Mapping urban and peri-urban breeding habitats of *Aedes* mosquitoes using a fuzzy analytical hierarchical process based on climatic and physical parameters

Muhammad Shahzad Sarfraz[1], Nitin K. Tripathi[2], Fazlay S. Faruque[3], Usama Ijaz Bajwa[4], Asanobu Kitamoto[5], Marc Souris[2,6]

[1]Department of Computer Science, National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Pakistan; [2]Remote Sensing and GIS Field of Study, School of Engineering and Technology, Asian Institute of Technology, Pathum Thani, Thailand; [3]GIS and Remote Sensing Program, University of Mississippi Medical Center, Jackson, USA; [4]Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad, Pakistan; [5]Digital Content and Media Sciences Research Division, National Institute of Informatics (NII), Tokyo, Japan; [6]Institut de Recherche pour le Dévelopement (IRD), Marseille, France

Abstract. The spread of dengue fever depends mainly on the availability of favourable breeding sites for its mosquito vectors around human dwellings. To investigate if the various factors influencing breeding habitats can be mapped from space, dengue indices, such as the container index, the house index and the Breteau index, were calculated from Ministry of Public health data collected three times annually in Phitsanulok, Thailand between 2009 and 2011. The most influential factors were found to be temperature, humidity, rainfall, population density, elevation and land cover. Models were worked out using parameters mostly derived from freely available satellite images and fuzzy logic software with parameter synchronisation and a predication algorithm based on data mining and the Decision Tree method. The models developed were found to be sufficiently flexible to accommodate additional parameters and sampling data that might improve prediction of favourable breeding hotspots. The algorithm applied can not only be used for the prediction of near real-time scenarios with respect to dengue, but can also be applied for monitoring other diseases influenced by environmental and climatic factors. The multi-criteria model presented is a cost-effective way of identifying outbreak hotspots and early warning systems lend themselves for development based on this strategy. The proposed approach demonstrates the successful utilisation of remotely sensed images to map mosquito breeding habitats.

Keywords: dengue fever, fuzzy analytic hierarchy process, larval density, data mining, climatic factors, health, geographical information system, Thailand.

Introduction

Dengue is a mosquito-borne, viral infection that can cause dengue fever (DF), occasionally leading to potentially lethal complications, such as dengue haemorrhagic fever (Simmons et al., 2012). This infection is found in tropical and sub-tropical climates worldwide, mostly in urban and semi-urban areas (Sarfraz et al., 2013) and the global incidence has grown dramatically in recent decades, so that about half of the world's population is now at risk (WHO, 2012; Bhatt et al., 2013). The dengue virus is a particular threat to humans at times of increased abundance of its vectors *Aedes aegypti* and *Ae. albopictus*. The transmission risk depends on climatic factors, which govern the expansion of mosquito habitats (Nakhapakorn and Tripathi, 2005; Hassan et al., 2012; Morin et al., 2013; Banu et al., 2014). There is no specific treatment for DF, but early detection and access to proper medical care can lower fatality rates to levels below 1% (WHO, 2012). Without chemotherapy and vaccines, dengue prevention depends solely on effective vector control. Information and communication technologies (ICT) have been recognised for enabling solutions to health and healthcare that can help with regard to these challenges.

Climatic parameters, especially rainfall, temperature and humidity along with land cover types have been reported to influence the densities of the vector (Craig et al., 1999; Ceccato et al., 2005; Machault et al.,

Corresponding author:
Muhammad Shahzad Sarfraz
Department of Computer Science
National University of Computer and Emerging Sciences
Chiniot-Faisalabad Campus, Chiniot, Punjab, Pakistan
Tel. +92 111 128-128; Fax +92 41 260-7272
E-mail: shahzad.sarfraz@nu.edu.pk

2011; Dambach et al., 2012) in localities with many open water containers that function as effective vector breeding sites when filled during rainfall (Arunachalam et al., 2010; Sarfraz et al., 2012). Thus, rainfall patterns have a clear spatio-temporal effect on breeding sites and can also improve larval survival by creating a favourable humid microclimate for the adult mosquito (Machault et al., 2011). *Ae. aegypti* is an anthropophagic mosquito that thrives in and around human dwellings, where a variety of sources of standing water provides ideal habitat for its larvae. *Ae. albopictus* also prefers rain-filled containers, in partcular tires and naturally formed cavities, such as tree holes (Vanwambeke et al., 2011). For this reason, many researchers have examined the use of precipitation data as a means for forecasting DF (Focks et al., 1995; Buczak et al., 2012).

Low temperatures limit the life-cycle of the dengue vectors (Beebe et al., 2009), while warmth enhances mosquito population densities and disease incidence (Reiter, 1996; Hsieh and Chen, 2009). Lambrechts et al. (2011) investigated the effect of realistic temperature profiles with respect to day and night on the capacity of mosquitoes to transmit the dengue virus. It has been found that the range of temperature fluctuations is inversely related to vector susceptibility and survival with respect to the dengue virus, so that both infection and survival rates decrease when the night-day temperature range increases. The trend of increasing temperatures in various regions of the world, predicted by global climate change models, points towards a possible pandemic DF threat (Banu et al., 2011).

Specific humidity, defined as the amount of vapour pressure in the air, is high only when both rainfall and temperatures are high. These are the very conditions that are conducive for the breeding and survival of vector populations and rapid virus replication (Focks et al., 1993). Thus, humidity has a direct impact on the occurrence and persistence of breeding sites and on larval survival rates (Aziz et al., 2011).

The land cover allows for the colonisation of new breeding habitats leading to a modified composition of the vector populations (Patz and Norris, 2004). Cultivation/agriculture is the largest driver for the changing land use across the globe. Together, pastures and croplands now cover more than 40% of the world's land surface and thus rival forests as the largest terrestrial biome (Ramankutty and Foley, 1999; Asner et al., 2004). Khormi and Kumar (2012) report that land use/land cover (LULC) patterns as well as socio-cultural and economic prac-

tices all have a strong relation to the incidence of vector-borne diseases including DF. Land cover and its vegetation characteristics can thus amplify the distribution of the dengue virus by modification of its vector populations. Changing land use is a major constituent of the global environmental alteration that potentially affects human health in relation to mosquito-borne diseases. Considering a dengue vector's flight range during its life span (generally no further than 1 km), buffering operations have been carried out to find the relationship between mosquito density-related indices and land cover classes (Sarfraz et al., 2013).

Increased human population densities result in higher levels of disease morbidity and mortality (Quintero et al., 2014). Indeed, growing anthropogenic influence on the natural environment, such as changes in land use and deterioration of ecosystems, are accelerating the problem. Increased exposure to infectious diseases has been identified as one of five major emerging public health threats developing from the above-mentioned changes of the natural environment (Myers and Patz, 2009). Urbanisation, in particular, is a major driver for the emergence of epidemic DF (Gubler, 2011). Technologies developed for the collection of Earth-observational data can detect such changes and help identifying potential risk zones for spread of the disease.

