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

CrossMark

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

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).

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 statewide 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
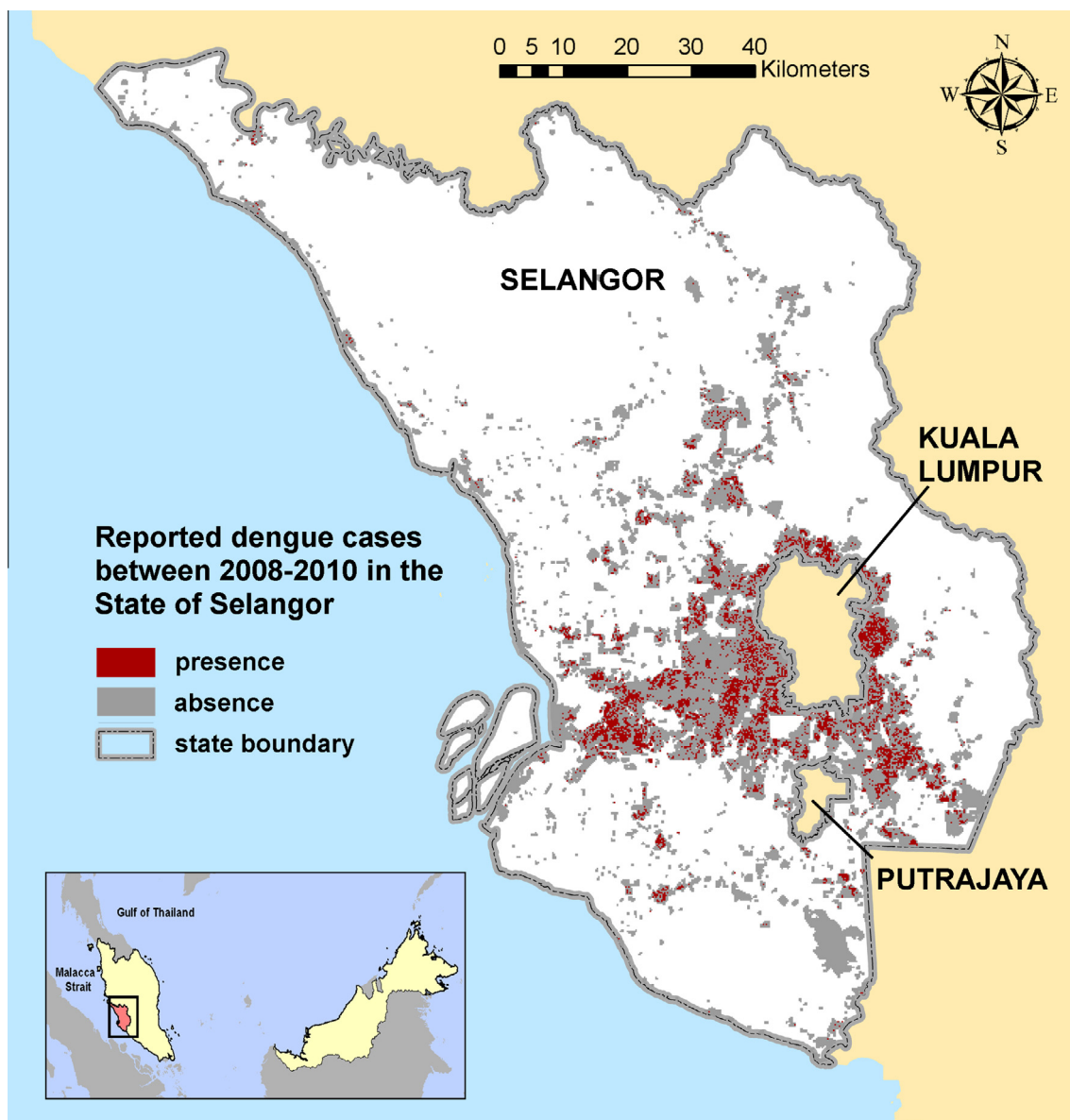
Fig. 1. Study area: State of Selangor, Malaysia with presence and absence of dengue cases in 200 m × 200 m grid cells.

flowers, fruits, vegetables and ornamental plants surrounding the human settlements where mosquitoes find several suitable breeding possibilities (Nazri et al., 2009). Furthermore, Aedes mosquitoes have been found in ornamental plants (Wan-Norafikah et al., 2012) including the water-holding bromeliad axils (Mocellin et al., 2009) in mixed horticulture. The presence of fruit trees and vegetable farms – which is classified here as orchard and farm land including vegetation farming, fruit trees and mixed cropland – has been seen to increase the probability of finding Aedes mosquitoes (Chareonviriyaphap et al., 2004; Vanwambeke et al., 2007b). In tea plantation areas Aedes mosquitoes have been found in discarded tires, tins and plastic cups near human inhabitations (Amala and Anuradha, 2011).

Aedes mosquitoes have also been found in old mining areas, in particular in flooded disused gold mine shafts and wells (Russell et al., 1996). They have been identified in suburban areas surrounded by oil palm plantation (Ponlawat and Harrington, 2005) and in the neglected grassland of a cemetery (Vezzani et al., 2005). Furthermore, Aedes mosquitoes have also been found in rubber plantations, particularly in the rubber tree hole (Paily et al., 2013), in suspended sap-collecting containers without rain guards (Sumodan, 2003) and discarded water storage containers (Chareonviriyaphap et al., 2004). Paddy fields and swamp forest containing standing water from 2.5 cm to 30 cm in depth serve as a rainfall catchment area, and are another suitable habitat for the dengue vector (Sarfraz et al., 2012). Aedes mosquitoes have been

captured from sites inside the forest, as far as 1 km from houses (Lourenco-de-Oliveira et al., 2004), and in tree holes, bamboo stumps and rock pools in the forest area (Amala and Anuradha, 2011).

Artificial water containers that provide nutrients for the Aedes mosquitoes have been found in open land, i.e. areas that are ready for construction or agriculture and have only little or no vegetation coverage with a dominance of sand and rocks (Chang et al., 1997; Dom et al., 2012; Gleiser and Zalazar, 2010). In several studies, Aedes mosquitoes were identified in human settlements where they were found to breed in a variety of water filled containers (Nyamah et al., 2010). Another favorable dengue vector habitat is natural reservoirs including various types of water bodies, e.g. roadside ponds, pools, streams and other containers found on river banks (Uribe et al., 2008).

