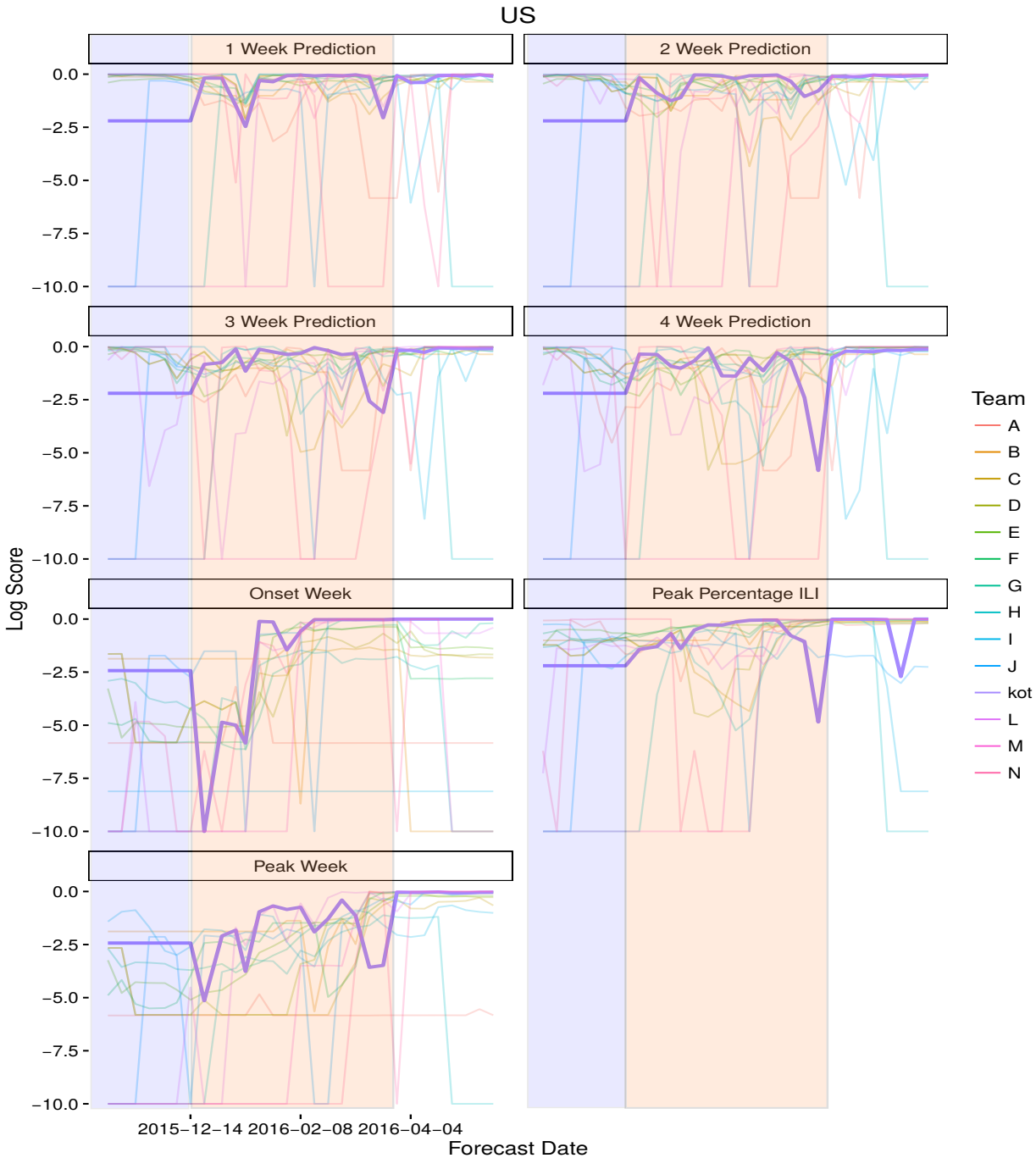
Team Kernel of Truth – Influenza Forecasting Discussion

Evan Ray

University of Massachusetts - Amherst

August 31 - September 1 2016


US
1 Week Prediction
2 Week Prediction
3 Week Prediction
4 Week Prediction
Onset Week
Peak Percentage ILI
Peak Week
0.0
−2.5
−5.0
−7.5
−10.0
Log Score
2015−12−14
2016−02−08
2016−04−04
Forecast Date
Team
A
B
C
D
E
F
G
H
I
J
kot
L
M
N