Many researchers report a relationship between the number of DF cases and climatic factors (Nakhapakorn and Tripathi et al., 2005; Jeefoo et al., 2010; Goto et al., 2013). Drawbacks associated with the use of case data include the asymptomatic infections, which cannot be used to support a real-time indication of future disease scenarios. Moreover, the results of studies comparing remotely sensed data with ground data interpolated from meteorological ground stations often do not correspond well with the annual mean climatic data due to limitations incurred by large distances between the stations as there is usually a fixed number of them per area (Machault et al., 2011). Interpolation techniques are used to produce approximate parameter values for the areas situated between two stations, which means that predictions directly based on larval densities in these locations must be more reliable (Sarfraz et al., 2012). However, the laborious collection of climatic parameters from ground stations and integration with health data are time consuming and cause intervention delays, while remote sensing approaches can provide useful data for any given region depending on the available resolution specifics. During the past few decades, geospatial and

remotely sensed data have been increasingly utilised in this respect (Beck et al., 2000; Machault et al., 2011). Adopting novel and freely available online data are most useful and can aid in achieving timely awareness of the status of various vector-borne diseases (Hay et al., 2013).

Preventive measures based on near-real time prediction of climatic and environmental fluctuation can save human lives. The objectives of this study were to create a simulation model, capable of predicting possible breeding habitats of the *Aedes* vectors by considering entomological information (dengue indices) together with climatic and environmental data.

Materials and methods

Study Area

Phetchabun Province, located in the lower, northern region of Thailand, (16.27 N, 101.09 E) about 350 km north of Bangkok was selected as the study area due to known variations in demography and climatic factors. Recent topographic data obtained from the Land Development Office in Thailand show Phetchabun as comprising 11 districts subdivided into 1,261 villages, the smallest public health unit in Thailand. The current study was carried out in 131 of these villages covering all 11 districts (Fig. 1).

The central part of Phetchabun Province is situated on the Pa Sak River basin with mountain ranges running along both the western and the eastern edges with an altitude varying between 43 and 1,766 m above the mean sea level (MSL). The province is agriculturally productive, today also known for tourism. According to National Statistical Office data for 2010, there are about 1 million people in Phetchabun with the majority between 15 to 59 years of age and a population density of approximately 79 per km². The climate is mainly pleasant varying between a maximum mean of 34.8 °C and minimum mean of 23.3 °C. The humidity varies between 94% and 39%, while the daily, maximum rainfall is 78.3 mm.
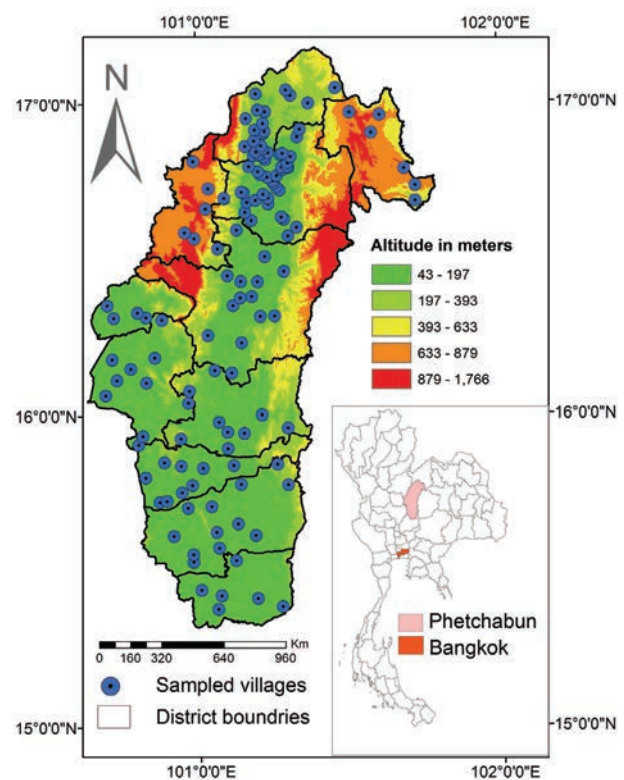

Fig. 1. Phetchabun Province with district boundaries and the location of samples villages. The blue buffers indicate the 131 villages, which were sampled three times (in March, May and July) during the period of 2009-2011. The inset shows the location od the province within Thailand.

Model parameters

Before developing the predictive algorithm for near real-time dengue habitat modelling, we felt that it necessary to provide a base by producing a model on the basis of the ground truth data. The variables considered included dengue larval density and a set of remotely sensed data. Due to limited resolution with regard to rainfall and humidity data, interpolation techniques were used. Among all the interpolations methods tested, the inverse distance weighted (IDW) (Sarfraz et al., 2012) approach resulted in the least

Table 1. Details of satellite resources used.

Source/information	Temperature	Rainfall	Humidity	Elevation
Satellite platform	AQUA	TRMM	AQUA	SRTM
Instrument	MODIS	Precipitation radar	AIRS	Synthetic aperture radar
Image name	MYD11C3	3B43	RetQuant	SRTM_57_09
Spatial resolution	0.05 degree	0.25 degree	13.5 km	90 m
Availability	Day/Night	3-hourly	5 days	11-day mission
Version	Version 5	Version 6	Version 5	Version 1

root-mean-square error and was therefore adopted. Details of remotely sensed data used are summarised in Table 1.

Dengue entomological data

The entomological data were used as a primary variable to develop this model. It consisted of the larval density of *Ae. aegypti* and *Ae. albopictus* inside and around residential homes. The specimens were collected according to the World Health Organization (WHO) standards through the Office of Disease Prevention and Control, Ministry of Public Health, Thailand. Sampling was done three times a year (March, May and July) during the peak time of dengue breeding in the study region. Data samples, collected from different districts within the province considering lifestyles of the local population and the diversity of the surrounding environment, were used to calculate the house index (HI); the container index (CI); and the Breteau index (BI). In the study villages, sampling houses were randomly selected. If locked, they were tested during the next visit and adjacent houses were sampled instead. This was felt acceptable as neighbouring houses would have an almost identical structure with the same surroundings

$$House\ Index = \frac{Number\ of\ infested\ houses}{Number\ of\ inspected\ houses} \times 100$$

<div align="right">equation (1)</div>

$$Container\ Index = \frac{Number\ of\ infested\ containers}{Number\ of\ inspected\ containers} \times 100$$

<div align="right">equation (2)</div>

$$Breteau\ Index = \frac{Number\ of\ infested\ containers}{Number\ of\ inspected\ containers} \times 100$$

<div align="right">equation (3)</div>

Temperature

The monthly average temperature, the maximum land surface day temperature (LST_{day}) and the minimum night temperature (LST_{night}) were extracted from the moderate resolution imaging spectroradiometer (MODIS) on-board the Aqua satellite in the form of a hierarchical data format (HDF) file (LP DAAC, 2012) re-projected for the study area using the conversion utility (MCKT) toolkit (http://www.nsidc.org/data/modis/tools.html) and an area mask. The temperatures were expressed as degrees Kelvin by applying a factor for linear re-scaling (0.02) of the sensor data,

which are normally referred to as digital numbers (DN) (http://www.icess.ucsb.edu/modis/LstUsrGuide/usrguide_mod11.html#sds) and transformed into centigrades by subtracting the 273.15 value. These steps are summarised in equation 4. The average temperatures in Fig. 2 were extracted using three-year averages.