2.4. Data preprocessing

In a first step, we geocoded the addresses of the dengue cases with totals of 11,664, 10,482 and 8300 for the years 2008, 2009 and 2010. The geocoding accounted for 92.16% (2008), 93.44% (2009) and 94.87% (2010) of the addresses; the others had to be excluded due to missing values. We used the Google Maps Application Programming Interface (API) that showed high quality geocoding services (Cui, 2013; Roongpiboonsopit and Karimi, 2010) and has successfully been used for the geocoding of health care facility locations (Gu et al., 2010) and mosquito survey in earlier studies (Neteler et al., 2011).

Secondly, we calculated a systematic 200 m grid of the projected coordinate system Kertau (RSO)/RSO Malaya (m) for the state of Selangor. This follows the official definition of the Ministry of Health Malaysia that a dengue outbreak is constituted by an incidence of two or more dengue cases in a location where the onset dates of the cases are less than 14 days apart and the cases are within 200 m of each other (MoHM, 1986, MoHM, 2009). Next, only those grids that are at least partly covered by human settlements were included in the model. We hereby avoid a modeling bias of dengue absence data in areas where no human settlement and therefore no potential cases could be recorded because we analyse the patient's home address that falls into the human settlements class of land use. We then classified 6344 grid cells as presence data (1) where at least one dengue case was reported between 2008 to 2010. The remaining 29,662 grid cells with a certain proportion of human settlements but no reported dengue cases were considered as absence (0) data. Hence, our target, i.e. response variable, is a binary presence/absence value of at least one reported dengue case in a 200 m grid cell that includes human settlement area.

We then used the focal statistics geoprocessing function of ArcGIS's Spatial Analyst (Environmental Systems Research Institute, Inc.) to obtain the proportion of coverage of each land use factor (Supplementary Fig. 1) within a rectangular moving window of 1 km radius surrounding each grid cell for the spatial neighborhood analysis. The radius was used in order to consider the maximum flight range of the mosquito during its life span and the potential dengue virus (DV) transmission with human mobility that has also been applied in other studies (Maciel-de-Freitas et al., 2006; Sarfraz et al., 2012). After data compilation, and in order to avoid using highly correlated variables in the subsequent models, we checked for the pair-wise Spearman rank correlation between all predictor variables.

2.5. Boosted Regression Tree analysis

To identify the association between the land use factors and the reported dengue cases we apply Boosted Regression Trees (BRT), a machine learning algorithm developed by Friedman et al. (2000). BRT are very flexible and capable of dealing with complex responses, including nonlinearities and interactions (Elith et al., 2008). Compared to generalized linear model (GLM) and generalized addictive model (GAM) models they have been shown to yield better results e.g. in terms of prediction capability (Coutts and Yokomizo, 2014; Oppel et al., 2012). BRT have performed well in disease modeling (Stevens and Pfeiffer, 2011) and have been applied in various study domains, from predicting the distribution of organism (Elith et al., 2008), to comparing the factors of cropland abandonment (Müller et al., 2013). We applied the BRT using the R version 3.0.2 (R Development Core Team, 2013), package "dismo" version 0.8–17 (Hijmans et al., 2013) and package "gbm" version 2.1 (Ridgeway, 2013).

BRT combine boosting and regression trees in a single algorithm. Regression trees grow with recursively binary splits of the data until some stopping criterion is met. Despite its easier visual interpretation of the interactions, capability to adapt predictor variables of any type (numeric, binary, categorical, etc.), and insensitivity to outliers, decision trees are prone to bias and not as accurate as other statistical methods (e.g. GAM). Boosting based on the stochastic gradient boosting (Friedman, 2002) improves the decision trees by minimizing the loss function (deviance) at each tree split. Stochastic gradient boosting applies randomness into the sequential fitting (De'ath, 2007). We use a bagging factor of 0.5 as suggested by Friedman (2001). The stagewise approach means that the first regression tree is fitted to the dependent variable while randomly selecting 50% of the training data (bagging). The following tree is then fitted to the residuals of the first tree and the fitted values are added to the logit of the fitted probability. The model is then updated to contain two trees and the residuals from this updated model are calculated. This process is then further iterated until reaching the defined stopping criterion (Elith et al., 2008).

