

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

Flávio Codeço Coelho<sup>1</sup>✉, Elisa Mussumeci<sup>1</sup>, Marcelo Orgler<sup>1</sup>, Linneu Holanda<sup>1</sup>

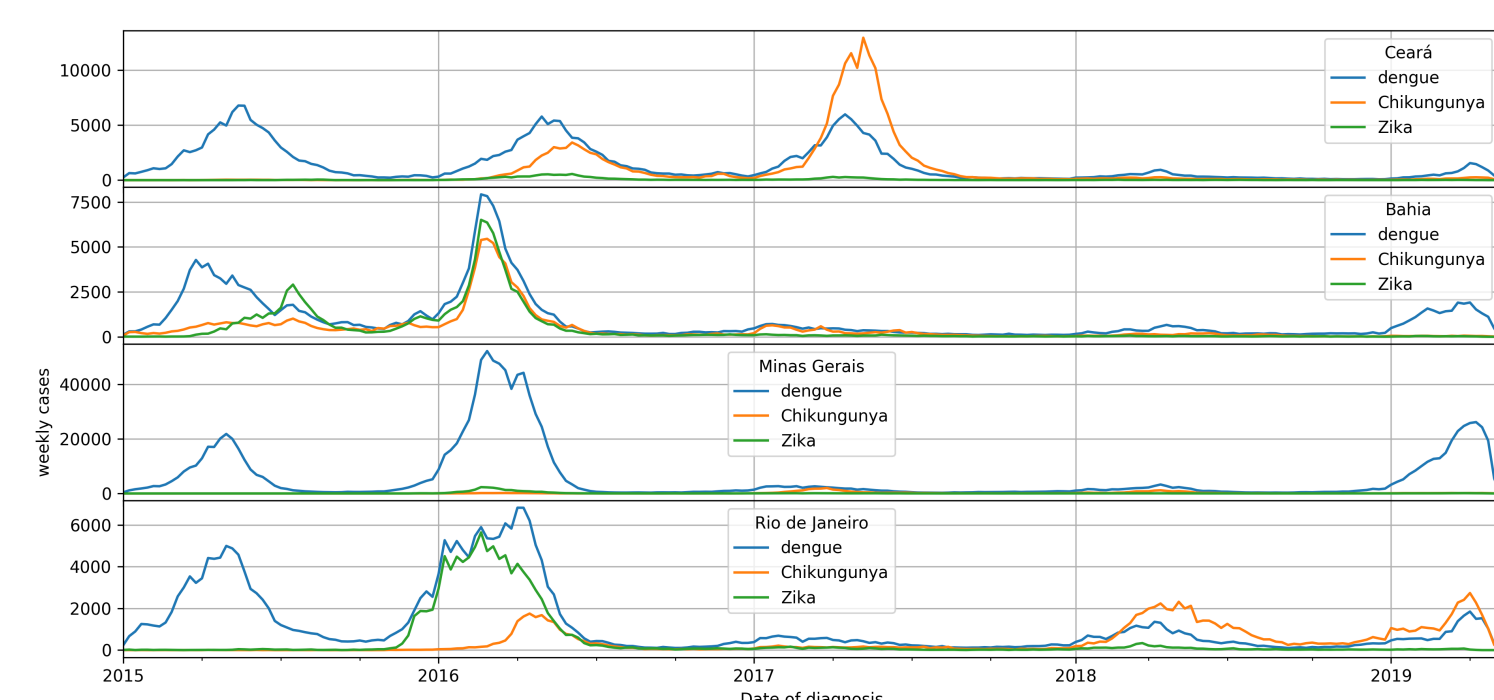
<sup>1</sup>Fundação Getulio Vargas, Brazil; Correspondence to: fc Coelho@fgv.br

Poster DOI: 10.6084/m9.figshare.9633653  
[github.com/AlertaDengue](https://github.com/AlertaDengue) — [@fccoelho](https://twitter.com/fccoelho)

FGV EMap

GEOMED  
2019 GLASGOW

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 cycle 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.