Summary of Results for National Predictions

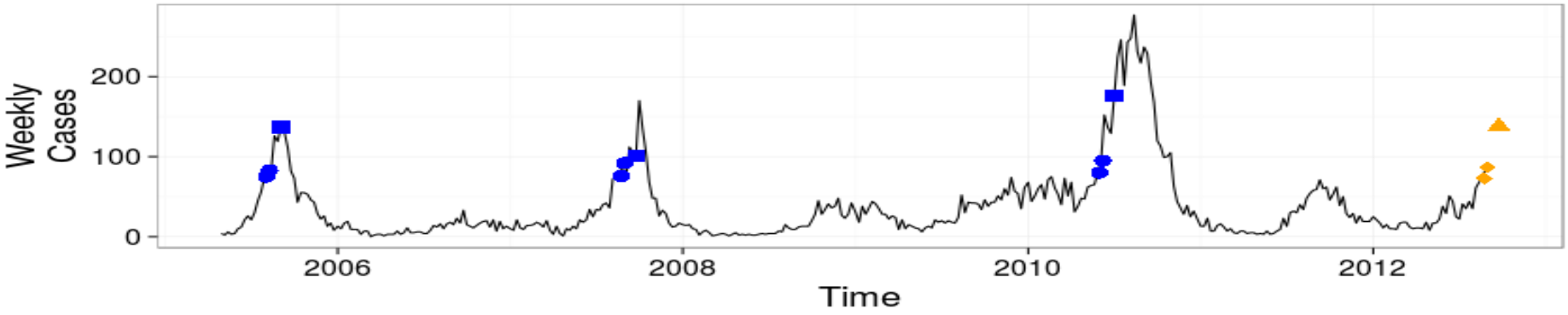
	Onset	Peak Week	Peak Week %	Combined target	Exp Combined
A	-5.61193	-5.78944	-0.58974	-3.99704	1.8
B	-3.59949	-1.38075	-0.58941	-1.85655	15.6
C	-2.76998	-3.669	-1.19191	-2.54363	7.9
D	-2.69239	-3.02018	-1.10988	-2.27415	10.3
E	-2.48957	-2.2959	-0.3909	-1.72546	17.8
F	-2.95212	-2.27653	-0.3788	-1.86915	15.4
G	-2.2877	-1.91221	-0.47389	-1.55794	21.1
H	-5.53289	-5.52674	-5.10347	-5.38770	0.5
I	-8.11462	-1.73125	-1.77745	-3.87444	2.1
J	-2.26622	-3.05179	-2.57078	-2.62959	7.2
Kernel of Truth	-1.55756	-1.55858	-1.06941	-1.39518	24.8
L	-3.89439	-3.34374	-1.10202	-2.78005	6.2
M	-6.22494	-5.35493	-5.01637	-5.53208	0.4
N	-3.29722	-6.66888	-2.50142	-4.15584	1.6