$$LST_{centigrade} = LST_{data}\ \text{x scale factor} - 273.15$$

<div align="right">equation (4)</div>

After acquiring the re-scaled temperatures according to Celsius returned by the equation 4, the next step was to relate the effect of a certain temperature on mosquito breeding. This was achieved by equation 5 below, which maps a given temperature in the continuous range of 0 (i.e. no influence on the dengue larval density) to 1 (i.e. high impact). Different ranges of temperature values and calculation of their corresponding impact were determined by relating the dengue larval density at a survey point to its measured temperature using a fuzzy logic approach as discussed in the section on dengue habitat modelling (see below).

$$Temperature\ (x) = \begin{cases} 0 & x < 20\ \text{or}\ x > 40 \\ \dfrac{x - 20}{10} & x \geq 21\ \text{or}\ x \leq 29 \\ 1 & x \geq 30\ \text{or}\ x \leq 35 \\ \dfrac{40 - x}{5} & x > 36\ \text{or}\ x < 40 \end{cases}$$

<div align="right">equation (5)</div>

Population density

The gridded populations of the world (GPW), version 3 raster data (CIESIN, 2012) provided by the Socio Economic Data and Applications Center (SEDAC) (http://www.ciesin.org/) was used for estimating the local population density (Table 1). The GPW data contains a near-global population dataset based on night-time light imagery found in various human settlements (Fig. 3). GPW data is the best possible solution for regions with unreliable population data. The population parameter is modified according to maximum pixel value of the raster image. A higher accuracy may be possible by using population density maps from local organisations. However, as no frequent changes in the population density was expected this parameter was kept fixed throughout the research study.
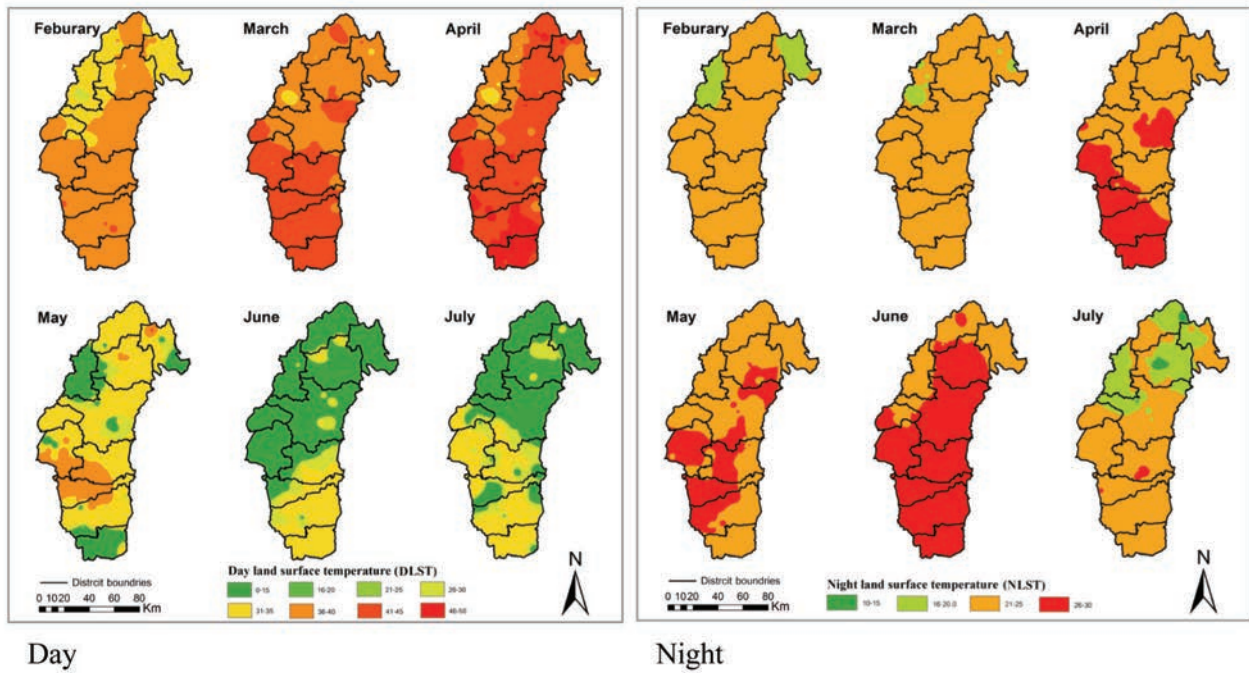
Fig. 2. Monthly average temperatures for 2010 in Phetchabun Province (MODIS data, rescaled and expressed in centigrades).
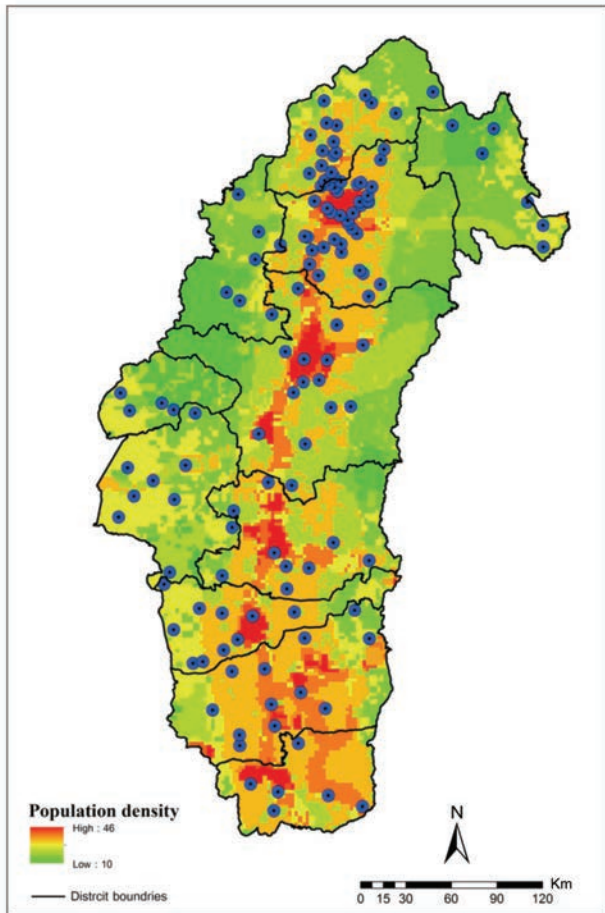

Fig. 3. Population density of Phetchabun Province. Data from the Gridded Population of the World 2010.