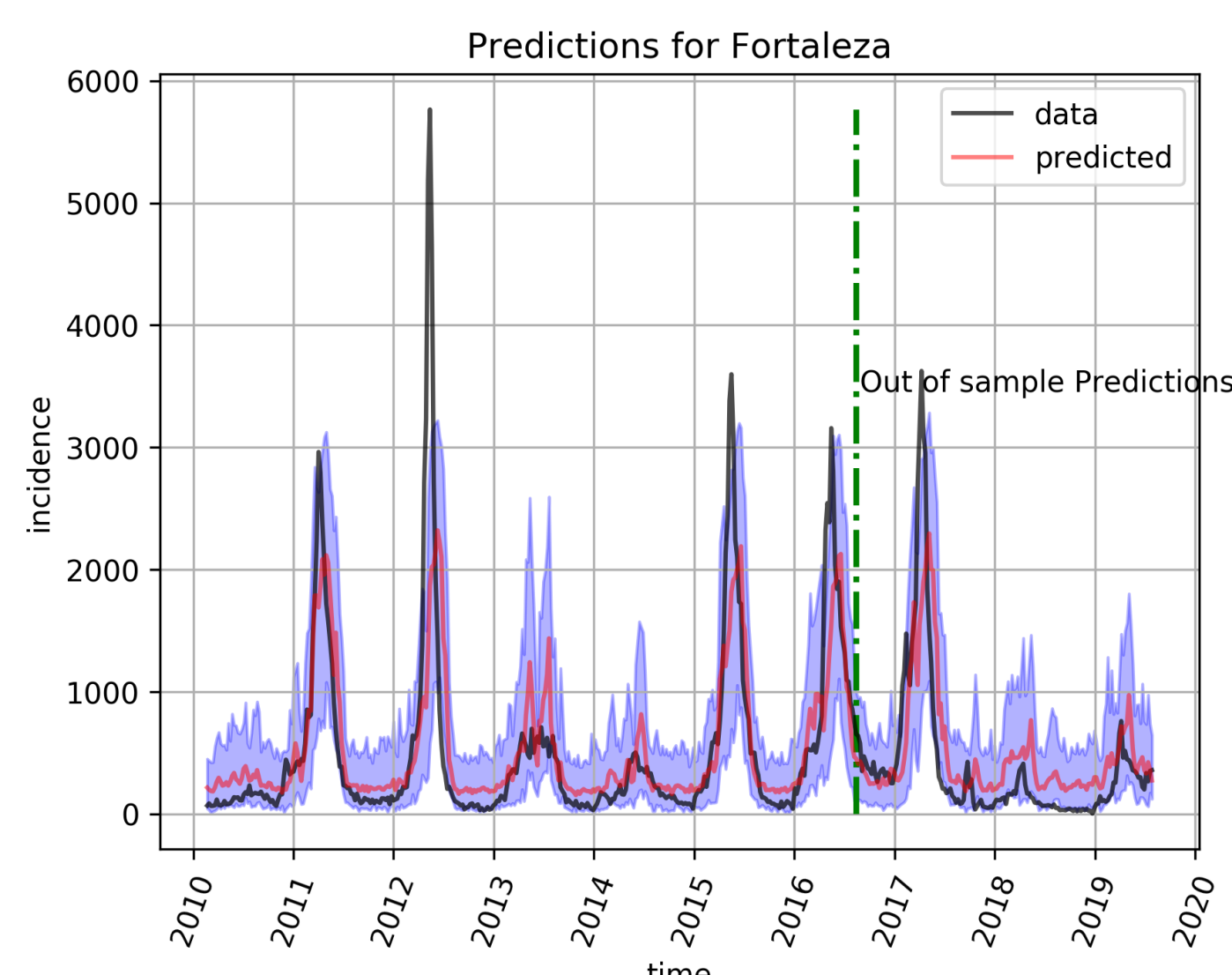
**Figure 1: Weekly incidence for the state of Ceará, Brazil.** The three arboviroses incidences are correlated.

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.

