

Dengue fever outbreak prediction

Phillip Koshute¹, Anna Buczak^{1*}, Steven Babin¹, Brian Feighner¹, Carlos Sanchez², Edwin Omar Napanga³ and Sheri Lewis¹

¹Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; ²Navy Medical Research Unit No. 6, Lima, Peru; ³General Directorate for Epidemiology, Ministry of Health, Lima, Peru

Objective

This paper addresses the problem of predicting high incidence rates of dengue fever in Peru several weeks in advance.

Introduction

Dengue fever is endemic in over 100 countries and there are an estimated 50–100 million cases annually (1). There is no vaccine for dengue fever yet, and the mortality rate of the severe form of the disease, dengue hemorrhagic fever, ranges from 10–20% but may be greater than 40% if dengue shock occurs (2). A predictive method for dengue fever would forecast when and where an outbreak will occur before its emergence. This is a challenging task, and truly predictive models for emerging infectious diseases are still in their infancy.

Methods

Predictive disease modeling attempts to exploit the complicated relationship between disease outbreaks and measurable environmental, biological, ecological and sociopolitical variables. Previous studies (3–5) identified factors associated with dengue outbreaks such as: past cases, ambient temperature, precipitation, Normalized Difference Vegetation Index, Enhanced Vegetation Index, Southern Oscillation Index, sea surface temperature anomalies and socioeconomic factors. We obtained and preprocessed these variables to get one value per district per week. The epidemiological dengue fever case data used span 2001–2009 and cover several districts in Loreto, Peru. We computed incidence rate per 1000 residents, enabling us to deal with significantly different population sizes in the different districts. The method predicts incidence rates 4 weeks in advance. We developed logistic regression (LR) models using part of the data set available. The second part, not previously used for model development, was used for testing.

The ROC curve and positive predictive value (PPV) for the test set are shown in Fig. 1. High specificity is easier to obtain

than high sensitivity. The preliminary results are encouraging: when sensitivity is 0.375, we obtain a specificity of 0.987 and a PPV of 0.6.

Conclusions

Effective methodologies to predict outbreaks of dengue fever may facilitate public health interventions to mitigate the impact of the disease. For best results, the researchers must have access to data streams with timely, detailed and accurate values of predictor variables. High PPV is of principal importance as health officials may be unlikely to spend resources on mitigation efforts based on model predictions without evidence of accuracy on past outbreaks.

Keywords

Dengue; disease prediction; logistic regression

Acknowledgments

Funding for this work is provided by the Armed Forces Health Surveillance Center (AFHSC) Division of GEIS Operations.

References

1. Guzman MG, Kouri G. Dengue: an update. *Lancet*. 2002;2:33–42.
2. Gibbons RV, Vaughn DW. Dengue: an escalating problem. *BMJ*. 2002;324:1563.
3. Lu L, Lin H, Tian L, Yang W, Sun J, Liu Q. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health*. 2009;9:395.
4. Hu W, Clements A, Williams G, Tong S. Dengue fever and El Nino/Southern Oscillation in Queensland, Australia: a time series predictive model. *Occup Environ Med*. 2010;67:307–11.
5. Choudhury Z, Banu S, Islam A. Forecasting dengue incidence in Dhaka, Bangladesh: a time series analysis. *Dengue Bull*. 2008;32:29–36.

*Anna Buczak

E-mail: anna.buczak@jhuapl.edu

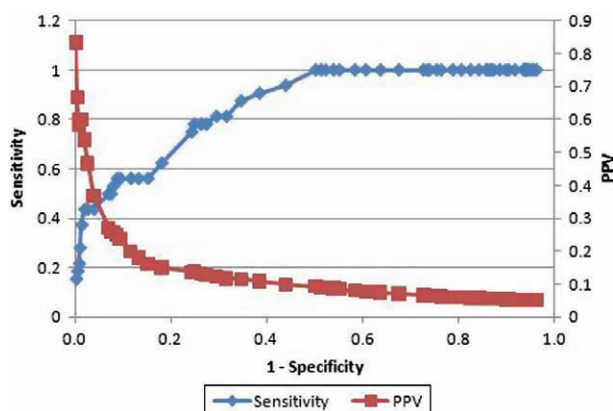


Fig. 1. ROC curve and PPV for dengue prediction.