Spatial and Spatio-temporal Epidemiology 10 (2014) 75–84

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees

Yoon Ling Cheong [a,b,*], Pedro J. Leitão [a], Tobia Lakes [a]

[a] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[b] Medical Research Resource Centre, Institute for Medical Research, Jalan Pahang, 50588 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Available online 11 June 2014

Keywords:
Aedes mosquitoes
Boosted Regression Trees
Horticulture
Human settlements
Environmental
Land use

ABSTRACT

The transmission of dengue disease is influenced by complex interactions among vector, host and virus. Land use such as water bodies or certain agricultural practices have been identified as likely risk factors for dengue because of the provision of suitable habitats for the vector. Many studies have focused on the land use factors of dengue vector abundance in small areas but have not yet studied the relationship between land use factors and dengue cases for large regions. This study aims to clarify if land use factors other than human settlements, e.g. different types of agricultural land use, water bodies and forest are associated with reported dengue cases from 2008 to 2010 in the state of Selangor, Malaysia. From the correlative relationship, we aim to generate a prediction risk map. We used Boosted Regression Trees (BRT) to account for nonlinearities and interactions between the factors with high predictive accuracies. Our model with a cross-validated performance score (Area Under the Receiver Operator Characteristic Curve, ROC AUC) of 0.81 showed that the most important land use factors are human settlements (model importance of 39.2%), followed by water bodies (16.1%), mixed horticulture (8.7%), open land (7.5%) and neglected grassland (6.7%). A risk map after 100 model runs with a cross-validated ROC AUC mean of 0.81 (±0.001 s.d.) is presented. Our findings may be an important asset for improving surveillance and control interventions for dengue.

1. Introduction

Dengue fever (DF) and dengue haemorrhagic fever (DHF) are the most important vector-borne diseases (VBD) in tropical areas (Gubler, 2006). In recent decades the risk of dengue infection has increased dramatically not only in tropical, but also in sub-tropical regions (World Health Organization, 2012). There are between 50 and 100 million dengue infections every year, and more than 500,000 cases are hospitalized (Gubler, 2006). Dengue transmission is influenced by a complex set of factors including the environment, climate and weather, human behavior and dengue virus serotype-specific herd immunity among the human population (Cheong et al., 2013; Halstead, 2008; Hay et al., 2000). Understanding the association between environmental factors and VBD is essential for better preventing and controlling disease transmission (Armien et al., 2008; Dambach et al., 2009).

Abbreviations: DF, dengue fever; DHF, dengue haemorrhagic fever; VBD, vector-borne diseases; BRT, Boosted Regression Trees; ELISA, IgM capture enzyme-linked immunosorbent assay; API, Application Programming Interface; GLM, generalized linear model; GAM, generalized addictive model; DV, dengue virus; lr, learning rate; tc, tree complexity; ROC AUC, Receiver Operating Characteristic Area Under the Curve; IVM, Integrated Vector Management.

* Corresponding author at: Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Tel.: +49 30 2093 6827; fax: +49 30 2093 6848.

E-mail addresses: yoon.ling.cheong@geo.hu-berlin.de (Y.L. Cheong), p.leitao@geo.hu-berlin.de (P.J. Leitão), tobia.lakes@geo.hu-berlin.de (T. Lakes).

http://dx.doi.org/10.1016/j.sste.2014.05.002
1877-5845/

In many countries, such as in Malaysia, dengue disease has been endemic (Halstead, 2008). Since the first reported case in Malaysia in 1901 (Skae, 1902) peaks of dengue cases have been reported in 1974, 1978, 1982 and 1990, and the total number of dengue cases has increased (Lam, 1993). All four serotypes circulated, the predominant serotype identified was DEN-3 from 1992 to 1995, DEN-1, DEN-2 and DEN-3 alternated in recent years (Arima and Matsui, 2011). According to the Ministry of Health Malaysia, the incidence rate of DF was the highest ever in 2010 (148.73 per 100,000 population) and the mortality rate of DHF was 0.42 per 100,000 population (Ministry of Health Malaysia, 2011). These figures significantly exceed the national target of Malaysia, which aims for the incidence rate of DF to be less than 50 cases per 100,000 population (Ministry of Health Malaysia, 2011).