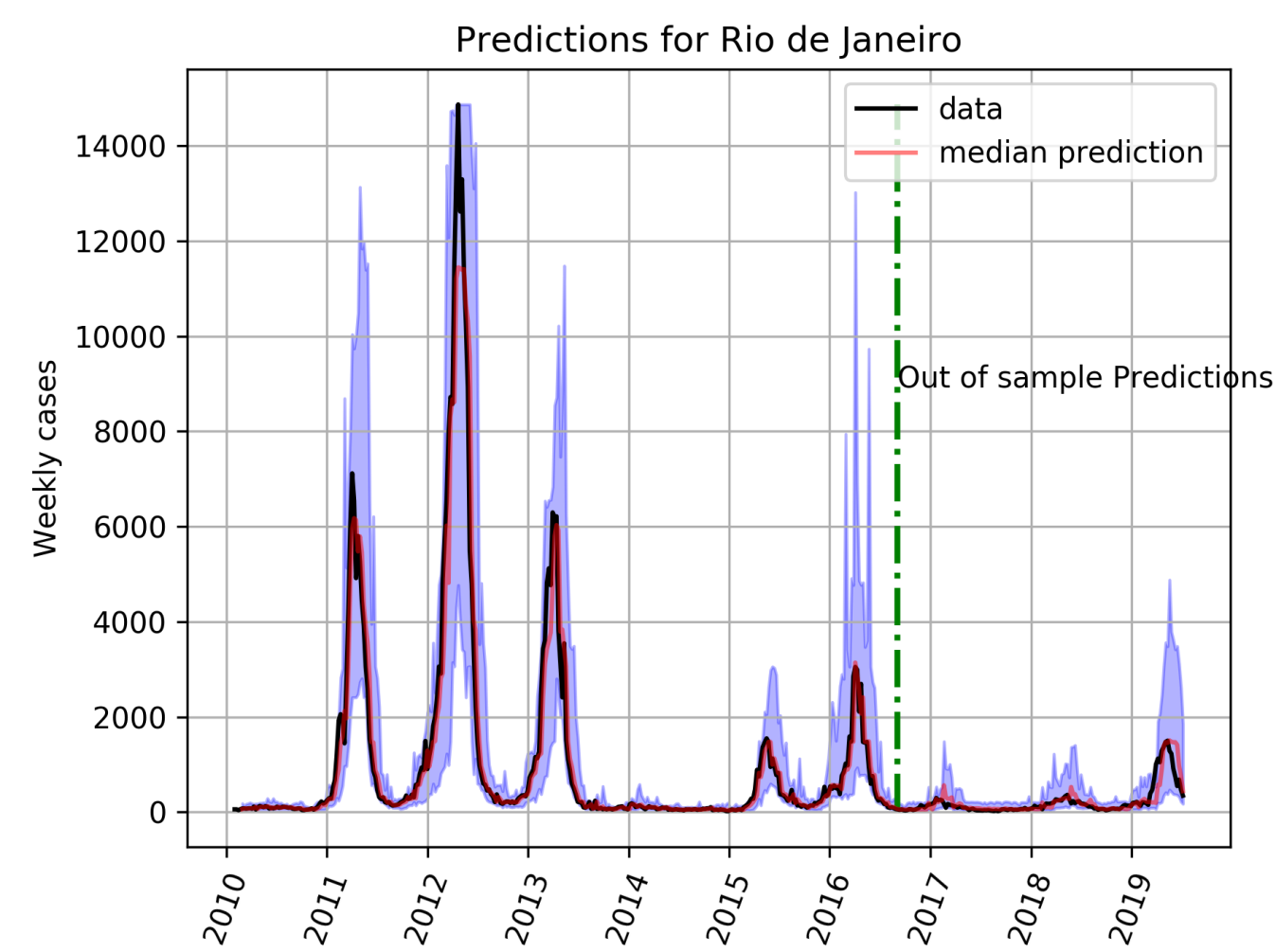


**Figure 2: LSTM forecast for Dengue incidence (black line) Prediction (red line) with 95% interval for the city of Fortaleza.** Every point in the red line corresponds to prediction from 4 weeks before.