The model building process is best with slower learning rates (lr), also known as shrinking rate. The number of splits levels of each tree, also called tree complexity (tc), determines the degree to which predictors may interact with each other in relation to the response. More levels of interactions are explained with a higher tc. In this study, we fit the model with a "bernoulli" distribution as we are dealing with presence and absence data, as explained above. We test for several combinations of the lr (0.025, 0.05, 0.1) and tc (3, 4, 6, 8, 9, 10) parameters. The model with the highest cross-validated Receiver Operating
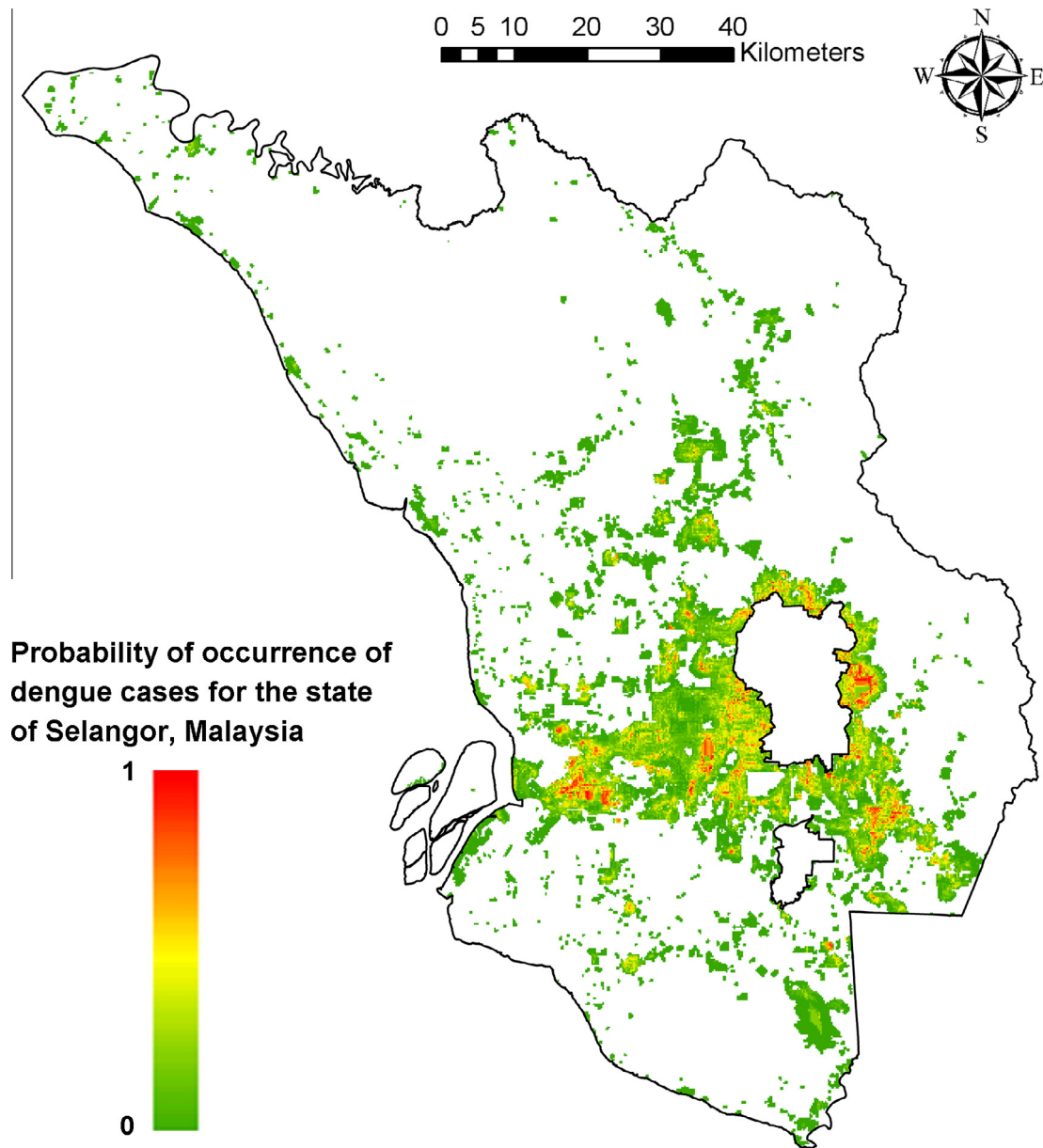
Fig. 2. Probability of dengue occurrence at 200 m × 200 m spatial resolution of the mean predicted map (area under the receiver operator curve of 0.81 (±0.001 s.d.)) from 100 boosted regression tree model iterations. High probability of dengue occurrences area is shown in red and low probability of dengue occurrences area in green.

Characteristic Area Under the Curve (ROC AUC) score was selected as the most optimal settings (Hastie et al., 2009).

To obtain a more valid and reliable result, the BRT model runs were repeated in 100 iterations (Bhatt et al., 2013). The mean and standard deviation of the ROC AUC scores among the 100 iterations were reported. We then calculated a risk map to show the spatial pattern of probability of occurrence of dengue cases on the 200 m grid level for the state of Selangor, Malaysia. We therefore get the mean of the 100 iterations of prediction value. As a result we receive a risk value scaled between 0 and 1 where 0 means low risk for dengue and 1 means high risk

for dengue based on the identified associations between land use variables and dengue cases.

BRT delivers the relative importance of the predictor variables in the model (Friedman, 2001). The relative importance is based on the number of times a variable is selected for splitting, weighted by the squared improvement and averaged over all trees (Friedman and Meulman, 2003). The relative importance of each variable is then scaled so that the sum adds to 100 as percentages. A higher percentage of a variable indicates a stronger relative importance of this variable on the response. With 15 predictor variables in our model, only the relative contribution above

the randomness threshold (100%/number of predictor variables, 100%/15 = 6.67%) can be interpreted (Müller et al., 2013). To better interpret the fitted functions, we used partial dependence plots to visualize the relative importance of the predictor variables on dengue cases (Friedman, 2001), as well as rug plots to show the percentile distribution of the response variables. To aid interpretation, all plots were smoothed using a spline function.

3. Results

From the correlation analysis we found no high rank correlations (rho > 0.70) between any pair of predictor variables so that we continued with a model that included all land use variables. The selected model was fitted using a *lr* of 0.05, *tc* of 9 and a total of 2150 decision trees, and showed a cross-validated ROC AUC score of 0.81. The ensemble prediction map after the 100 iterations resulted in a performance ROC AUC score of 0.81 (±0.001 s.d.). These predictions were represented in the risk map of dengue where we identify distinct spatial patterns of different probabilities of occurrence of dengue cases for each grid cell (Fig. 2). The spatial patterns of predicted dengue cases overall coincide very well with the observed cases that are shown in Fig. 1. Most of the reported high dengue risk areas are identified, while only few are not correctly identified. Our predictions of high dengue risk (Fig. 2) seems to be more focused on the central region of Selangor, with clusters at the border to Kuala Lumpur, west and southeast of Selangor. The presence of dengue cases that were scattered in the northern and southern part of Selangor (Fig. 1) is mainly predicted with low dengue risk (Fig. 2). Besides, a large share of the presence of dengue cases in the central region of Selangor is predicted with the probability of occurrences of dengue cases between high and low risk (Fig. 2).

