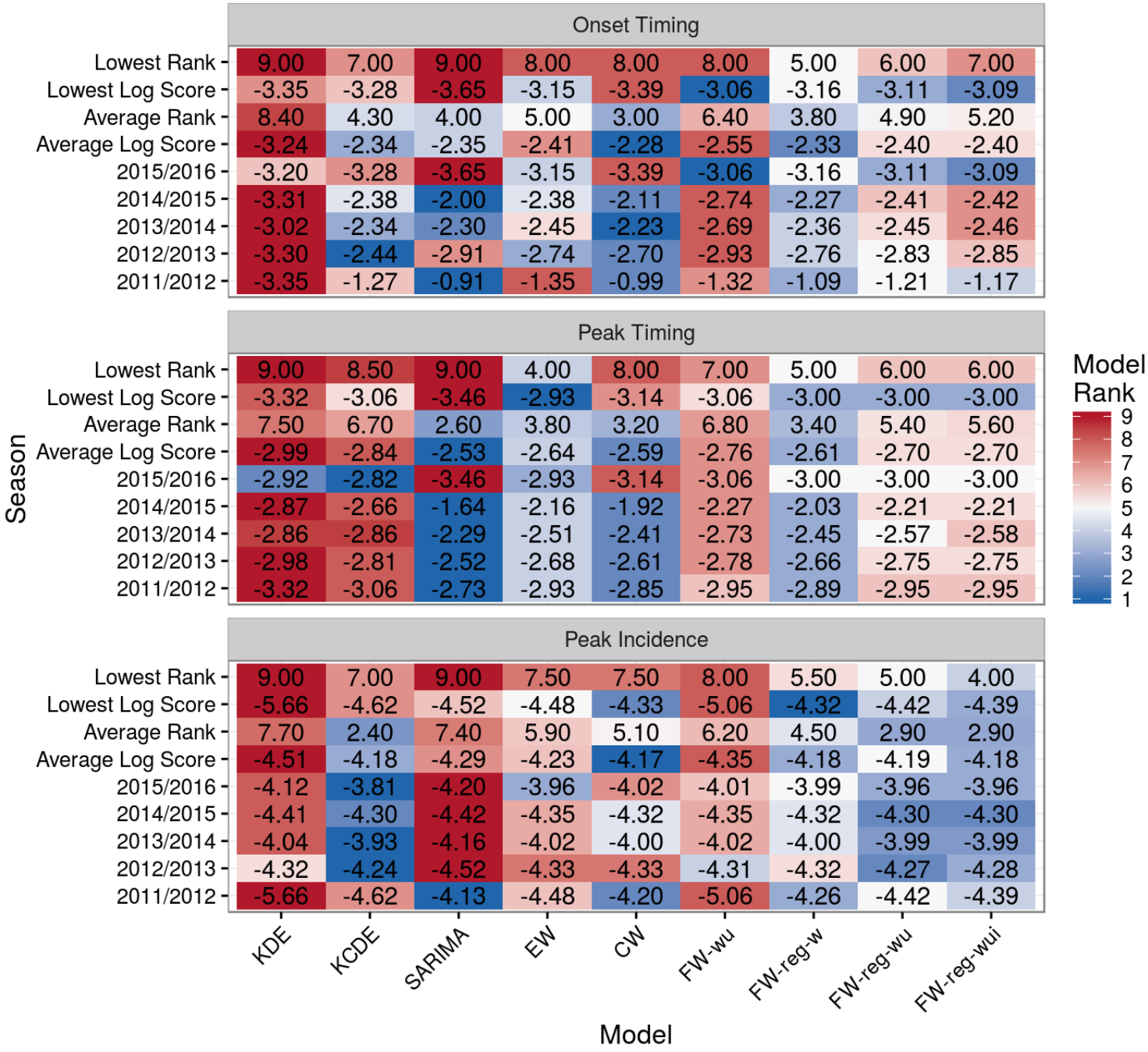

Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE KCDE SARIMA EW CW FW-wu FW-reg-w FW-reg-wu FW-reg-wui
Model
Model Rank
9 8 7 6 5 4 3 2 1


Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE
KCDE
SARIMA
EW
CW
FW-wu
FW-reg-w
FW-reg-wu
FW-reg-wui
Model
Model Rank
9
8
7
6
5
4
3
2
1


Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE
KCDE
SARIMA
EW
CW
FW-wu
FW-reg-w
FW-reg-wu
FW-reg-wui
Model
Model Rank
9
8
7
6
5
4
3
2
1


Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE KCDE SARIMA EW CW FW-wu FW-reg-w FW-reg-wu FW-reg-wui
Model
Model Rank
9 8 7 6 5 4 3 2 1

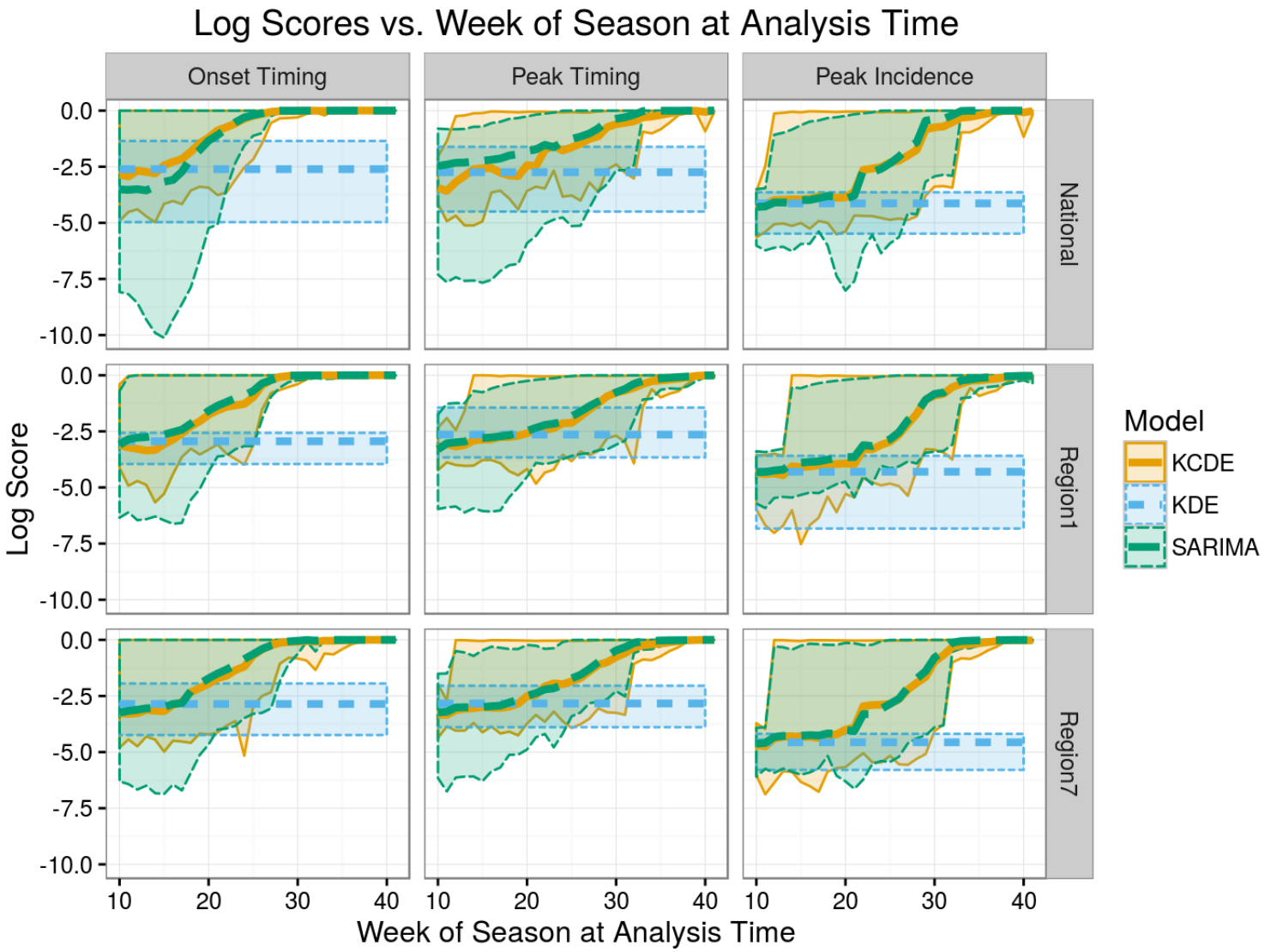

Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE
KCDE
SARIMA
EW
CW
FW-wu
FW-reg-w
FW-reg-wu
FW-reg-wui
Model
Model Rank
9
8
7
6
5
4
3
2
1