**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

$$\hat{y}_{t+\tau} = \beta_t T_t + \epsilon_t \quad (1)$$

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

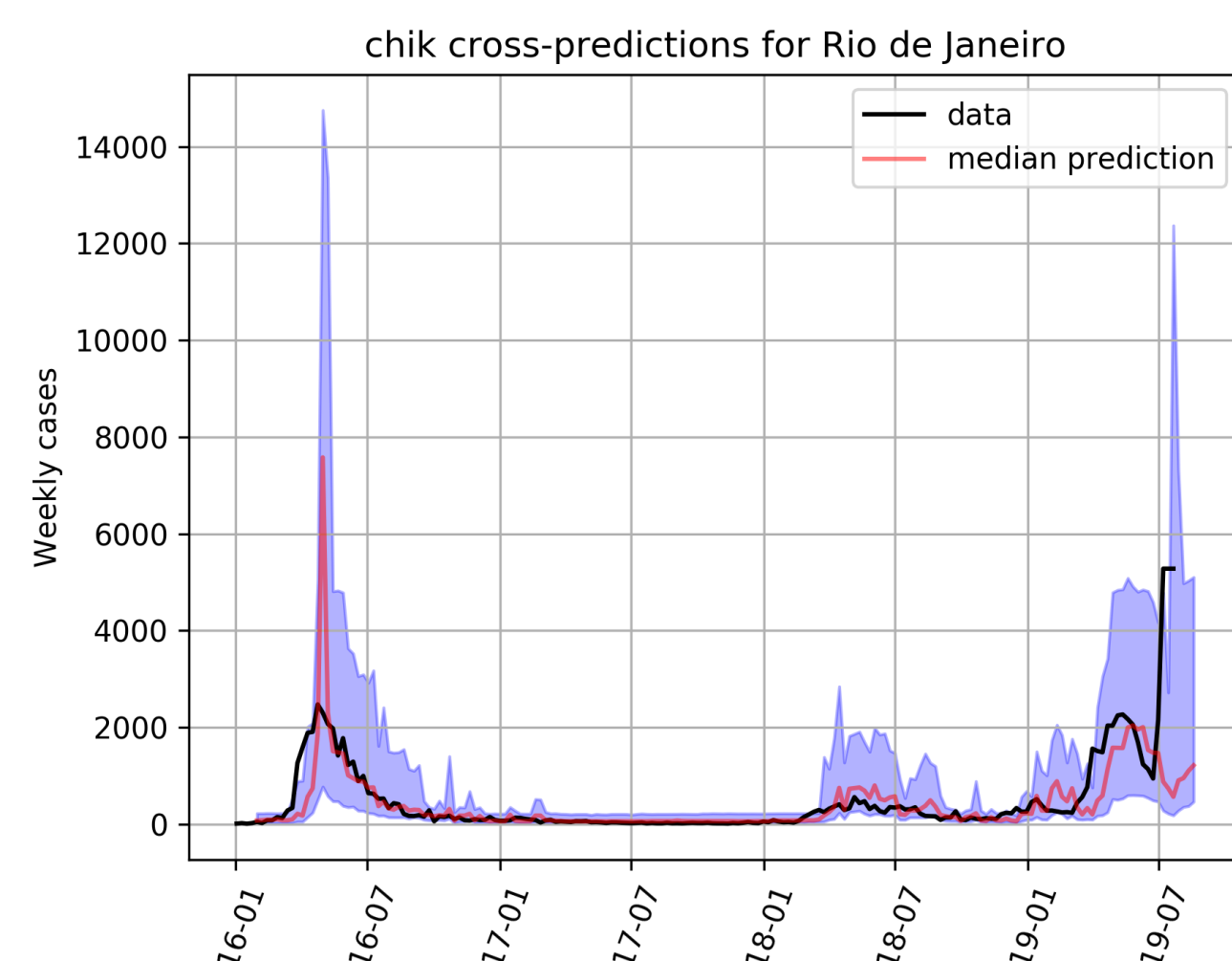


**Figure 3: Baseline forecast for dengue in Rio de Janeiro.** Random quantile forest model trained on data from 2010 to mid 2016. Red line is the median prediction, with 95% intervals in light purple.