In addition to the spatial pattern of probability of occurrence of dengue cases we identified the relative importance of the predictors in the BRT model (Fig. 3) and the correspondent partial dependence plots for each predictor and the risk of dengue presence (Fig. 4). The selected model demonstrated that only 5 out of 15 predictor variables had a considerable importance, i.e. above the randomness threshold of 6.67% (Fig. 3). The partial dependence plots in Fig. 4 further show the influence of each of the five predictor variables on the dengue occurrences, taken that every other variable has been kept to its mean value. We observed that the proportion of cover of human settlements in a grid cell substantially contributed to the models with a relative importance of 39.2% (Fig. 3), with a nearly linear increasing association with the occurrence of dengue cases (Fig. 4(a)). Water bodies including drains, lakes and rivers followed as the second most important predictor of the dengue cases (relative importance of 16.1%) (Fig. 3) with a slightly decreasing non-linear association (Fig. 4(b)). The model indicates an optimum for the occurrence of dengue in a grid cell with the presence of water bodies of around the value of 25 to 50 ha of area (ca. 5 to 10% of the 1 km grid cell) (Fig. 4(b)). Mixed horticulture, normally related to the suburban orchards and gardens surrounding the human
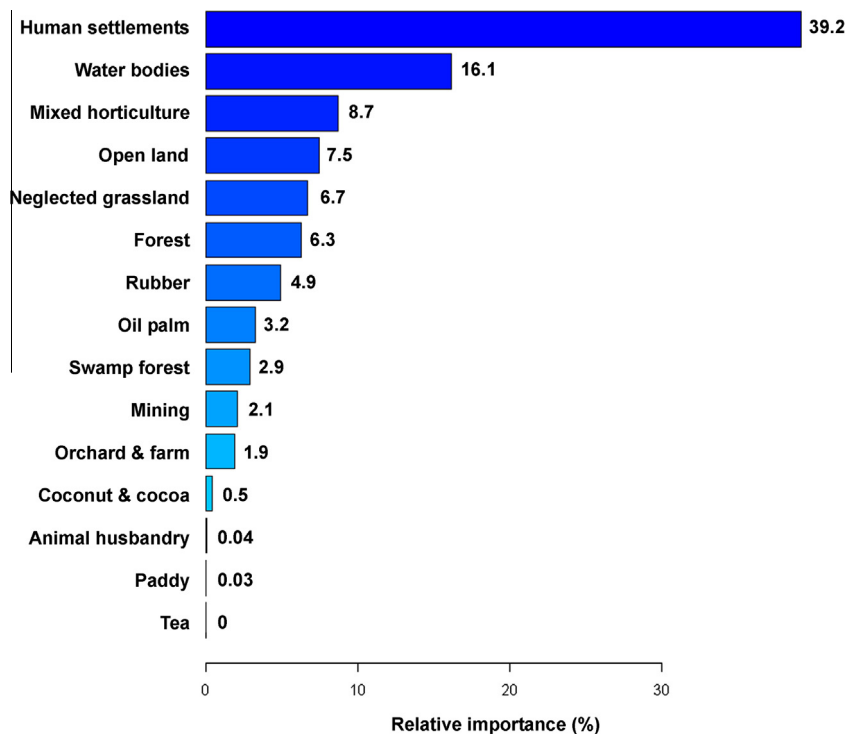

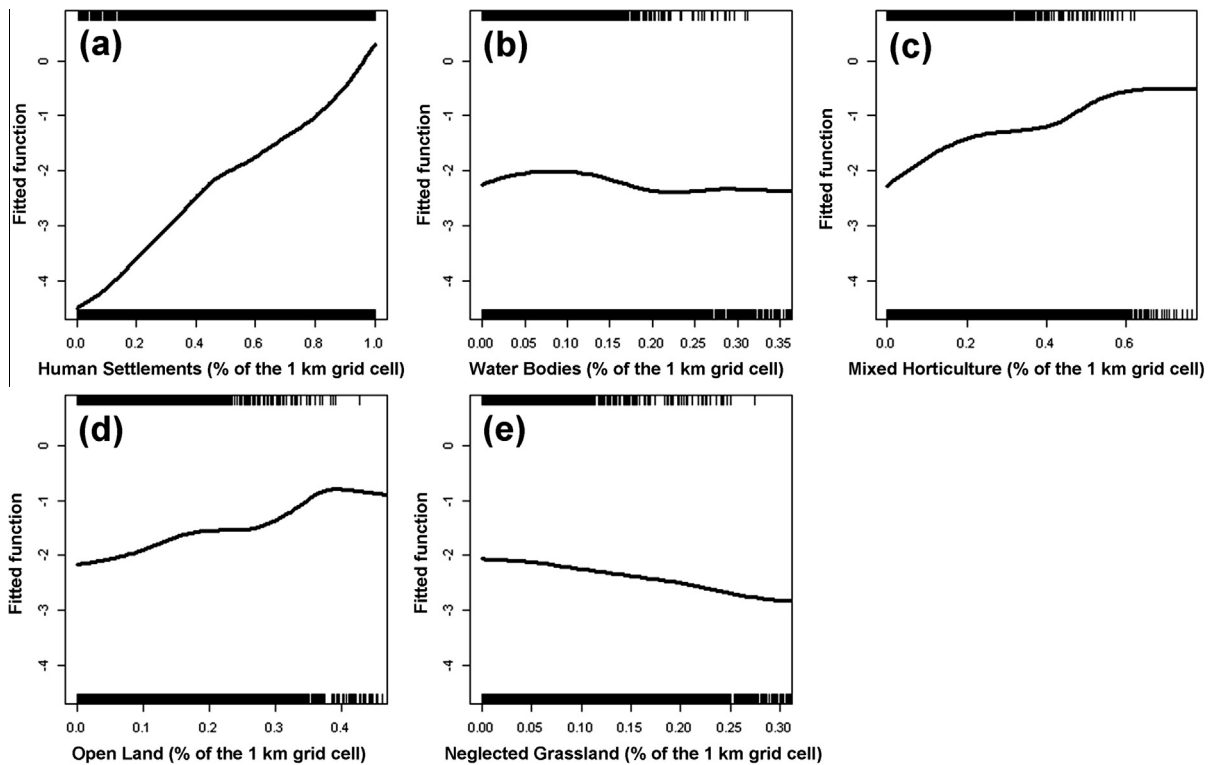**Fig. 3.** Relative importance of explanatory variables.