Rainfall

Accurate rainfall measurements are generally problematic, since it is impossible to place rain gauges throughout large areas. The Tropical Rainfall Measuring Mission (TRMM) (http://www.trmm.gsfc.nasa.gov/) attempts to provide accurate measurements of global precipitation (Kummerow et al., 1998). The TRMM satellite was launched in November 1997 and has collected data since the following month with a spatial resolution of 0.25 to 5.0 degrees. The radar rainfall product (3B42) provided indirect rain gauge data along with the same data based on a combination of datasets for the current study (Table 1). Three-hourly, rainfall measurements of variable quality were summed over one month, creating a monthly multi-satellite (MS) product. By summing the gauge and the MS product, a post-real-time, monthly satellite-gauge combination (SG) was developed (Fig. 4). Reasoning along the same lines as for done the temperature, Eq. 6 maps rainfall values in a 0 to 1 range were produced:

$$Rainfall\ (x) = \begin{cases} 0 & x < 10 \text{ or } x \geq 180 \\ \dfrac{x-1}{40} & x \geq 11 \text{ or } x \leq 40 \\ 1 & x \geq 41 \text{ or } x \leq 80 \\ \dfrac{180-x}{100} & x > 81 \text{ or } x < 180 \end{cases}$$
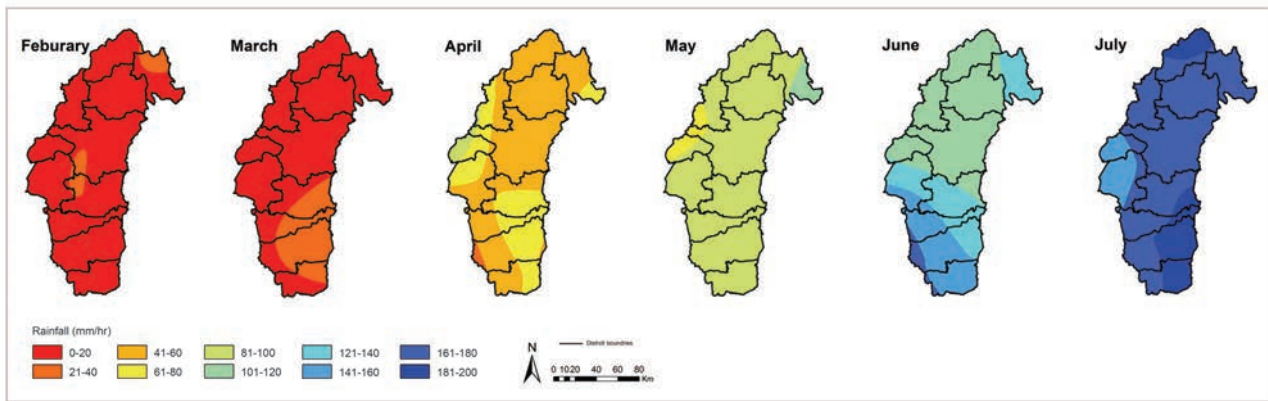
equation (6)

Fig. 4. Monthly average of rainfall in Phetchabun Province (TRMM 2010 data).

Relative humidity

Day and night humidity data were acquired through Atmospheric Infrared Sounder (AIRS) (https:// www.airs.jpl.nasa.gov/), a satellite-based, continuously-operating, cross-track scanning sounder with a telescope feeding an "echelle" spectrometer. The spatial footprint of the infrared channels has a 1.1 degree diameter, which corresponds to about 15 x 15 km in the nadir. (Table 1). Humidity increases at night compared to day time levels, so both day and night humidity values (Fig. 5) were used for the analysis. As with temperature and rainfall, the same range (0 to 1) was applied for calculating this parameter:

$$Humidity\ (x) = \begin{cases} 0 & x < 50 \text{ or } x > 90 \\ \dfrac{x - 49}{20} & x \geq 50 \text{ or } x \leq 68 \\ 1 & x > 68 \text{ or } x \leq 79 \\ \dfrac{90 - x}{11} & x \geq 80 \text{ or } x \leq 90 \end{cases}$$
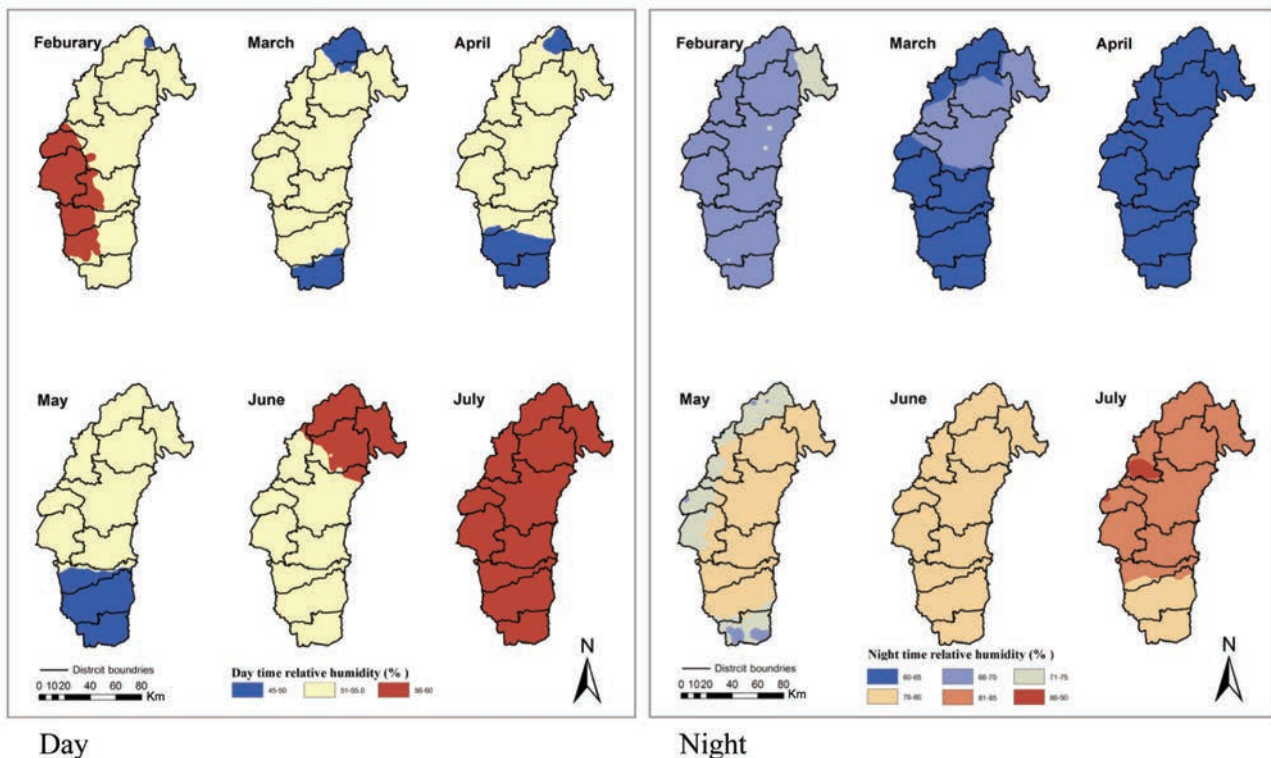
equation (7)

Fig. 5. Monthly average humidity in Phetchabun Province (AIRS 2010 data).

