The Mosquito Aedes aegypti is a vector for multiple viruses around the globe. Its distribution is restricted to tropical and subtropical climates, which reflects its sensitivity to temperature, humidity and other weather constraints. The modulation of its life cicle by climate, shapes the seasonality of the diseases it transmits. In Brazil, Aedes aegypti has been mainly associated with the transmission of dengue, making it a marked seasonal disease. In recent years, A. aegypti has also been notably responsible for epidemics of the Zika and Chikungunya virus. In this paper we explore the performance of dengue forecast models trained on the longer available incidence timeseries to predict the weekly incidence of Zika and Chikungunya as well. We will use a LSTM (long short term memory) recursive neural network model, which we have shown, in a previous work, to yield accurate forecasts for weekly dengue incidence. We will also compare it to a Random Quantile Forest model. Climate variables such as temperature, humidity, and atmospheric pressure are also used as predictors. A spatial component built from the incidence at neighboring cities is also included. We present results of the forecast of total incidence of arboviral disease as well as of each disease separately and discuss the relative performances of the model for each of these tasks.

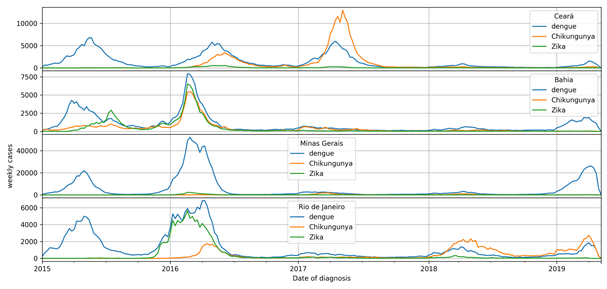

Figure 1: Weekly incidence for the state of Ceará, Brazil. The three arboviroses incidences are correlated.

The data used in this work comes from the Infodengue project, which monitors these arboviroses in Brazil. In figure 1 we can see the incidence series for states in Brazil which had significant outbreaks of the 3 arboviroses since 2016 when Chikungunya and Zika arrived in Brazil.

Dengue forecast models

Both LSTM and RQF models were trained to predict 4 weeks ahead of the last data point(w_{t+4}). Both models use 4 weeks of historical data to generate forecasts. Forecasts are done in a rolling window fashion. Both models use as predictors, the following series: number of cases, Effective reproduction number (R_t),Temperature, Humidity and Atmospheric pressure.

LSTM. A LSTM model is a recurrent deep neural network model developed to handle predictions of timeseries. We used a LSTM model with 3 LSTM units followed by 3 dropout layers. The model was trained for 300 epochs using a mean-log squared-error (MLSE) loss function and a Nesterov Adam optimizer[3]. Figure 2 shows dengue forecasts by the LSTM model for Fortaleza.

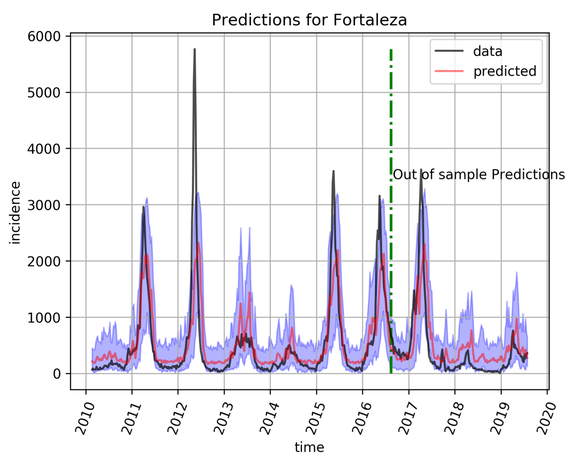

Figure 2: LSTM forecast for Dengue incidence (black line) Prediction (red line) with 95% interval for the city of Fortaleza. Every point in the red line corresponds to prediction from 4 weeks before.

Random Quantile Forest (RQF). Random Forest models calculate an ensemble of regression trees from random subsets of data. RQF models are an extension to regular random forests in which the full conditional distribution of Y given $X = x$ is calculated. As a result, it is a non-parametric, consistent and accurate way to determine conditional quantiles from high-dimensional predictors[2]. Let T be an array containing the $\mathcal{D} = 4$ most recent observations from each series in the predictor matrix. Thus the regression model can be simply represented by

$$\hat{y}_{t+\tau} = \beta_t T_t + \epsilon_t \tag{1}$$

Figure 3 shows dengue forecasts for Rio de Janeiro by the RQF model.

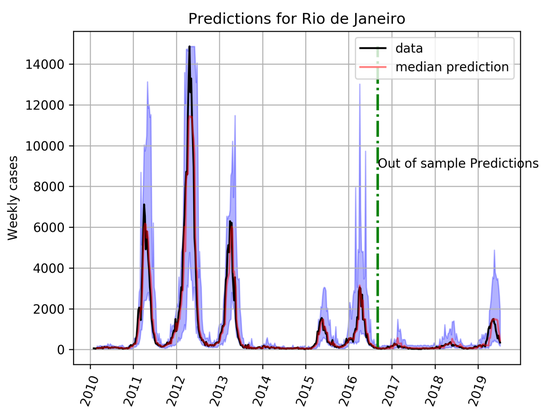

Figure 3: Baseline forecast for dengue in Rio de Janeiro. Random quantile forest model trained on data from 2010 to mid 2016. Red line is the median prediction, with 95% intervals in light purple.

Results

Evaluating dengue forecasting models to predict Zika and Chikungunya in Brazil.

Flávio Codeço Coelho[1]✉, Elisa Mussumeci[1], Marcelo Orgler[1], Linneu Holanda[1]
[1]Fundação Getulio Vargas, Brazil; Correspondence to: fccoelho@fgv.br

Poster DOI: ADD here from figshare
github.com/AlertaDengue — @fccoelho

FGV EMAp

GEOMED 2019 GLASGOW