Creating statistical models that generate accurate predictions of infectious disease incidence over multiple time points is a challenging problem. Forecasts of infectious disease outbreaks can be used to inform targeted intervention and prevention strategies such as increased healthcare staffing or vector control measures. I will present a new approach to predicting infectious disease that uses kernel conditional density estimation (KCDE) and copulas. The method was a top performer in the 2015-2016 FluSight influenza forecasting challenge run in real-time by the CDC. The method works by obtaining predictive distributions for incidence in individual weeks using KCDE and then tying those distributions together into joint distributions using copulas. This strategy enables us to create predictions for the timing of and incidence in the peak week of a given season. Our implementation of KCDE also incorporates two novel kernel components: a periodic component that captures seasonality in disease incidence, and a component that allows for a full parameterization of the bandwidth matrix with discrete variables. We apply the method to predicting dengue fever and influenza.

Overall, KCDE compares favorably to other standard prediction methods including a seasonal autoregressive integrated moving average (SARIMA) model and a previously published generalized linear model for infectious disease incidence known as HHH4. Finally, I will discuss ongoing work that integrates this method into an ensemble model averaging framework using model weights that vary based on the time of year predictions are made.