Digital elevation data

More DF cases occur at lower altitudes than in high-land areas (Aziz et al., 2011). To evaluate elevation affects on mosquito breeding habitats, values derived from the Digital Elevation Model (DEM) (http://www2.jpl.nasa.gov/srtm/) with 90-m resolution from the Shuttle Radar Topography Mission (SRTM) were used (Table 1). Three arc second (90 m) data of elevation are freely available globally. Similar to the other parameters, our fuzzy modelling approach, relating the given elevation data to the dengue larval densities found, divided the elevation data into three different linear ranges and provided mappings for each range (again the 0-1 approach was used) intervals as shown in equation 8:

$$Elevation\ (x) = \begin{cases} 0 & x < 44 \text{ or } x \geq 1501 \\ \dfrac{x-42}{360} & x \geq 44 \text{ or } x \leq 400 \\ 1 & x \geq 401 \text{ or } x \leq 1100 \\ \dfrac{1500-x}{400} & x > 1101 \text{ or } x < 1500 \end{cases}$$

equation (8)

Land use/land cover

Landscape features, such as land use/land cover (LULC) affect the availability of suitable mosquito breeding habitats and hence provide an indication of the abundance of vectors. LULC also determines the location of people in relation to the vector habitats and therefore affect human exposure. The MODIS land cover product (MCD12) is freely available (https://www.lpdaac.usgs. gov/) and therefore lends itself for this type of study. This product contains multiple classification schemes, which include 11 natural vegetation classes, three non-vegetation ones and three developed land class mosaics. To achieve high accuracy with respect to the relationship between LULC and the mosquito breeding potential, LULC data consisting of more than 150 classes were obtained from the Land Development Office of Thailand. Detailed information on LULC methods are available in a previous study conducted for the Phitsanulok health division (Sarfraz et al., 2012). Study results obtained by this previous study were merged with the new data from the environmental parameters mentioned above to see if the combined effect on mosquito breeding habitats would gain additional, useful information.

Preprocessing

Because different environmental, climatic and demographic data were collected at diverse spatio-temporal scales, they had to be converted into a uniform scale. The temporal scale was one month and the village was considered the unit for spatial distribution. A buffering operation of 200, 500 and 1,000 m was performed using geographical information systems (GIS) software for the 131 villages as these buffers were considered to cover the range from minimum to maximum mosquito flight range during its life span. The percentage of each buffer area was calculated for further statistical analysis to see the affected land cover with respect to the Breteau Index. All images were re-projected onto the map projection UTM zone 47 N and datum WGS 84.

Habitat modelling and machine-learning

To examine factors affecting the dengue larval habitat potential, a machine-learning workbench using the Konstanz Information Miner (KNIME) was used for the current study (Berthold et al., 2008). KNIME allows the rapid exploration of a given dataset using a variety of machine-learning schemes and data mining. It is an open source, graphical user interface programme, widely used for pharmaceutical research (Holzinger et al., 2013). We utilised KNIME here for habitat modelling to develop a methodology for climatic factors contributing to favourable breeding sites. Through experience with this tool, we learnt that successful application of machine-learning involves much more than merely executing a learning algorithm on some data. The process model used for our machine-learning applications is presented below (Fig.6).
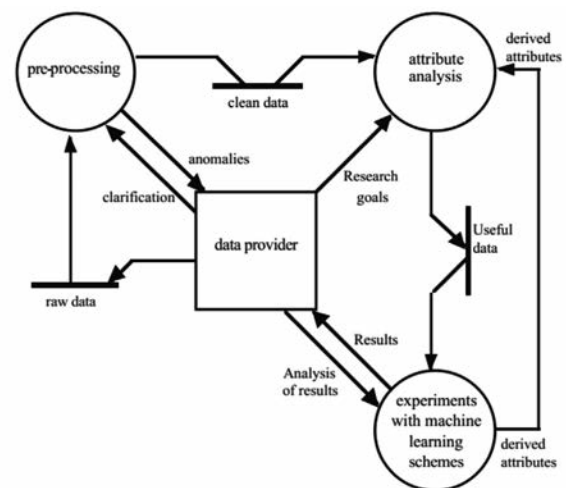

Fig. 6. Flow diagram for the extraction of suitable ranges provided by the data mining software.
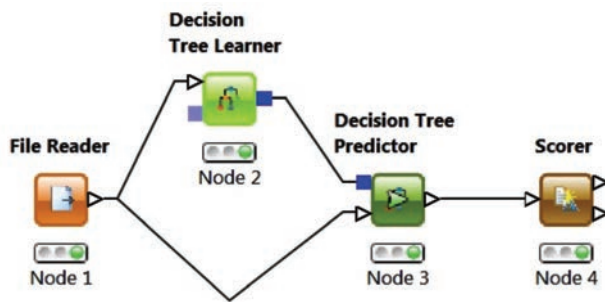
Fig. 7. Decision score flow process of Decision Tree learning method.

Different mining algorithms (e.g. Bayesian, Cluster, Neural Network and Decision Tree) were applied using the KNIME. The Decision Tree algorithm provided a minimum re-substitution error rate and this algorithm resulted in high accuracy for all parameters investigated here. Using the "CONFIGURE" menu, the dengue parameters.xls data file of dependent and independent parameters (larval density, temperature, population density, rainfall, relative humidity and elevation) were imported with the "FILE READER" component, while the "LEARNER" component was used for the induction of the Decision Tree on the whole dataset to find the suitable and the not suitable ranges. The Decision Tree predictor with classifier was used for the learning set. The confusion (error) matrix, a table layout allowing visualisation of the performance of algorithms in general, was computed using the "SCORER" component (Fig. 7).

Fuzzy analytic hierarchy process

A disease involves several levels of uncertainty and imprecision. When the knowledge discovered is in the form of association rules, the methodology is called association rule mining (ARM), which describes a relationship among different attributes (Agrawal et al., 1993). For the development of a decision support system for the mosquito habitat mapping on the basis of the above datasets, fuzzy logic was applied. Decision support systems can be helpful in introducing automation and intelligence for real-time modelling (Buczak et al., 2012). Among the various habitat models and automated outbreak prediction systems developed, the vector-borne disease prediction model is considerably complex since it requires evaluation and predication under uncertainty based on social, climatic, environmental and behavioural factors (Sarfraz et al., 2013). The analytic hierarchy process (AHP) is a sound way to address various problems as this method is qualitative and easy to implement from the validation and data requirement point of view. However, the AHP is not useful in solving prediction problems based on uncertainty such as dengue outbreaks. In order to exclude this constraint, we developed the fuzzy-based habitat suitability approach, a technique that eliminates the requirement of extra aggregation as well as ranking procedures.

