Statistical modeling of the spatial distribution of dengue fever - an investigation of the validity of different selection strategies when using presence only and pseudo absence data

Ying Liu[*]

April 2014

Abstract

Dengue fever is a widely dispersed vector-borne infectious disease with an uncertain global geographic distribution. Recent studies have created risk maps identifying global risk areas of dengue transmission to enhance surveillance, control, risk awareness and local and international policies. Being a climate sensitive disease, the potential change in the risk area of dengue has also been studied under climate change projections to the end of the 21st century. The validity of prediction and projections of dengue depends on the data quality, how researchers avoid systematic bias, and the type modeling approach taken. Boosted regression tree (BRT) modeling has been credited to perform species distributions and disease presence and absence mapping. Here a BRT model is used to investigate climatic conditions and human population as possible predictors of dengue fever transmission. There are two forms of information about dengue fever utilized: presence only (PO) and pseudo absence (PA) data. The locations where dengue fever has been reported globally (totally 1537 different geographical locations) is referred to as presence only (PO) data. The set of geographical locations where dengue has not been reported constitutes a set of potential, but not confirmed is pseudo absence (PA) data. This thesis aims to 1) model the spatial distribution for dengue fever; 2) use different methods to generate pseudo absence data in order to compare how different strategies of simulating PA data affect BRT model fits and the importance of predictor variables; and 3) discuss the implications of this to risk mapping strategies of dengue. Two combinations of strategies are used to randomly select PA data. One strategy uses a selection based on the geographical distance to PO, the other strategy selects the data according to evidence based consensus regions of dengue absence. The result shows that different PA selection methods do affect the distribution of dengue. The risk maps show that the risk areas of dengue are larger under selection according to evidence-based consensus compared to selection at random.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:yili9939@gmail.com. Supervisor: Kristoffer Spricer.