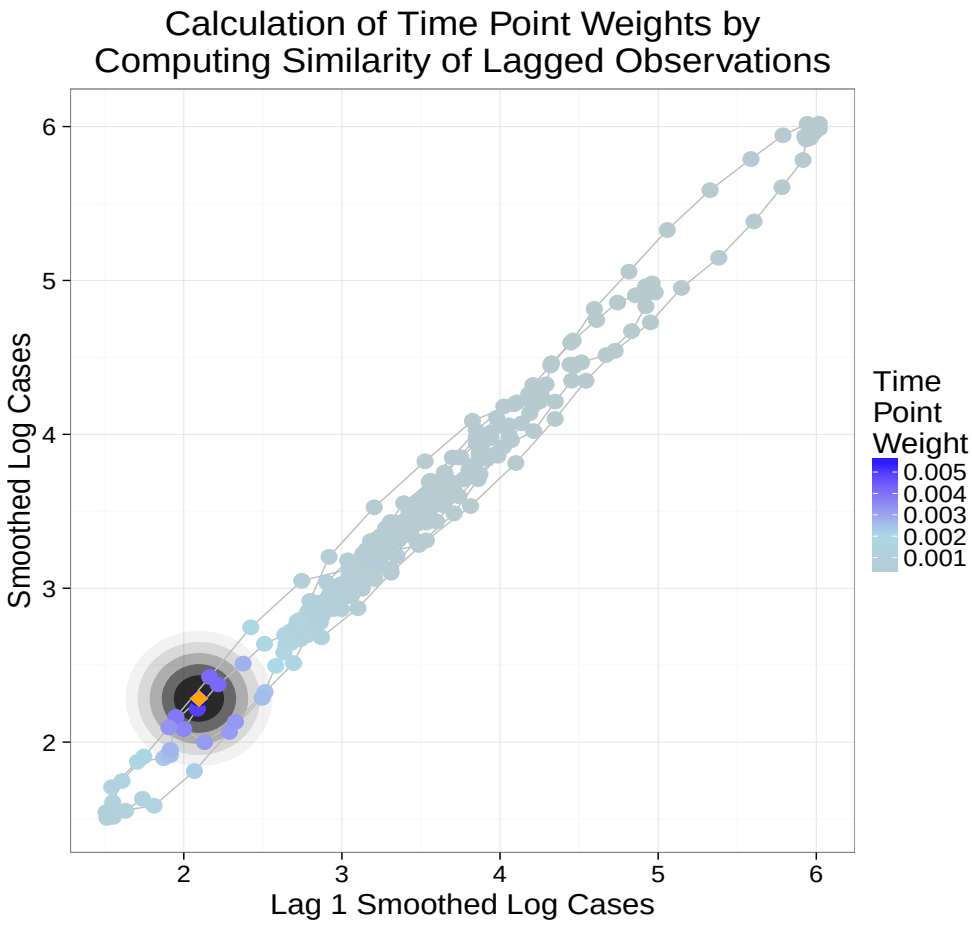
	1wk	2wk	3wk	4wk	Combined target	Exp Combined
A	-1.38183	-1.31384	-1.40009	-1.53934	-1.40878	24.4
B	-0.44044	-0.5662	-0.65333	-0.72242	-0.59560	55.1
C	-0.41802	-0.84522	-1.21178	-1.47449	-0.98738	37.3
D	-0.18399	-0.38742	-0.69879	-0.91415	-0.54609	57.9
E	-0.41499	-0.49524	-0.57115	-0.64528	-0.53167	58.8
F	-0.17807	-0.32362	-0.43757	-0.63008	-0.39233	67.5
G	-0.23845	-0.37888	-0.4748	-0.55298	-0.41128	66.3
H	-4.35103	-4.45316	-4.6123	-4.77534	-4.54796	1.1
I	-0.47669	-0.71313	-0.86633	-1.13026	-0.79660	45.1
J	-1.8355	-1.86009	-1.90443	-1.94016	-1.88504	15.2
Kernel of Truth	-0.83017	-0.76615	-0.92935	-1.18107	-0.92668	39.6
L	-0.50116	-0.86639	-1.63725	-1.57596	-1.14519	31.8
M	-4.94675	-5.06935	-5.68862	-5.67702	-5.34544	0.5
N	-2.62992	-2.01352	-2.74985	-3.04177	-2.60877	7.4

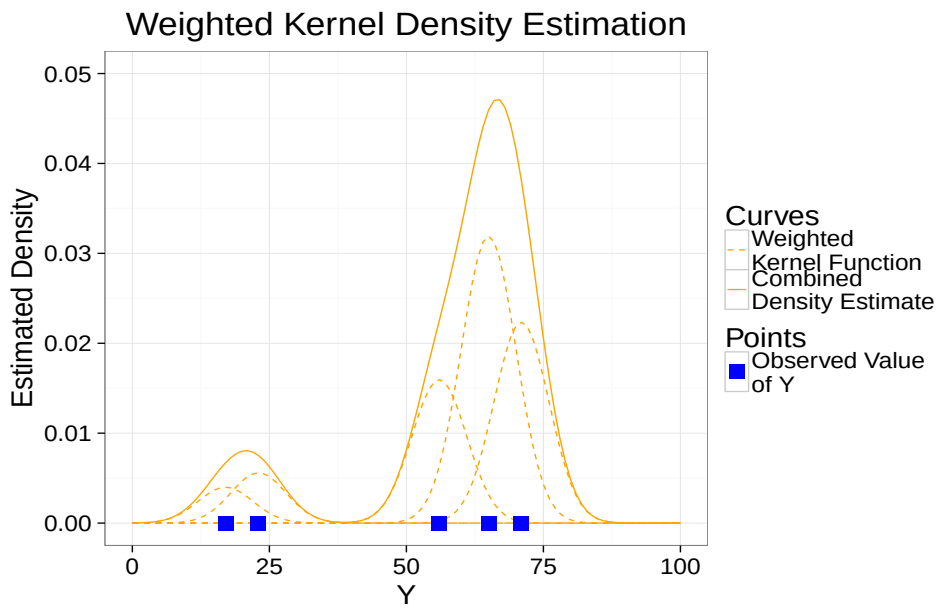
Overview of Method

1. Use kernel conditional density estimation (KCDE) to obtain separate predictive distributions for incidence in each future week.
2. Use copulas to combine these individual distributions to get joint predictive distribution for the trajectory of incidence over the rest of the season.
3. Obtain predictive distributions for onset week, peak ILI, and peak week from this joint predictive distribution.