Using the AHP approach, we accepted all locations falling within the given criteria; if not (even if very close to the set threshold), they were excluded from analysis. Spatial features often lack clearly defined boundaries and concepts like "favourable", "moderately favourable", or "unfavourable" can be better expressed as degrees of membership to a fuzzy dataset than with a binary yes/no classification. Humans often use uncertain or vague concepts as our thinking and language are non-binary. In real life, we add much more variation to our judgments and classifications than what can be given by a simple positive or negative expression. Instead of classes with sharp boundaries (membership: yes or no), each feature is represented using fuzzy datasets with degrees of membership. In this way, all locations can be considered; even those with minor effects on mosquito breeding habitats. For example, locations with a relatively low influence were included in the analysis but with a low degree of membership. We assigned the degree of membership to a class as a value between 0 (not suitable as habitat) and 1 (highly suitable). Values falling between these endpoints were mapped as degrees of membership (from lowly to highly favourable breeding environments). For the dengue vector habitat risk zone identification, the raster calculator (Spatial Analyst) was used. All fuzzy datasets were computed on the basis of equations obtained from machine-learning methods by modifying the raster calculator commands of each layer.

Results

The assessment of climatic and environmental factors using the confusion (error) matrix, computed using the "SCORER" component of the flow process of the Decision Tree learning method resulted in an overall accuracy of 80% (Table 2). This score was used in the development of near real-time monitoring application for dengue vector habitats using the fuzzy-based approach estimating the suitable range of factors suitable for habitats using the data mining approach. The scores and rules of those found suitable on the basis of the data mining algorithm were transferred to the equations used.

Table 2. Accuracy assessment chart for climatic and environmental factors.

|	Confusion matrix		
	Low	Medium	High
Low	8	12	2
Medium	3	91	2
High	1	6	6
Correctly classified	105	Accuracy	80.15 %
Wrongly classified	26	Error	19.84%

Fig. 8. Fuzzy logic graph for population distribution.
*Represents the degree of truth as an extension of assessment (pixel values vary between 0 and 1).

Using the GPW layer as source resulted in a maximum human population density of 46 persons, per 100 m². Increasing numbers of containers led to more mosquito habitats close to human dwellings and higher mosquito populations, thus intensifying the infection risk. The human population density showed a linear relationship with respect to its effect on the dengue larval density as depicted in Fig. 8.

The elevation results showed that altitude plays an important role in limiting the *Ae. aegypti* distribution. The maximum effect of elevation on *Aedes* breeding capacity was found to be between 400 and 1,100 m above MSL (Fig. 9). The SRTM data gave the maximum elevation for *Aedes* mosquitoes as approximately 1,800 m above MSL.

Humidity values in the range between 68 and 80% influenced the role of the vector highly in the study area (Fig. 10).

Results from the data mining approach showed that a temperature range between 30-35 °C had a high impact on dengue vector breeding (Fig. 11). The current global forecast of a 2.0 °C to 4.5 °C average worldwide warming by 2100 will certainly impact this and other vector-borne diseases (Kyle and Harris, 2008).

The rainfall graph presents 40-81 mm/hr as a favourable range for the dengue vectors in the study area. Clearly, rainfall in this range had a positive impact on breeding (Fig. 12).

Taken together, the *Aedes* habitats were found to be spatially clustered around places characterised by high human population densities and large-surface water bodies, such as marsh/swamps and rice paddies but also in areas with poor housing conditions and at elevations from 400 to 1,100 m above MSL, at humidity levels between 68 and 79%, at temperatures of 30-35 °C and where rainfall were between 41 and 80 mm/hr.

Fig. 13 shows the results of dengue habitat probability due to factors considered in the model. The period of April-July was found to be a period of significant,
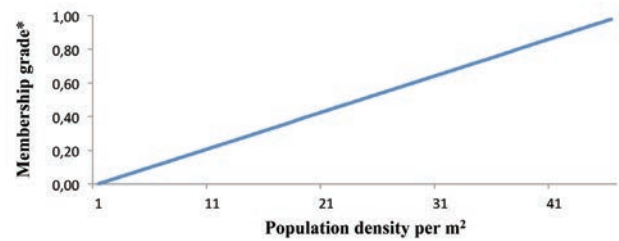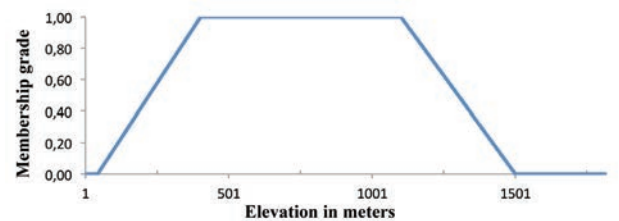

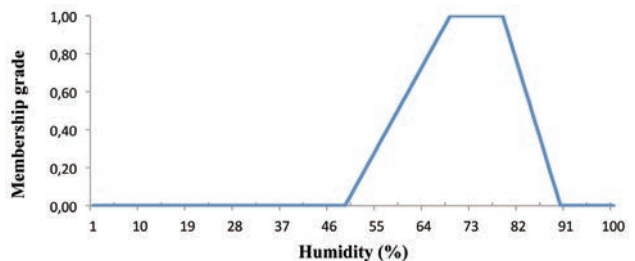Fig. 9. Fuzzy logic graph for altitude distribution.

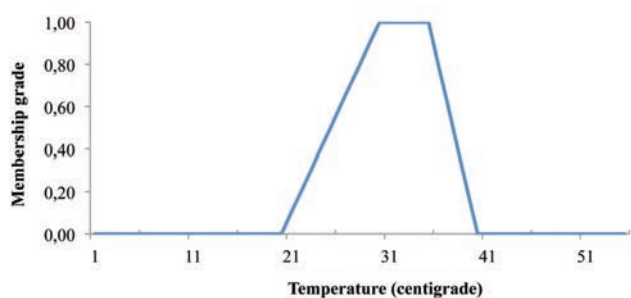Fig. 10. Fuzzy logic graph for humidity distribution.