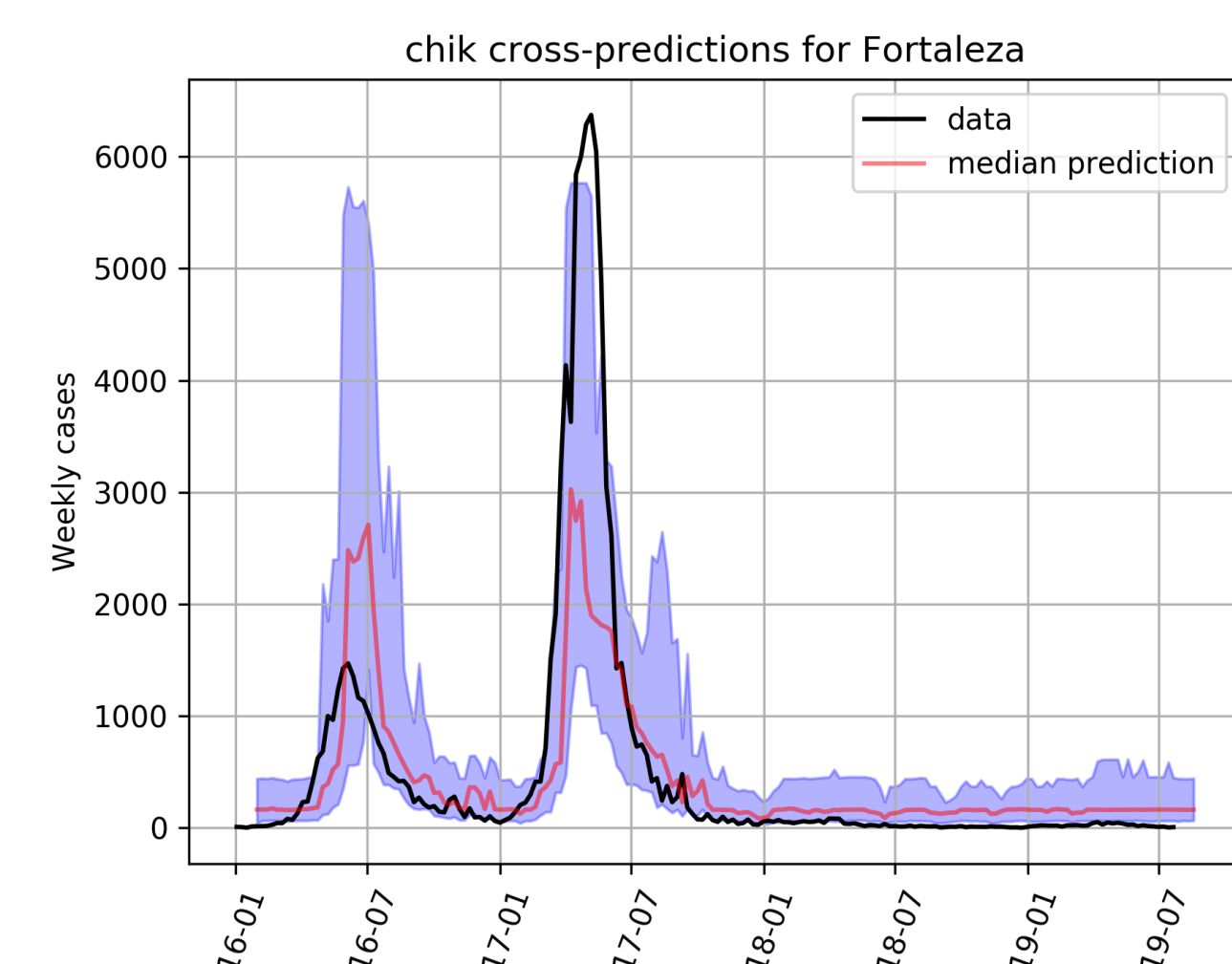
## Results

Both models were fed available Chikungunya and Zika data, yielding reasonable predictions. Below are the forecasts for Chikungunya for the cities of Rio de Janeiro and Fortaleza.