Fig. 4. Smoothed partial dependence plots for the five most influential land use factors for dengue cases; (a) human settlements, (b) water bodies, (c) mixed horticulture, (d) open land, (e) neglected grassland. The density of presence and absence of dengue cases are represented in rug plots on the top and bottom along the variable axis.

settlements, showed a relative importance of 8.7%, with an increasing non-linear association with the reported dengue cases (Figs. 3 and 4(c)). The coverage of mixed horticulture, particularly above ca. 200 ha (ca. 40% of the 1 km grid cell) depicted a sudden increase of dengue risk. Similarly, the proportion of area covered by open land, with a relative importance of 7.5% (Fig. 3), also showed an increasing non-linear association with the occurrence of dengue cases (Fig. 4(d)), with particularly higher dengue risk in grid cells with ca. 150–200 ha of open land (ca. 30% to 40% of the 1 km grid cell). Finally, neglected grassland, with a relative importance of 6.7% (Fig. 3) in the model, showed a decreasing linear association with the occurrence of dengue (Fig. 4(e)).

The other predictor variables with relative importance below the randomness threshold are coconut and cocoa, animal husbandry, orchard and farm, tea plantation, mining area, oil palm, rubber plantation, paddy field, swamp forest, and forest. The dependency profiles for these predictor variables are not shown in the Fig. 4.

4. Discussions

In this study, we aimed to characterize associations between different land use factors and dengue cases in Selangor state, Malaysia. Moreover, we wished to show the spatial patterns of dengue risk based on the identified correlative relationships. Our results demonstrate that we are able to explain the different associations between land use factors and dengue cases from a state level perspective.

The study showed that human settlements and non-agricultural areas largely determined the occurrence of dengue cases. Larger shares of human settlement coverage in the neighborhood are associated with higher numbers of dengue cases. One of the reasons for this may be higher population density in areas with more human settlements, leading to higher human biting rates. Increased human biting rates offer opportunities for the Aedes mosquitoes to acquire DV by biting an infected person and then to transmit the virus after becoming infected (Scott and Morrison, 2010). Our findings are in line with those of other authors, such as e.g. a study in one suburban city in Selangor, Subang Jaya, where they found that most dengue cases occurred in urban areas (Nazri et al., 2009). The habitats of Aedes mosquitoes were found in artificial containers in areas related to human activities, but disparities were identified between types of human settlements (Vanwambeke et al., 2007a). Although no variation of human settlements was investigated in this study, previous entomological surveillance has provided insightful information. Aedes mosquitoes were found indoors and outdoors in human settlements, especially in culverts, water compartments, metal drums, plastic drums, pails, gully traps, discarded containers, construction sites, solid-waste dumps, open spaces and factories (Chen et al., 2009; Nyamah et al., 2010; Teng and Singh, 2001).

In addition to human settlements, the spatial patterns of occurrence of dengue cases were also greatly influenced by water bodies. In a 1 km buffer zone surrounding dengue cases, a share of water bodies of about 25–50 ha exercise the highest influence on the occurrences of dengue cases. Although delineation of the type of water bodies was not possible in this study, we infer that drains, small ponds and streams would be the water bodies of about 25–50 ha. Of particular relevance in this regard is the fact that Malaysia has an open concrete drainage system (Ghani et al., 2000; Zakaria et al., 2004). Our study is hence in line with other entomological studies in Malaysia that found that shallow clean stagnant water in drains served as a suitable habitat for the Aedes mosquitoes (Chen et al., 2005; Lee, 1991; Wan-Norafikah et al., 2012). Other water bodies including roadside ponds, pools and streams are also considered as natural reservoirs for Aedes mosquitoes (Uribe et al., 2008).

Mixed horticulture also contributes to the high frequency of dengue cases. The higher the coverage of mixed horticulture, particularly above ca. 200 ha, the higher the influence on dengue cases. Mixed horticulture is distributed throughout the suburban area. Orchards provide shade and temporary stagnant water on plants or in artificial containers that are suitable habitats for Aedes mosquitoes (Vanwambeke et al., 2007a). Vanwambeke et al. (2007a) also showed that the probability of finding the A. albopictus mosquito larvae in orchards, peri-urban settlements or villages was higher than in other land cover types. Furthermore, the gardens or ornamental plants surrounding the human settlements were also positively associated with dengue cases. Teng et al. (1999) found that containers used for planting and containers for watering plants formed part of the breeding habitats for Aedes mosquitoes.

More open land coverage in the 1 km vicinity shows a positive association with dengue cases, especially when open land covers an area of 150 to 200 ha. Natural gullies filled with rainfall are a good reservoir for Aedes mosquitoes (Delatte et al., 2013). Such natural reservoirs are available all year round and are often neglected as vector control largely targets human settlements (Delatte et al., 2013). Furthermore, potential breeding sites have been found in artificial containers discarded by workers in open land (Chang et al., 2011).

Aedes mosquitoes are free to proliferate without much dengue control intervention, especially in abandoned land cover types such as neglected grassland. Our study shows that a small patch of neglected grassland of ca. 25 ha would be a suitable reservoir for Aedes mosquitoes. Microhabitats that are shaded and vegetated are favorable habitats for Aedes mosquitoes to breed in (Tun-Lin et al., 1995; Vezzani et al., 2005). Another study also found that wet grassland was associated with suitable habitats for dengue vector reproduction (Li et al., 2013).

Our results suggest that land use factors are an important component to be considered in the strategic planning and implementation of vector control. This finding can be used as empirical evidence from the environmental perspective for Integrated Vector Management (IVM) (World Health Organization, 2012) at the state level. IVM has been found more effective than other vector control approaches in that it represents a rational decision-making process that allows the optimal use of resources for vector control, addressing all risk factors from parasite to vector, human to environment (Erlanger et al., 2008; World Health Organization, 2012). Although Malaysia is now adopting this strategy, there is still a lot to be done to ensure effective management with cost-effective programmes, intersectoral action, regulatory and operational standards, subsidiarity and sustainability (Horstick et al., 2010; Mnzava et al., 2006). For Selangor, we suggest more combined vector control intervention should be focused on human settlements, water bodies, mixed horticulture, open land and neglected grassland.