More detail about KCDE

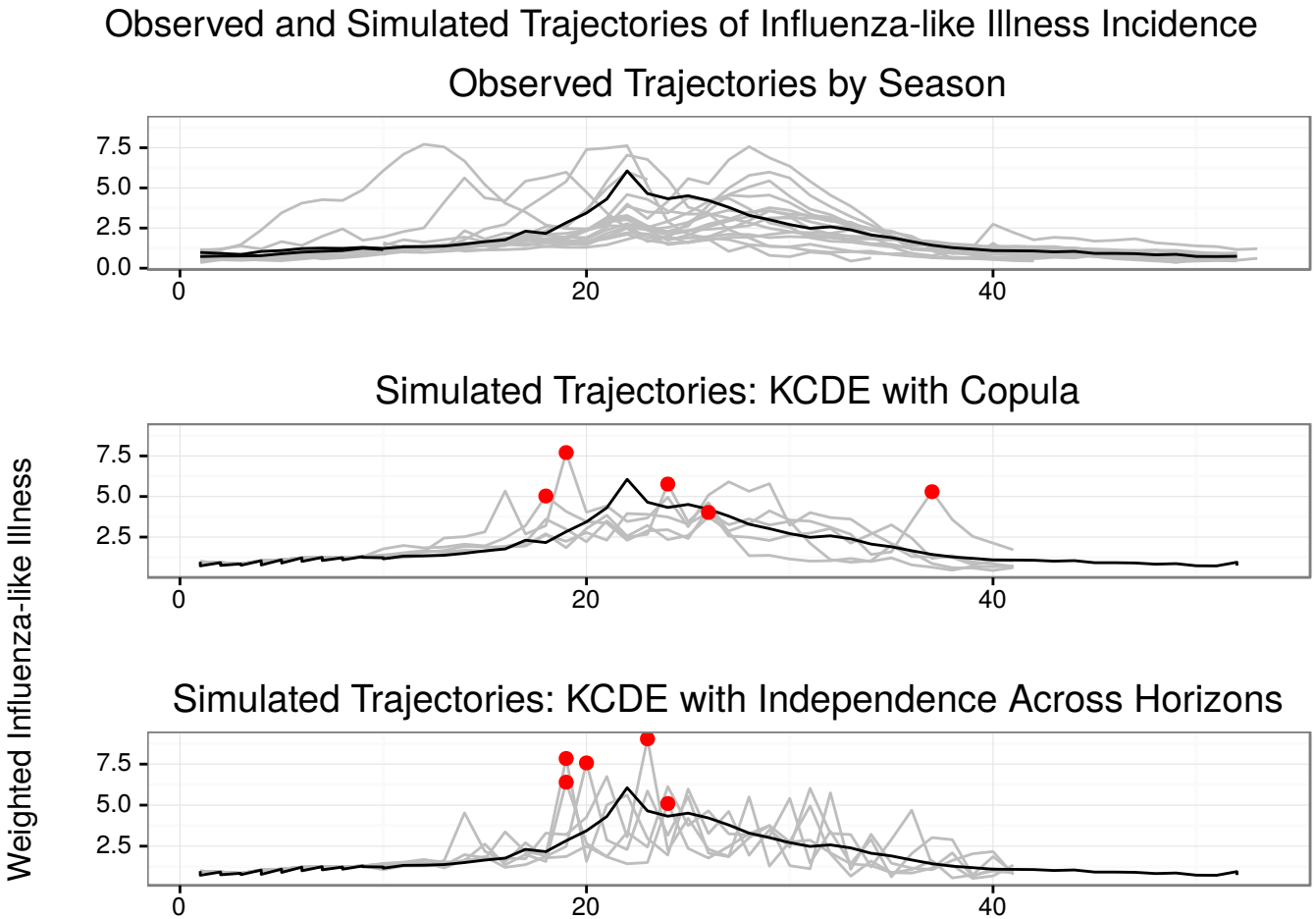






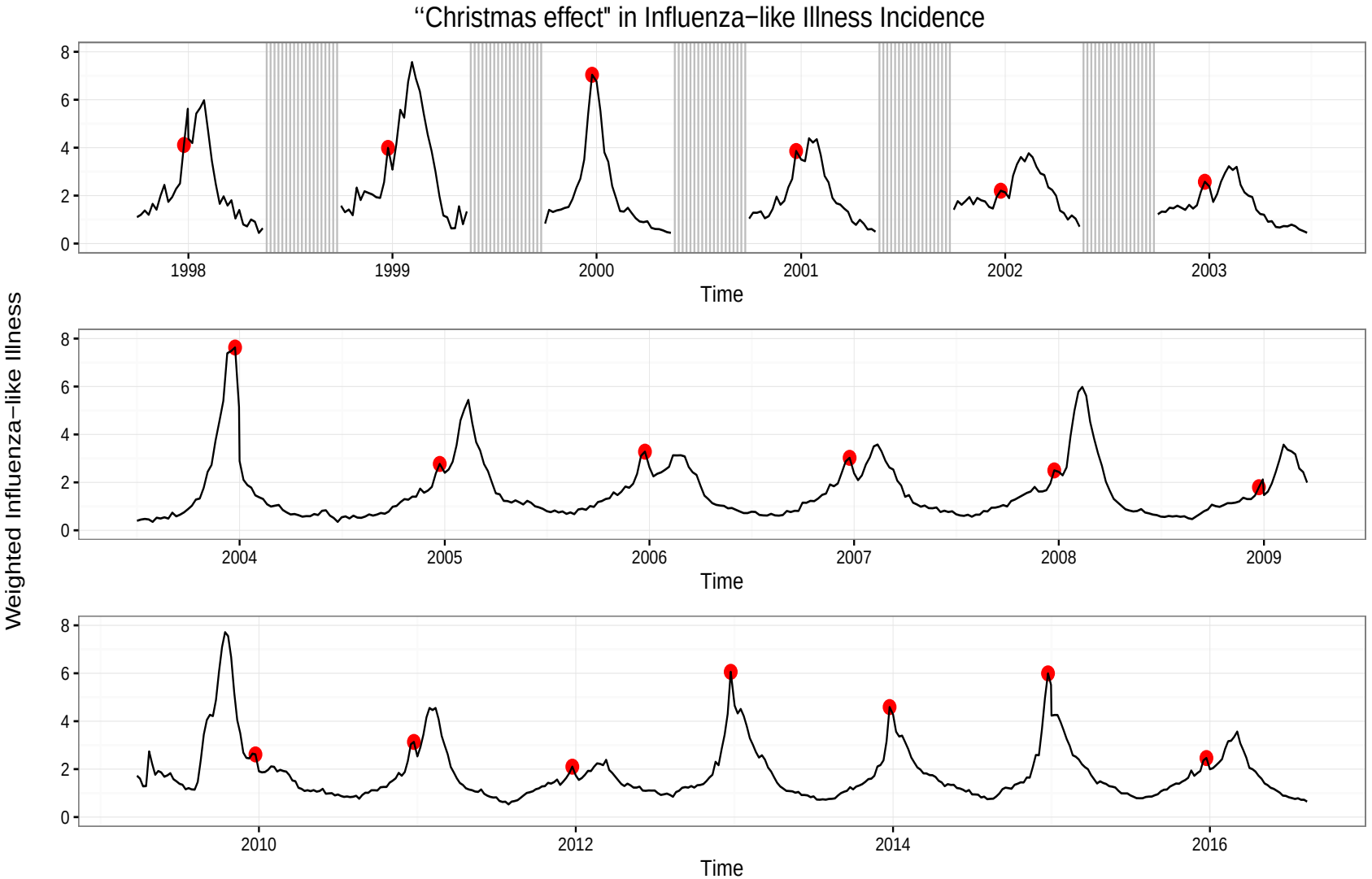
(Slightly) more detail about copulas

- Sklar's Theorem: For any random vector $(X_1, \ldots, X_D)$, there exists a function C such that
  $F_{X_1,\ldots,X_D}(x_1, \ldots, x_D) = C\{F_{X_1}(x_1), \ldots, F_{X_D}(x_d)\}$
- We have used the parametric specification for C that arises in a multivariate normal distribution



A Weakness: The Christmas Effect

- Our method did not capture the Christmas effect:



Future Directions

- Participate fully (all regions, consistent model formulation throughout)
- Ensembles to combine models with varying strength at different points in the season
  - (Unconditional) KDE near the beginning of the season?
  - SARIMA near Christmas?
  - KCDE+copulas elsewhere?
- We have some other ideas, but won't get to them by the start of the next round...