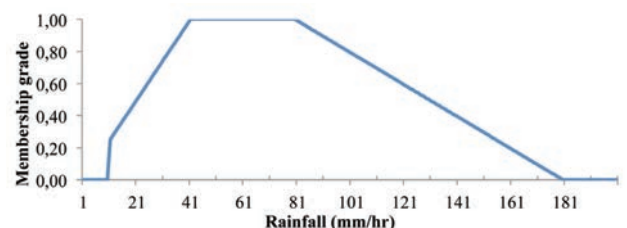Fig. 11. Fuzzy logic graph for temperature distribution.
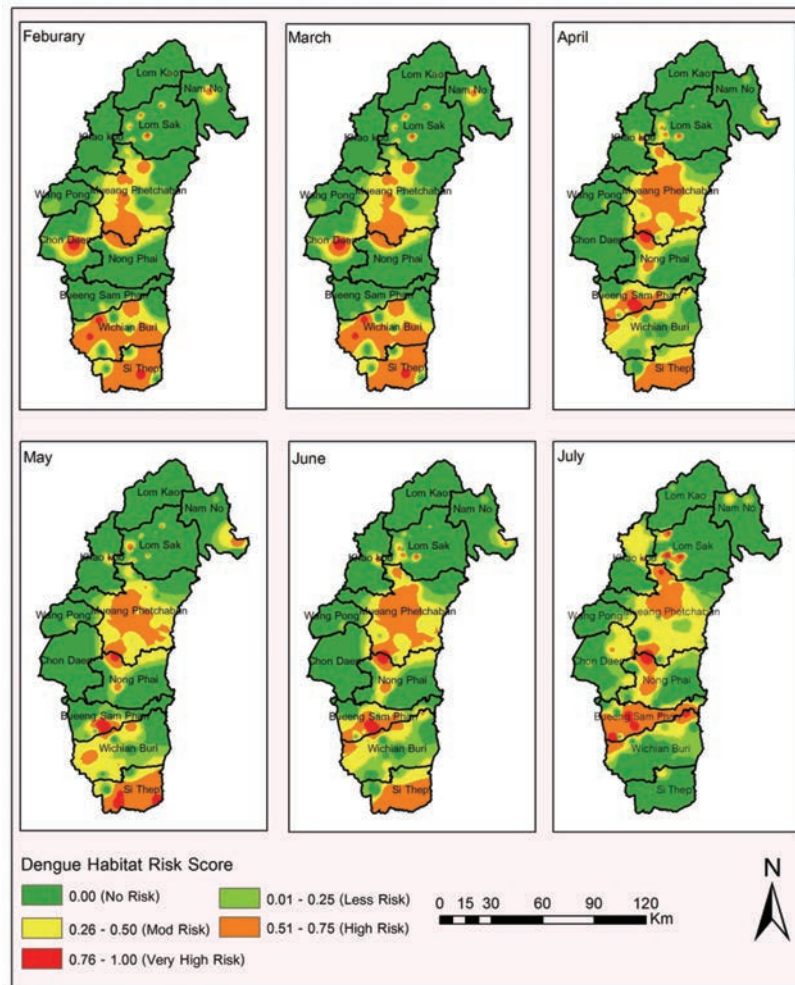

Fig. 12. Fuzzy logic graph for rainfall distribution.

Fig. 13. Dengue vector breeding habitat probability estimated.

high suitability, after which time the heavy rains begin and the number of successful breeding habitats suddenly decreases. The proposed model can be adopted to monitor dengue vector breeding throughout the year allowing health administrators to monitor favourable areas for mosquito breeding on a daily basis and to institute immediate, precautionary measures.

Discussion

Disease prediction is one of the more interesting and challenging tasks, where a data mining tool can be useful. Computers powered with automated applications and collected datasets can rapidly provide large volumes of medical data. Knowledge discovery in databases (KDD), which includes data mining techniques, has become a popular research tool for medical researchers to identify and exploit patterns and relationships among large number of variables. This

has made it possible to predict the outcome of a disease using stored, historical cases (Fathima et al., 2011). Data mining can thus discover inherent and previously unknown information and it can also support interactive data investigation. Some of the specific attributes include preprocessing attributes, evaluation and comparison of results for different schemes as well as designing of off-line comparative experiments.

In our fuzzy model, each parameter mentioned in the previous paragraph represents a data layer. We calculated fuzzy scores for these layers using ArcGIS fuzzy tool (setting gamma at 0.9) and produced overlays where the analysis determined the risk due to the breeding potential based on numerous sets, i.e. a multi-criteria overlay investigation. In our case, fuzzy overlay determined the phenomenon layers by combining all possible layers influencing the mosquito habitats by analysing the associations between the various degrees of membership of all considered parameters. It is also possible to fine-tune the model by

including expert opinions for changing the influence of various parameter levels. Consequently, decision-makers can add/express constraints through the use of natural language interfaces providing a means to analysis data and create maps without having an in-depth knowledge of the tools involved in building the application.

Most of the mosquito larval densities were found together with higher human population densities. Higher population accuracy is theoretically possible by using population density maps from local organisations. However, as there was not much change in the population density during the time of the study, this variable was kept fixed throughout.

Researchers in India report that the altitude range most suitable of for *Ae. aegypti* is <500 m above MSL and altitudes >1,200 m above MSL is the upper limit for *Aedes* (Kalra et al., 1997). In other regions of Southeast Asia, an altitude of >1,500 m above MSL limits *Ae. aegypti* breeding, while elevations above 1,500 m have been found to be favourable in a few regions (Patz et al., 1998). Dengue transmission typically occurs in periods of rainy days when temperatures and humidity levels are favourable for the increase of *Aedes* breeding in secondary habitats. Humidity with suitable temperature also increases mosquito survival (Rotela et al., 2007).

Besides biotic factors (i.e. virus, vector and host), abiotic factors like temperature, humidity and rainfall also play an important role for favourable breeding of the dengue vector. In arid regions, where rainfall is insufficient during the dry season, breeding mainly occurs in man-made storage containers rather than natural containers. The important role of the latter in the rainy season was confirmed in this study.

Limitations have to do with the dengue indices assessing dengue transmission risk, such as CI, HI and BI. However, these indices may constrain estimations, e.g. by defining immature forms of the vector found in the containers. In order to investigate the pragmatic trend for expected dengue risk, pupal surveys (mature forms of vectors) may aid in predicting dengue incidence potentially. Improvements in elevation data (SRTM-4), which have been processed to fill voids can yield better results for such types of studies. Furthermore, available satellite sensor data are limited by resolution issues and it is sometimes difficult to get cloud-free images, which restricts estimating factors for specific time periods. Furthermore, the values of locations in the fuzzy output map derived from data mining software using fuzzy inference processes have limitations due to the complexity of defining the habi-

tats of dengue mosquitos. In the future, adding infrastructure and socio-economic layers for settlements may further increase accuracy of the model around populated areas. In addition, results extracted at the local scale may not be used to extrapolate to the regional scale because factors influencing dengue vector densities vary due to ecological and social factors. Utilisation of geospatial methods in epidemiology will only be fulfilled once remote sensing data providers adapt to the needs of public health concerns.

Conclusions

This study emphasises the possibility of using free, satellite-based climate data and integrating it with geospatial data collected on the ground. The approach reduces the dependency on time-consuming and strenuous field data collection of climatic parameters from ground stations and integration with health data thus reducing the delay in prediction. Geospatial and remotely sensed data utilisation with proper statistical methods, such as fuzzy-logic analysis, could effectively predict the trend of vector-borne diseases.

Acknowledgements

References

Agrawal R, Imielinski T, Swami A, 1993. Mining association rules between sets of items in large databases. ACM SIGMOD Record 22, 207-216.

Asner G, Elmore A, Olander L, Martin R, Harris A, 2004. Grazing systems, ecosystem responses and global change. Annu Rev Environ Resour 29, 261-299.

Aziz AA, Rusli A, Hamidah I, 2011. A model of knowledge mapping in visualizing the hotspot of dengue. In: Open Systems (ICOS), IEEE Conference, 444-449 pp.