The risk map that we provide can offer a helpful resource for the local vector control programme for the state of Selangor (Fig. 2). The prediction map differentiates between areas of higher and lower risk. Moreover, the prediction model can be transferred to other locations with similar environmental settings as applied in other studies for predicting avian influenza risk (Van Boeckel et al., 2012) and fish species distribution (Pittman and Brown, 2011).

However, our study is constrained by four main aspects. Firstly, we use correlative models which cannot define a causal relationship. Although the relationships identified are not causal, they are suggestive for vector control intervention and are in line with earlier studies that use field knowledge and focus on the processes within the disease cycle (Bhatt et al., 2013; Martin et al., 2011). Secondly, special care needs to be taken in interpreting the relationship between land use factors, dengue cases and habitats of Aedes mosquitoes. The link between habitat and land cover characteristics is not always straightforward and is influenced by landscape structure (Vanwambeke et al., 2007b). Nevertheless, knowledge of mosquito ecology from other studies helps explain this relationship. Thirdly, we could not include human lifestyles and habits in the study due to a lack of data, e.g. on housing with air-conditioning, human mobility from house to workplace, recreational habits both outdoor or indoor, and self-awareness of personal protection – all of which could influence the occurrence of dengue cases and would be worth including in a more detailed study (Barmak et al., 2011). Fourthly, this study focused on the spatial and did not consider temporal effects associated with rapid climatic changes, changes in the composition of animal host population, abrupt land use/cover change, political and economic change (Lambin et al., 2010). However, our study indicated that there is a significant relationship between land use factors associated with higher frequencies of dengue cases and, moreover, we further quantified these relationships and depicted the risk of dengue on a state-wide level.

5. Conclusions

Our study has shown that BRT are a useful tool to model the nonlinear and complex association between dengue and land use factors and to derive a risk map. This approach can not only be applied to other vector-borne

diseases, but may also be transferred to other endemic regions. More detailed consideration of processes related to each land use type is a possible direction for future research. An equal focus on parasite, vector, human and environmental factors is needed for planning, implementing and evaluating the vector control interventions to reduce or eliminate dengue transmission.

Competing interests

Authors' contributions

CYL contributed to the design of the study, data collection and analysis and drafted the manuscript. PJL contributed to the design and analysis of the study and assisted with the manuscript preparation. TL conceived the design, coordination and implementation of the study and assisted with the manuscript preparation. All authors read and approved the final manuscript.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sste.2014.05.002.

References

Abdullah SA, Nakagoshi N. Changes in landscape spatial pattern in the highly developing state of Selangor, peninsular Malaysia. Landscape Urban Plan 2006;77(3):263–75.

Amala S, Anuradha V. A study on diversity of mosquitoes in the forest ecosystem of Sirumalai Hills, Dindigul, Tamilnadu. The ECOSCAN – An Int Q J Environ Sci 2011;2:219–22.

Arima Y, Matsui T. Epidemiologic update on the dengue situation in the Western Pacific Region, 2010. West Pac Surveill Response J 2011;2(2):4–8.

Armien B, Suaya JA, Quiroz E, Sah BK, Bayard V, Marchena L, et al. Clinical characteristics and national economic cost of the 2005 dengue epidemic in Panama. Am J Trop Med Hyg 2008;79(3):364–71.

Barmak DH, Dorso CO, Otero M, Solari HG. Dengue epidemics and human mobility. Phys Rev E 2011;84(1):011901.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature 2013;496(7446):504–7.

Chadwick D, Arch B, Wilder-Smith A, Paton N. Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. J Clin Virol 2006;35(2):147–53.

Chang MS, Christophel EM, Gopinath D, Abdur RM. Challenges and future perspective for dengue vector control in the Western Pacific Region. West Pac Surveill Response J 2011;2(2):9–16.

Chang MS, Hii J, Buttner P, Mansoor F. Changes in abundance and behaviour of vector mosquitoes induced by land use during the development of an oil palm plantation in Sarawak. Trans R Soc Trop Med Hyg 1997;91(4):382–6.

Chareonviriyaphap T, Akratanakul P, Huntamai S, Nettanomsak S, Prabaripai A. Allozyme patterns of *Aedes albopictus*, a vector of dengue in Thailand. J Med Entomol 2004;41(4):657–63.

Chen CD, Benjamin S, Saranum MM, Chiang YF, Lee HL, Nazni WA, Sofian-Azirun M. Dengue vector surveillance in urban residential and settlement areas in Selangor, Malaysia. Trop Biomed 2005;22(1):39–43.

Chen CD, Teoh PS, Othman W-N, Sofian-Azirun M, Lee HL. The effectiveness of a floor gully with grating to prevent the oviposition of *Aedes* (Stegomyia) *aegypti* (Linneaus): a laboratory study. Jurnal Sains Kesihatan Malaysia 2009;7(1):73–80.

Cheong YL, Burkart K, Leitão PJ, Lakes T. Assessing weather effects on dengue disease in Malaysia. Int J Environ Res Public Health 2013;10(12):6319–34.

Coutts SR, Yokomizo H. Meta-models as a straightforward approach to the sensitivity analysis of complex models. Popul Ecol 2014;56(1):7–19.

Cui Y. A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. Appl Geogr 2013;41:87–95.

Dambach P, Sie A, Lacaux J, Vignolles C, Machault V, Sauerborn R. Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district Burkina Faso. Glob Health Action 2009:2–5.

De'ath G. Boosted trees for ecological modeling and prediction. Ecology 2007;88(1):243–51.

Delatte H, Toty C, Boyer S, Bouetard A, Bastien F, Fontenille D. Evidence of Habitat Structuring *Aedes albopictus* populations in Réunion Island. PLoS Negl Trop Dis 2013;7(3):e2111.

Department of Statistics Malaysia. Compendium of Environment Statistics Malaysia. Department of Statistics Malaysia, Putrajaya; 2011.

Dom NC, Latif ZA, Ahmad AH, Ismail R, Pradhan B. Manifestation of GIS tools for spatial pattern distribution analysis of dengue fever epidemic in the city of Subang Jaya, Malaysia. Environ Asia 2012;5(2):82–92.