Onset Timing
Lowest Rank 9.00 7.00 9.00 8.00 8.00 8.00 5.00 6.00 7.00
Lowest Log Score -3.35 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
Average Rank 8.40 4.30 4.00 5.00 3.00 6.40 3.80 4.90 5.20
Average Log Score -3.24 -2.34 -2.35 -2.41 -2.28 -2.55 -2.33 -2.40 -2.40
2015/2016 -3.20 -3.28 -3.65 -3.15 -3.39 -3.06 -3.16 -3.11 -3.09
2014/2015 -3.31 -2.38 -2.00 -2.38 -2.11 -2.74 -2.27 -2.41 -2.42
2013/2014 -3.02 -2.34 -2.30 -2.45 -2.23 -2.69 -2.36 -2.45 -2.46
2012/2013 -3.30 -2.44 -2.91 -2.74 -2.70 -2.93 -2.76 -2.83 -2.85
2011/2012 -3.35 -1.27 -0.91 -1.35 -0.99 -1.32 -1.09 -1.21 -1.17
Peak Timing
Lowest Rank 9.00 8.50 9.00 4.00 8.00 7.00 5.00 6.00 6.00
Lowest Log Score -3.32 -3.06 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
Average Rank 7.50 6.70 2.60 3.80 3.20 6.80 3.40 5.40 5.60
Average Log Score -2.99 -2.84 -2.53 -2.64 -2.59 -2.76 -2.61 -2.70 -2.70
2015/2016 -2.92 -2.82 -3.46 -2.93 -3.14 -3.06 -3.00 -3.00 -3.00
2014/2015 -2.87 -2.66 -1.64 -2.16 -1.92 -2.27 -2.03 -2.21 -2.21
2013/2014 -2.86 -2.86 -2.29 -2.51 -2.41 -2.73 -2.45 -2.57 -2.58
2012/2013 -2.98 -2.81 -2.52 -2.68 -2.61 -2.78 -2.66 -2.75 -2.75
2011/2012 -3.32 -3.06 -2.73 -2.93 -2.85 -2.95 -2.89 -2.95 -2.95
Peak Incidence
Lowest Rank 9.00 7.00 9.00 7.50 7.50 8.00 5.50 5.00 4.00
Lowest Log Score -5.66 -4.62 -4.52 -4.48 -4.33 -5.06 -4.32 -4.42 -4.39
Average Rank 7.70 2.40 7.40 5.90 5.10 6.20 4.50 2.90 2.90
Average Log Score -4.51 -4.18 -4.29 -4.23 -4.17 -4.35 -4.18 -4.19 -4.18
2015/2016 -4.12 -3.81 -4.20 -3.96 -4.02 -4.01 -3.99 -3.96 -3.96
2014/2015 -4.41 -4.30 -4.42 -4.35 -4.32 -4.35 -4.32 -4.30 -4.30
2013/2014 -4.04 -3.93 -4.16 -4.02 -4.00 -4.02 -4.00 -3.99 -3.99
2012/2013 -4.32 -4.24 -4.52 -4.33 -4.33 -4.31 -4.32 -4.27 -4.28
2011/2012 -5.66 -4.62 -4.13 -4.48 -4.20 -5.06 -4.26 -4.42 -4.39
Season
KDE
KCDE
SARIMA
EW
CW
FW-wu
FW-reg-w
FW-reg-wu
FW-reg-wui
Model
Model Rank
9
8
7
6
5
4
3
2
1