### RQF cross-predictions

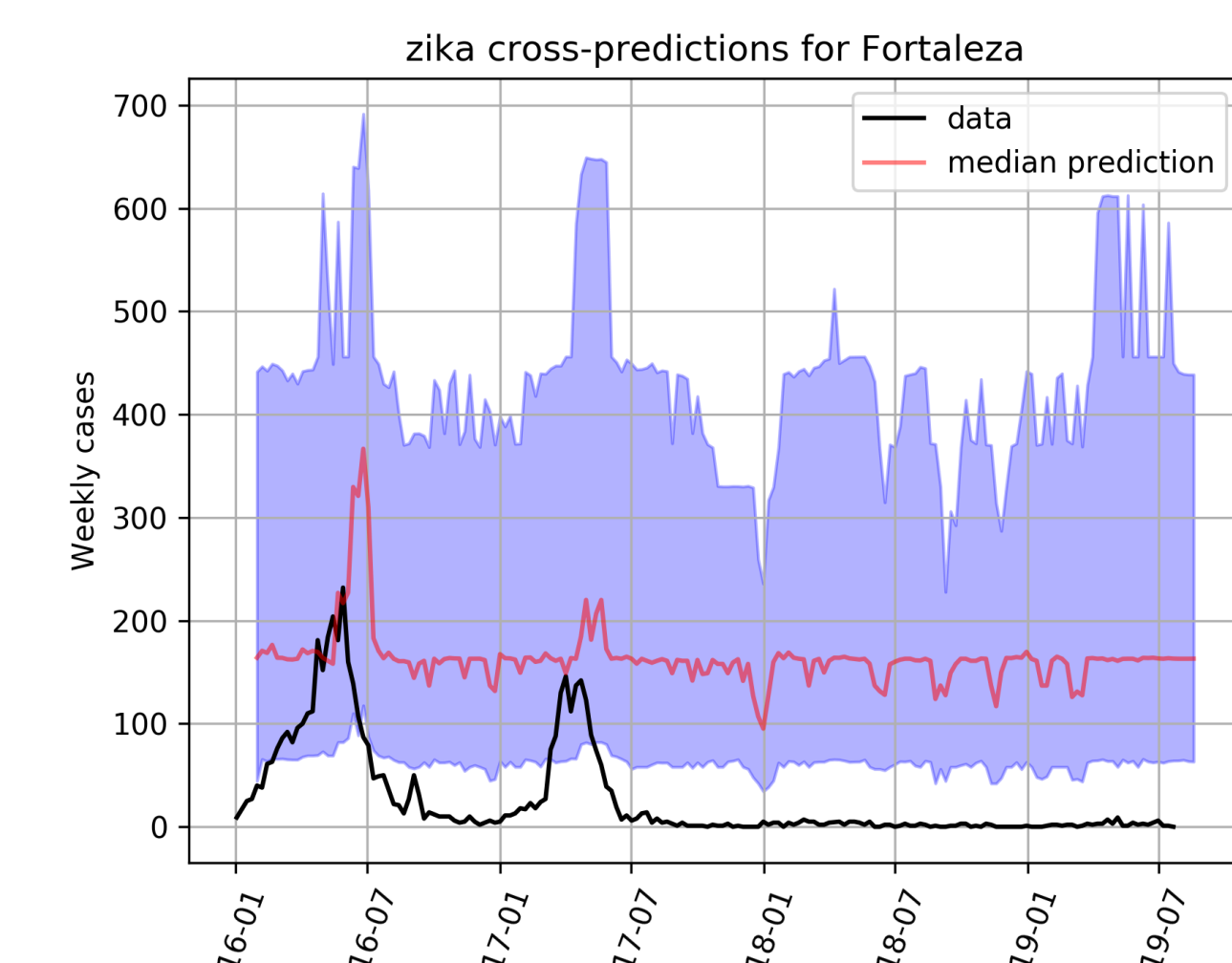


**Figure 4: Chikungunya forecasts based on dengue RQF model.** The 95% confidence intervals are wider than those for dengue data, but include the observed data.