Banu S, Hu W, Hurst C, Tong S, 2011. Dengue transmission in the Asia-Pacific region: impact of climate change and socio-

environmental factors. Trop Med Int Health 16, 598-607.

Banu S, Wenbiao H, Yuming G, Cameron H, Shilu T, 2014. Projecting the impact of climate change on dengue transmission in Dhaka, Bangladesh. Environ Int 63, 137-142.

Beck LT, Lobitz BM, Wood BL, 2000. Remote sensing and human health: new sensors and new opportunities. Emerg Infect Dis 6, 217-227.

Beebe NW, Cooper RD, Mottram P, Sweeney AW, 2009. Australia's dengue risk driven by human adaptation to climate change. PLoS Negl Trop Dis 3, e429.

Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B, 2008. KNIME: The Konstanz information miner. In: Data analysis, machine learning and applications. Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds). Freiburg im Breisgau: Spinger, 319-326 pp.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O et al., 2013. The global distribution and burden of dengue. Nature 496, 504-507.

Buczak AL, Koshute PT, Babin SM, Feighner BH, Lewis SH, 2012. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. BMC Med Inform Decis Mak 12, 124.

Ceccato P, Connor SJ, Jeanne I, Thomson MC, 2005. Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. Parassitologia 47, 81-96.

CIESIN - Center for International Earth Science Information Network, 2012. Columbia University. Available at: http://www.sedac.ciesin.columbia.edu/gpw (accessed on March 2012).

Craig MH, Snow RW, Sueur D, 1999. A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitol Today 15, 105-111.

Dambach P, Machault V, Lacaux JP, Vignolles C, Sié A, Sauerborn R, 2012. Utilization of combined remote sensing techniques to detect environmental variables influencing malaria vector densities in rural West Africa. Int J Health Geogr 11, 8.

Fathima AS, Manimegalai D, Nisar H, 2011. A Review of data mining classification techniques applied for diagnosis and prognosis of the arbovirus-dengue. Int J Comput Sci 8, 322-328.

Focks DA, Daniels E, Haile DG, Keesling JE, 1995. A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. Am J Trop Med Hyg 53, 489-506.

Focks DA, Haile D, Daniels E, Mount G, 1993. Dynamic life table model for *Aedes aegypti* (L) (Diptera: Culicidae). Analysis of the literature and model development. J Med Entomol 30, 1003-1017.

Goto K, Kumarendran B, Mettananda S, Gunasekara D, Fujii Y, Kaneko S, 2013. Analysis of effects of meteorological factors on dengue incidence in Sri Lanka using time series data. PLoS One 8, e63717.

Gubler DJ, 2011. Dengue, urbanization and globalization: the unholy trinity of the 21st century. Trop Med Health 39, 3.

Hassan H, Shoaimi S, Hashim NR, 2012. Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. Geospat Health 7, 21-25.

Hay SI, George DB, Moyes CL, Brownstein JS, 2013. Big data opportunities for global infectious disease surveillance. PLoS Med 10, e1001413.

Holzinger A, Zupan M, 2013. KNODWAT: A scientific framework application for testing knowledge discovery methods for the biomedical domain. BMC Bioinformatics 14, 191.

Hsieh YH, Chen CW, 2009. Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks. Trop Med Int Health 14, 628-638.

Jeefoo P, Tripathi NK, Souris M, 2010. Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao Province, Thailand. Int J Environ Res Public Health 8, 51-74.

Kalra NL, Kaul SM, Rastogi RM, 1997. Prevalence of *Aedes aegypti* and *Aedes albopictus*-vectors of dengue haemorrhagic fever in north, north-east and central India. Dengue Bull 21, 84-92.

Khormi HM, Kumar L, 2012. Assessing the risk for dengue fever based on socioeconomic and environmental variables in a geographical information system environment. Geospat Health 6, 171-176.

Kummerow C, Barnes W, Kozu T, Shiue J, Simpson J, 1998. The Tropical Rainfall Measuring Mission (TRMM) sensor package. J Atmos Oceanic Technol 15, 809-817.

Kyle JL, Harris E, 2008. Global spread and persistence of dengue. Ann Rev Microbiol 62, 71-92.

Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, Thomas MB, Scott TW, 2011. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. Proc Natl Acad Sci U S A 108, 7460-7465.

LP DAAC - Land Processes Distributed Active Archive Center, 2012. Available at: https://www,lpdaac.usgs.gov/products/modis_products_table (accessed on April 2012).

Machault V, Vignolles C, Borchi F, Vounatsou P, Briolant S, Lacaux JP, Rogier C, 2011. The use of remotely sensed environmental data in the study of malaria. Geospat Health 5, 151-168.

Morin CW, Andrew CC, Kacey E, 2013. Climate and dengue transmission: evidence and implications. Environ Health Perspect 121, 1264-1272.

Myers SS, Patz JA, 2009. Emerging threats to human health from global environmental change. Annu Rev Environ Resour 34, 223-252.

Nakhapakorn K, Tripathi NK, 2005. An information value

based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. Int J Health Geogr 4, 13.

Patz JA, Martens WJ, Focks DA, Jetten TH, 1998. Dengue fever epidemic potential as projected by general circulation models of global climate change. Environ Health Perspect 106, 147-153.

Patz JA, Norris DE, 2004. Land use change and human health. In: Ecosystems and land use change, geophysical monograph 153. DeFries RS, Asner G, Houghton RA (eds). Washington DC, American Geophysical Union, 159-167 pp.

Quintero J, Brochero H, Manrique-Saide P, Barrera-Pérez M, Basso C, Romero S, Petzold M, 2014. Ecological, biological and social dimensions of dengue vector breeding in five urban settings of Latin America: a multi-country study. BMC Infect Dis 14, 38.

Ramankutty N, Foley J, 1999. Estimating historical changes in global land cover: croplands from 1700 to 1992. Glob Biogeochem Cycles 13, 997-1027.

Reiter P, 1996. Global warming and mosquito-borne disease in USA. Lancet 348, 622.

Rotela C, Fouque F, Lamfri M, Sabatier P, Introini V, Zaidenberg M, Scavuzzo C, 2007. Space-time analysis of the dengue spreading dynamics in the 2004 Tartagal outbreak, Northern Argentina. Acta Trop 103, 1-13.

Sarfraz MS, Tripathi NK, Kitamoto A, 2013. Near real-time characterisation of urban environments: a holistic approach for monitoring dengue fever risk areas. Int J Digital Earth 7, 916-934.

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

Simmons CP, Farrar JJ, Nguyen VC, Wills B, 2012. Dengue. N Engl J Med 366, 1423-1432.

Vanwambeke SO, Bennett SN, Kapan DD, 2011. Spatially dis-aggregated disease transmission risk: land cover, land use and risk of dengue transmission on the island of Oahu. Trop Med Int Health 16, 174-185.

WHO, 2012. Dengue and severe dengue. Fact sheet no 117, retrieved May 2012. Geneva: World Health Organization. Available at: http://www.who.int/mediacentre/factsheets/fs117/en/ (accessed on May 2012).