Dengue disease is often called an "urban" disease, with the Aedes mosquitoes, both *Aedes aegypti* and *Aedes albopictus,* mainly found breeding in artificial containers in areas where high population density appears (Chen et al., 2005; Gubler and Clark, 1995). However, *A. albopictus* has also been found in natural environments (Gubler, 1998; World Health Organization, 2008). Entomological studies showed that dengue vectors have been captured in vegetated areas (Hayden et al., 2010; Vezzani et al., 2005), orchards (Vanwambeke et al., 2007b), rubber plantations (Paily et al., 2013; Sumodan, 2003), marshy swamp (Sarfraz et al., 2012) and even in brackish waters (Idris et al., 2013; Ramasamy et al., 2011). In a study of dengue seroprevalence taken from the national database of the Malaysian cohort study of 2008 (Muhammad Azami et al., 2011), from the 1000 randomly selected adults aged 35–74, there was no significant difference in the seroprevalence rate between adults living in urban and rural areas. Therefore, diverse land use types in the neighborhood of human settlements may also provide a suitable habitat for Aedes mosquitoes. Although human settlements have been identified in earlier studies as being highly associated with dengue cases (Gubler, 2006), we hypothesized that other land use types are also associated with the occurrences of dengue cases. In this study, our research questions are:

(i) Which land use factors are associated with dengue cases in Selangor state, Malaysia?
(ii) What is the spatial pattern of dengue risk based on the identified correlative relationships?

Many studies have focused on the relationship between Aedes mosquitoes density and environmental factors in specific local hotspots of dengue (Chen et al., 2005; Rohani et al., 2001; Wan-Norafikah et al., 2012). There are also some studies that have evaluated the spatial risk factors of dengue cases on a local scale (Nazri et al., 2009; Shafie, 2011). To our knowledge, this is the first study assessing the land use factors associated with dengue cases in Malaysia on a state-wide level.

2. Data and methods

2.1. Study area

Selangor covers an area of 7930.20 km^2 and is located between 2°35′N to 3°60′N and 100°43′E to 102°5′E (Fig. 1). We selected the state of Selangor because of its large number of dengue cases, which accounted for 41.1% (56,305 cases) of all reported dengue cases in Malaysia (Department of Statistics Malaysia, 2011). Secondly, Selangor has the highest population density and gross domestic product per capita of all states in Malaysia (Abdullah and Nakagoshi, 2006). Thirdly, Selangor is geographically heterogeneous and displays a large variety of land use and related environmental characteristics with large urban areas, agricultural use, forests and wetlands (Abdullah and Nakagoshi, 2006).

2.2. Data sources

We obtained dengue data for the state of Selangor for the years 2008, 2009 and 2010 from the Disease Control Division, Ministry of Health Malaysia. We used only those dengue cases that were confirmed by the serological tests IgM capture enzyme-linked immunosorbent assay (ELISA) with single positive IgM, following earlier studies (Chadwick et al., 2006; Krishnan et al., 2012).

We obtained a land use map of the year 2006 from the Department of Agriculture Malaysia. This map is based on imagery from Landsat 7 (30 m × 30 m resolution) and SPOT 4 (20 m × 20 m resolution) satellite imagery, a topographic map (L7030, 1:50,000) and reference data from field trips. According to the official accuracy assessment, less than 5% of the land use classes have been incorrectly classified.

2.3. Explanatory variables

We selected 15 land use variables that could be associated with suitable habitats for Aedes mosquitoes according to a detailed literature review and dengue expert knowledge. The land use variables are coconut and cocoa plantation, animal husbandry, mixed horticulture, orchard and farm, tea plantation, mining, oil palm plantation, neglected grassland, rubber plantation, paddy field, swamp forest, forest, open land, human settlements and water bodies. The detailed map of each land use variable is shown in Supplementary Fig. 1.

In coconut and cocoa plantations, natural breeding habitats such as plant axils, coconut husks, coconut shells and coconut floral spathes containing organic debris were identified as suitable habitats for Aedes mosquitoes (Chareonviriyaphap et al., 2004; Rohani et al., 2001; Thavara et al., 2001). Land being used for animal husbandry is a potential habitat for the dengue vector as *A. albopictus* also feed on domestic chickens (Richards et al., 2006). Mixed horticulture describes the area of mixed cultivation of gardens, orchards and nurseries with