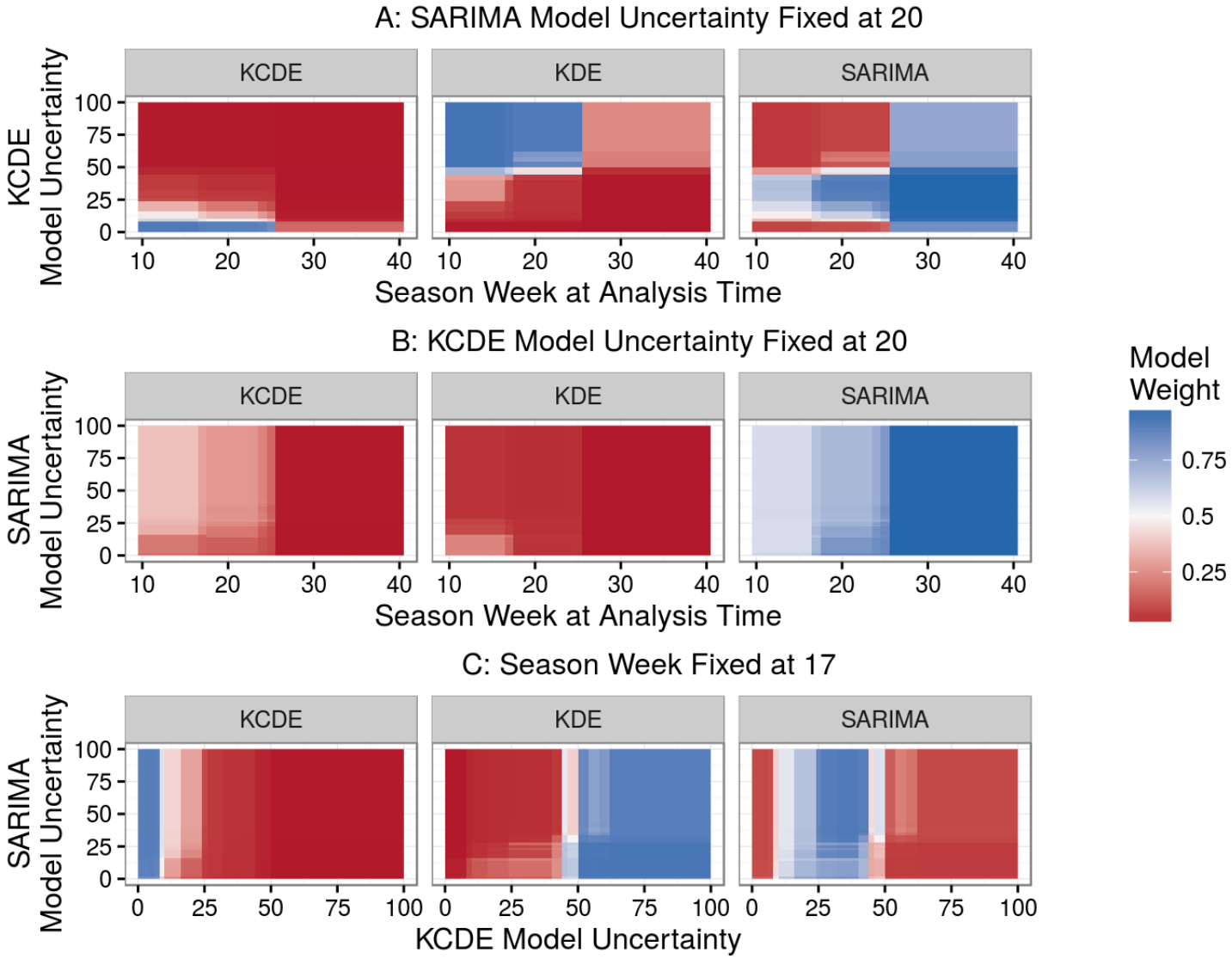
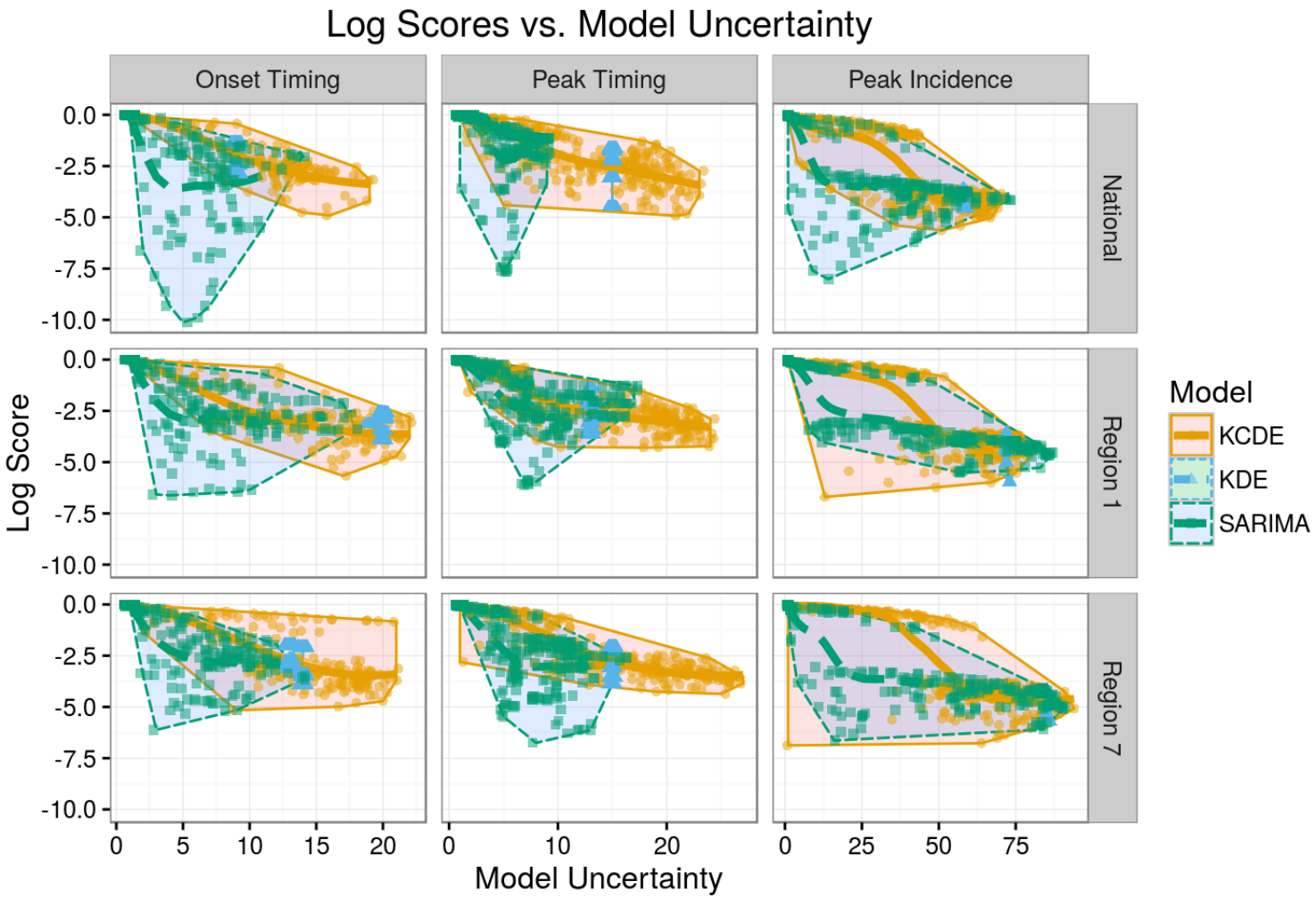
Conclusions

- All reasonable ensembles had *similar overall performance* as the best of the component models
- **FW-reg-w** ensemble had *more consistent performance* across seasons than the component models
- Limitations:
  - Only 3 component models – would prefer to have more diversity
  - 5 test phase seasons
- Future Directions:
  - Additional component models (additional predictive covariates, model disease transmission mechanisms)
  - Conditionally adaptive kernel conditional density estimation via ensemble framework

Relative performance varies with the week



Model Weight Functions Are Reasonable (Part 2)



Relative performance varies with model uncertainty