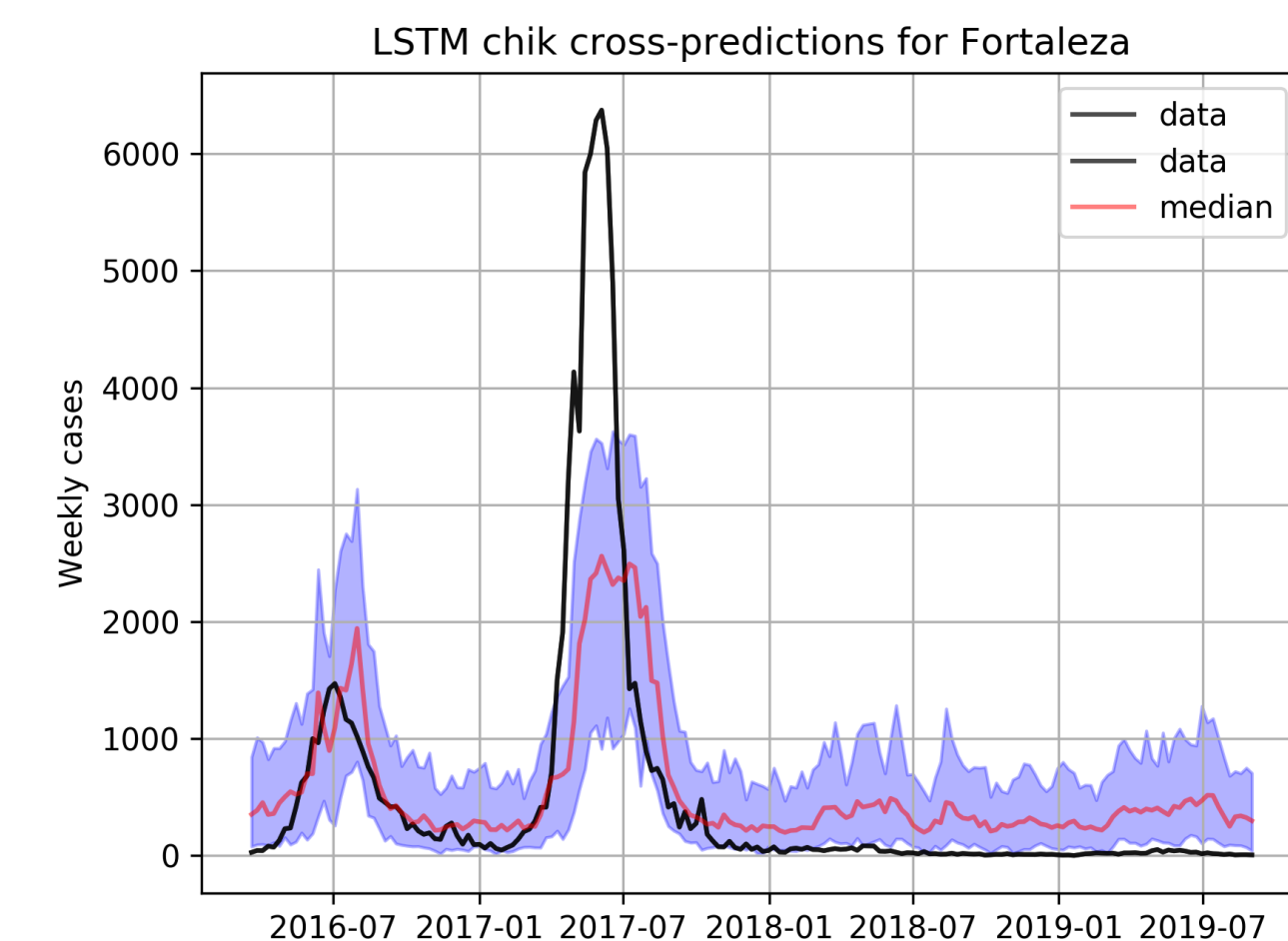
**Figure 5: Chikungunya forecasts based on dengue RQF model.** The 95% confidence intervals are wider than those for dengue data, but include the observed data.

Predictions were run from 2016 on since before this year Chikungunya and Zika did not circulate in Brazil.