Elith J, Leathwick J, Hastie T. A working guide to boosted regression trees. J Anim Ecol 2008;77(4):802–13.

Erlanger TE, Keiser J, Utzinger J. Effect of dengue vector control interventions on entomological parameters in developing countries: a systematic review and meta-analysis. Med Vet Entomol 2008;22(3):203–21.

Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat 2000;28(2):337–74.

Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29(5):1189–232.

Friedman JH. Stochastic gradient boosting. Comput Stat Data An 2002;38(4):367–78.

Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. Stat Med 2003;22(9):1365–81.

Ghani AA, Zakaria NA, Kassim M, Nasir BA. Sediment size characteristics of urban drains in Malaysian cities. Urban Water 2000;2(4):335–41.

Gleiser RM, Zalazar LP. Distribution of mosquitoes in relation to urban landscape characteristics. Bull Entomol Res 2010;100(2):153–8.

Gu W, Wang X, McGregor S. Optimization of preventive health care facility locations. Int J Health Geographics 2010;9(1):17.

Gubler DJ. Dengue/dengue haemorrhagic fever: history and current status. Novartis Found Symp 2006;277:3–16.

Gubler DJ. Dengue and dengue hemorrhagic fever. Clin Microbiol Rev 1998;11(3):480–96.

Gubler DJ, Clark GG. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. Emerg Infect Dis 1995;1(2):55–7.

Halstead SB. Dengue virus–mosquito interactions. Annu Rev Entomol 2008;53:273–91.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. 2nd ed. New York: Springer-Verlag; 2009. pp. 337–384.

Hay SI, Myers MF, Burke DS, Vaughn DW, Endy T, Ananda N, et al. Etiology of interepidemic periods of mosquito-borne disease. Proc Natl Acad Sci USA 2000;97(16):9335–9.

Hayden MH, Uejio CK, Walker K, Ramberg F, Moreno R, Rosales C, et al. Microclimate and human factors in the divergent ecology of *Aedes*

aegypti along the Arizona, US/Sonora, MX Border. Ecohealth 2010;7(1):64–77.

Hijmans RJ, Phillips S, Leathwick J, Elith J, dismo: species distribution modeling. R package version 0.8-17. Available from <http://cran.r-project.org/web/packages/dismo/index.html> (accessed August 2013).

Horstick O, Runge-Ranzinger S, Nathan MB, Kroeger A. Dengue vector-control services: how do they work? A systematic literature review and country case studies. Trans R Soc Trop Med Hyg 2010;104(6):379–86.

Idris FH, Usman A, Surendran SN, Ramasamy R. Detection of *Aedes albopictus* pre-imaginal stages in brackish water habitats in Brunei Darussalam. J Vector Ecol 2013;38(1):197–9.

Krishnan N, Purswani M, Hagmann S. Severe dengue virus infection in pediatric travelers visiting friends and relatives after travel to the Caribbean. Am J Trop Med Hyg 2012;86(3):474–6.

Lam SK. Two decades of dengue in Malaysia. Trop Med 1993;35(4):195–200.

Lambin E, Tran A, Vanwambeke S, Linard C, Soti V. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int J Health Geographics 2010;9(1):54.

Lee HL. A nationwide resurvey of the factors affecting the breeding of *Aedes aegypti* (L.) and *Aedes albopictus* (Skuse) (Diptera: Culicidae) in urban town of peninsular Malaysia – 1988–1989. Trop Biomed 1991;8:185–9.

Li S, Tao H, Xu Y. Abiotic determinants to the spatial dynamics of dengue fever in Guangzhou. Asia-Pacific J Public Health/Asia-Pacific Acad Consort Public Health 2013;25(3):239–47.

Lourenco-de-Oliveira R, Castro MG, Braks MAH, Lounibos LP. The invasion of urban forest by dengue vectors in Rio de Janeiro. J Vector Ecol 2004;29(1):94–100.

Maciel-de-Freitas R, Neto RB, Goncalves JM, Codeco CT, Lourenco-de-Oliveira R. Movement of dengue vectors between the human modified environment and an urban forest in Rio de Janeiro. J Med Entomol 2006;43(6):1112–20.

Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, et al. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. Plos Pathog 2011;7(3):e1001308.

Ministry of Health Malaysia. Health Facts 2010. Ministry of Health Malaysia, Putrajaya; 2011.

Mnzava A, Curtis C, Manga L, Touré Y. Application of genetically modified mosquitoes in national vector control programmes: thoughts on integrated control. In: Knols BGJ, Louis C, editors. Bridging laboratory and field research for genetic control of disease vectors. The Netherlands: Springer; 2006. p. 211–4.

Mocellin MG, Simoes TC, do Nascimento TFS, Teixeira MLF, Lounibos LP, de Oliveira RL. Bromeliad-inhabiting mosquitoes in an urban botanical garden of dengue endemic Rio de Janeiro. Are bromeliads productive habitats for the invasive vectors Aedes aegypti and Aedes albopictus? Mem I Oswaldo Cruz 2009;104(8):1171–6.

Muhammad Azami NA, Salleh S, Neoh H.m., Syed Zakaria S, Jamal R. Dengue epidemic in Malaysia: not a predominantly urban disease anymore. BMC Res Notes 2011;4(1):216.

Müller D, Leitão PJ, Sikor T. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agr Syst 2013;117:66–77.

Nazri C, Hashim A, Rodziah I. Distribution pattern of a dengue fever outbreak using GIS. J Environ Health Res 2009;9(2):89–96.

Neteler M, Roiz D, Rocchini D, Castellani C, Rizzoli A. Terra and Aqua satellites track tiger mosquito invasion: modelling the potential distribution of *Aedes albopictus* in north-eastern Italy. Int J Health Geographics 2011;10(1):49.

Nyamah MA, Sulaiman S, Omar B. Categorization of potential breeding sites of dengue vectors in Johor, Malaysia. Trop Biomed 2010;27(1):33–40.

Oppel S, Meirinho A, Ramírez I, Gardner B, O'Connell AF, Miller PI, et al. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biol Conserv 2012;156:94–104.

