

## A Data-Driven Compartmental Model for Dengue Forecasting and Covariate Inference

**Authors :** Yichao Liu, Peter Fransson, Julian Heidecke, Jonas Wallin, Joacim Rocklöv

**Abstract :** Dengue, a mosquito-borne viral disease, poses a significant public health challenge in endemic tropical or subtropical countries, including Sri Lanka. To reveal insights into the complexity of the dynamics of this disease and study the drivers, a comprehensive model capable of both robust forecasting and insightful inference of drivers while capturing the co-circulating of several virus strains is essential. However, existing studies mostly focus on only one aspect at a time and do not integrate and carry insights across the siloed approach. While mechanistic models are developed to capture immunity dynamics, they are often oversimplified and lack integration of all the diverse drivers of disease transmission. On the other hand, purely data-driven methods lack constraints imposed by immuno-epidemiological processes, making them prone to overfitting and inference bias. This research presents a hybrid model that combines machine learning techniques with mechanistic modelling to overcome the limitations of existing approaches. Leveraging eight years of newly reported dengue case data, along with socioeconomic factors, such as human mobility, weekly climate data from 2011 to 2018, genetic data detecting the introduction and presence of new strains, and estimates of seropositivity for different districts in Sri Lanka, we derive a data-driven vector (SEI) to human (SEIR) model across 16 regions in Sri Lanka at the weekly time scale. By conducting ablation studies, the lag effects allowing delays up to 12 weeks of time-varying climate factors were determined. The model demonstrates superior predictive performance over a pure machine learning approach when considering lead times of 5 and 10 weeks on data withheld from model fitting. It further reveals several interesting interpretable findings of drivers while adjusting for the dynamics and influences of immunity and introduction of a new strain. The study uncovers strong influences of socioeconomic variables: population density, mobility, household income and rural vs. urban population. The study reveals substantial sensitivity to the diurnal temperature range and precipitation, while mean temperature and humidity appear less important in the study location. Additionally, the model indicated sensitivity to vegetation index, both max and average. Predictions on testing data reveal high model accuracy. Overall, this study advances the knowledge of dengue transmission in Sri Lanka and demonstrates the importance of incorporating hybrid modelling techniques to use biologically informed model structures with flexible data-driven estimates of model parameters. The findings show the potential to both inference of drivers in situations of complex disease dynamics and robust forecasting models.

**Keywords :** compartmental model, climate, dengue, machine learning, social-economic

**Conference Title :** ICEID 2024 : International Conference on Emerging Infectious Diseases

**Conference Location :** London, United Kingdom

**Conference Dates :** July 29-30, 2024