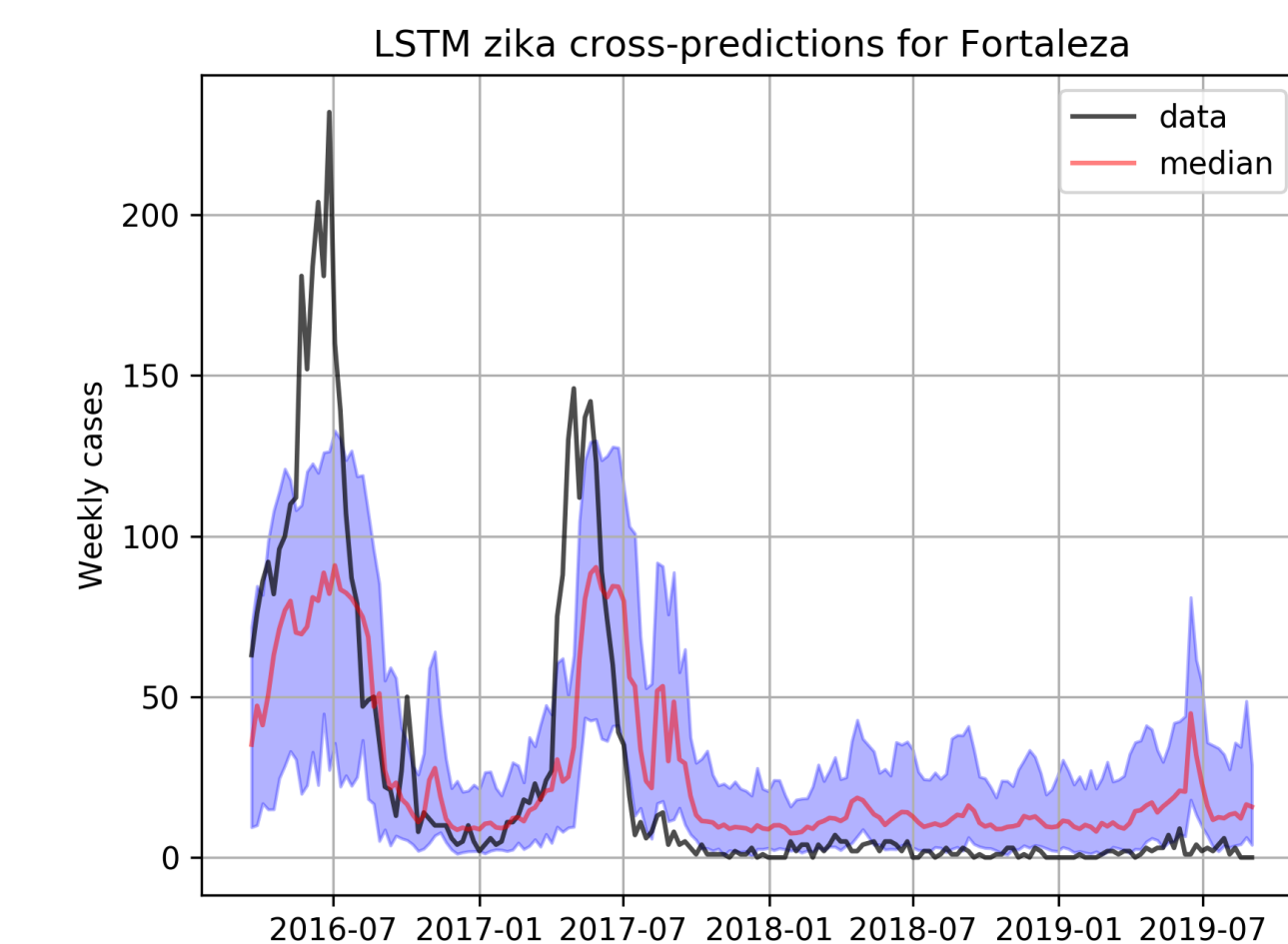


**Figure 6: Cross-predicting for Zika.** Zika forecasts based on dengue RQF, for the city of Fortaleza.

### LSTM cross-predictions



**Figure 7: Chikungunya forecasts based on dengue LSTM model.** City of Fortaleza.



**Figure 8: Zika forecasts based on dengue LSTM model.** City of Fortaleza.

## Discussion and Conclusion

Cross-disease forecasting models like the ones presented here are important because they allow us to take advantage from longer historic records from other diseases which share similar transmission mechanisms and environmental determinants.

The results presented here, expectedly show more uncertainty associated to cross-predictions when compared to forecasting the disease they were trained on. Nevertheless the models show reasonable accuracy when cross-predicting Chikungunya (figures 4 and 5), perhaps more than the same models would be capable of if they were trained on the scarce available data for Chikungunya.

For Zika, cross-predictions based of the RQF model did not work as well (figure 6). Perhaps the fact that Zika can also be transmitted sexually[1] makes its dynamic sufficiently different from dengue to make cross-predicting ineffective. The LSTM model showed similar performance when compared to RQF for Chikungunya (fig. 7). But for Zika (fig. 8) the LSTM model performed much better than RQF for cross predictions.

Machine learning models have shown great potential for infectious disease forecasting. However, long enough timeseries, essential to train such models are not easy to come by. In this work we have shown the potential of cross-disease forecasts, for diseases with similar transmission mechanisms. Despite the encouraging initial results there is still plenty of room for improvements of such models.

Supplementary figures with results for many other Brazilian cities are available from this github repository: [github.com/AlertaDengue/Geomed\\_2019](https://github.com/AlertaDengue/Geomed_2019)

## References

- [1] Flavio Codeço Coelho, Betina Durovni, Valeria Saraceni, Cristina Lemos, Claudia Torres Codeco, Sabrina Camargo, Luiz Max De Carvalho, Leonardo Bastos, Denise Arduini, Daniel AM Villela, et al. Higher incidence of Zika in adult women than adult men in Rio de Janeiro suggests a significant contribution of sexual transmission from men to women. *International Journal of Infectious Diseases*, 51:128–132, 2016.
- [2] Nicolai Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999, 2006.
- [3] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.

## Funded by