Paily KP, Chandhiran K, Vanamail P, Kumar NP, Jambulingam P. Efficacy of a mermithid nematode *Romanomermis iyengari* (Welch) (Nematoda: Mermithidae) in controlling tree hole-breeding mosquito *Aedes albopictus* (Skuse) (Diptera: Culicidae) in a rubber plantation area of Kerala, India. Parasitol Res 2013;112(3):1299–304.

Pittman SJ, Brown KA. Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS ONE 2011;6(5):e20583.

Ponlawat A, Harrington LC. Blood feeding patterns of *Aedes aegypti* and *Aedes albopictus* in Thailand. J Med Entomol 2005;42(5):844–9.

R Development Core Team. R: a language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria; 2013.

Ramasamy R, Surendran SN, Jude PJ, Dharshini S, Vinobaba M. Larval development of *Aedes aegypti* and *Aedes albopictus* in peri-urban brackish water and its implications for transmission of arboviral diseases. PLoS Negl Trop Dis 2011;5(11):e1369.

Richards S, Ponnusamy L, Unnasch T, Hassan H, Apperson C. Host-feeding patterns of *Aedes albopictus* (Diptera: Culicidae) in relation to availability of human and domestic animals in suburban landscapes of central North Carolina. J Med Entomol 2006;43(3):543.

Ridgeway G., Gbm: generalized boosted regression models. R package version 2.1. Available from <http://cran.r-project.org/web/packages/gbm/> (accessed June 2013).

Rohani A, Abdullah A, Ong Y, Saadiyah I, Zamree I, Lee H. Survey of mosquito larvae distribution in container habitats collected from urban and rural areas in major towns of Malaysia. Trop Biomed 2001;18(1):41–9.

Roongpiboonsopit D, Karimi HA. Comparative evaluation and analysis of online geocoding services. Int J Geogr Inf Sci 2010;24(7):1081–100.

Russell BM, Muir LE, Weinstein P, Kay BH. Surveillance of the mosquito *Aedes aegypti* and its biocontrol with the copepod *Mesocyclops aspericornis* in Australian wells and gold mines. Med Vet Entomol 1996;10(2):155–60.

Sarfraz MS, Tripathi N, Tipdecho T, Thongbu T, Kerdthong P, Souris M. Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. BMC Public Health 2012;12(1):853.

Scott TW, Morrison AC. Longitudinal field studies will guide a paradigm shift in dengue prevention. In: Atkinsons PW, editor. Vector biology, ecology and control. The Netherlands: Springer; 2010. p. 139–61.

Shafie A. Evaluation of the spatial risk factors for high incidence of dengue fever and dengue hemorrhagic fever using GIS application. Sains Malaysiana 2011;40(8):937–43.

Skae FMT. Dengue Fever in Penang. Br. Med. J. 1902;2(2185):1581–2.

Stevens KB, Pfeiffer DU. Spatial modelling of disease using data- and knowledge-driven approaches. Spat Spatio-Temporal Epidemiol 2011;2(3):125–33.

Sumodan PK. Potential of rubber plantations as breeding source for *Aedes albopictus* in Kerala, India. Dengue Bull 2003;27:197–8.

Teng AK, Singh S. Epidemiology and new initiatives in the prevention and control of dengue in Malaysia. Dengue Bull 2001;25:7–14.

Teng HJ, Wu YL, Lin TH. Mosquito fauna in water-holding containers with emphasis on dengue vectors (Diptera: Culicidae) in Chungho, Taipei County. Taiwan. J. Med. Entomol. 1999;36(4):468–72.

Thavara U, Tawatsin A, Chansang C, Kong-ngamsuk W, Paosriwong S, Boon-Long J, et al. Larval occurrence, oviposition behavior and biting activity of potential mosquito vectors of dengue on Samui Island, Thailand. J Vector Ecol 2001;26(2):172–80.

Tun-Lin W, Kay BH, Barnes A. The premise condition index: a tool for streamlining surveys of *Aedes aegypti*. Am J Trop Med Hyg 1995;53(6):591–4.

Uribe LAA, Pineda SV, Martinez OG, Flores JL, Davila MF, Harris MK. Culicid distribution in Coahuila, Mexico. Southwestern Entomol 2008;33(3):219–22.

Van Boeckel TP, Thanapongtharm W, Robinson T, Biradar CM, Xiao X, Gilbert M. Improving risk models for avian influenza: the role of intensive poultry farming and flooded land during the 2004 Thailand epidemic. PLoS ONE 2012;7(11):e49528.

Vanwambeke SO, Lambin EF, Eichhorn MP, Flasse SP, Harbach RE, Oskam L, et al. Impact of land-use change on dengue and malaria in northern Thailand. Ecohealth 2007a;4(1):37–51.

Vanwambeke SO, Somboon P, Harbach RE, Isenstadt M, Lambin EF, Walton C, et al. Landscape and land cover factors influence the presence of Aedes and Anopheles larvae. J Med Entomol 2007b;44(1):133–44.

Vezzani D, Rubio A, Velazquez SM, Schweigmann N, Wiegand T. Detailed assessment of microhabitat suitability for *Aedes aegypti* (Diptera: Culicidae) in Buenos Aires, Argentina. Acta Trop 2005;95(2):123–31.

Wan-Norafikah O, Nazni WA, Noramiza S, Shafa'ar-Ko'Ohar S, Heah SK, Nor-Azlina AH, et al. Distribution of Aedes mosquitoes in three selected localities in Malaysia. Sains Malaysiana 2012;41(10):1309–13.

World Health Organization. Dengue and dengue haemorrhagic fever. Health in Asia and the Pacific; 2008. p. 244–255.

World Health Organization. Handbook for integrated vector management. WHO library cataloguing-in-publication data; 2012a.

World Health Organization. DengueNet database and geographic information system. In: World Health Organization; 2012b.

Zakaria NA, Ghani AA, Abdullah R, Sidek LM, Kassim H, Ainan A. MSMA A new Urban Stormwater Management Manual for Malaysia. In: Altinakar MS, Wang SSY, Holz KP, Kawahara M, editors. Proceedings of 6th International Conference on Hydro-Science and Engineering (ICHE), May 31–June 2, 2004. Australia: Brisbane; 2004. p. 1–10.