Time-series modelling of dengue incidence in the Mekong Delta region of Viet Nam using remote sensing data

Nga TT Pham, a Cong T Nguyen a and Maria Ruth B Pineda-Cortel b

Objective: This study aims to enhance the capacity of dengue prediction by investigating the relationship of dengue incidence with climate and environmental factors in the Mekong Delta region (MDR) of Viet Nam by using remote sensing data.

Methods: To produce monthly data sets for each province, we extracted and aggregated precipitation data from the Global Satellite Mapping of Precipitation project and land surface temperatures and normalized difference vegetation indexes from the Moderate Resolution Imaging Spectroradiometer satellite observations. Monthly data sets from 2000 to 2016 were used to construct autoregressive integrated moving average (ARIMA) models to predict dengue incidence for 12 provinces across the study region.

Results: The final models were able to predict dengue incidence from January to December 2016 that concurred with the observation that dengue epidemics occur mostly in rainy seasons. As a result, the obtained model presents a good fit at a regional level with the correlation value of 0.65 between predicted and reported dengue cases; nevertheless, its performance declines at the subregional scale.

Conclusion: We demonstrated the use of remote sensing data in time-series to develop a model of dengue incidence in the MDR of Viet Nam. Results indicated that this approach could be an effective method to predict regional dengue incidence and its trends.

According to the World Health Organization (WHO), 1 Viet Nam is among the top 10 countries with the highest reported number of dengue cases in the world (91,321 cases in 2012). Studies have shown that dengue epidemics in Viet Nam occurred cyclically every 3–5 years and peaked approximately every 10 years. 2 These cycles are thought to be influenced by the circulating viral serotypes, host immunity and climate oscillations. 3 Dengue transmission occurs throughout the year in Viet Nam with peak numbers of cases reported in the rainy season from May to October. 4 Since 2007, dengue has been recorded in 55 of the 63 provinces in Viet Nam, increasing from north to south with the Mekong Delta region (MDR) experiencing the highest incidence recorded during the years 2000 to 2016.

Several recent studies have aimed to better understand the dynamics of dengue and the influences of environmental factors on the disease and to better predict outbreaks. Climate factors, in addition to multiple human, biological and ecological determinants, influence the emergence and re-emergence of infectious diseases, including dengue, 5 which is transmitted by both the primary vector Aedes aegypti and the secondary vector Aedes albopictus. 6,7 Studies have found a significant correlation between rain and dengue incidence in Metropolitan Manila, Philippines from 1996 to 2005, 8 and a correlation between temperature, rain and dengue incidence in southern Thailand by multiple regression analysis. 9 On a regional scale, a review of the impacts of climate change on human health provided more evidence of the burden of climate change–attributable diseases and emphasized the uncertainty in attributing diseases to climate change, owing to a lack of long-term, high-quality data. 10 Climate change is likely to affect the seasonal and geographical distribution of dengue fever in...
the Asia–Pacific region, but more studies are needed that adjust for regional and subregional socio-environmental factors in the assessment of climate effects on dengue transmission.\textsuperscript{5} Climate is only one of many environmental factors; changes in land cover by human settlements, the presence of water bodies, and vegetation type also affect dengue transmission processes.\textsuperscript{11}

A range of approaches, including statistical modelling, mathematical modelling and spatial analysis, have been applied to demonstrate relationships between dengue and climate variables and to predict dengue cases and outbreaks.\textsuperscript{12, 13} Statistical models that are commonly constructed to predict dengue incidence cannot precisely predict the time and place of a dengue outbreak. However, they are able to quantitatively associate climatic factors such as rain, temperature and humidity with dengue epidemics at certain geographic areas with specific time lags.\textsuperscript{14}

We assumed that there was a strong association between dengue incidences and climate variables; therefore, we applied a time-series autoregressive integrated moving average (ARIMA) model for dengue prediction in the MDR of Viet Nam. To minimize the limitations of climate data from meteorological stations in spatial and time scales, we used the most accessible remote sensing data for climate variables: the Global Satellite Mapping of Precipitation (GSMaP) for rain data and the Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (\textit{LST}_d\textsuperscript{17} and \textit{LST}_n\textsuperscript{17}) with a 1 km spatial resolution as a proxy for air temperature. In addition to climate variables, one of the most commonly used remote sensing–derived environmental variables, the normalized difference vegetation index (NDVI) from MODIS (MOD13Q1) with a 250 m spatial resolution, was also used in the model for its influence on dengue.\textsuperscript{19} These remote sensing–based parameters were aggregated to compute mean monthly variables for each province as examples presented in Fig. 2, showing clear spatial variations among provinces and between variables.

Statistical analysis

We used the Box-Jenkins methodology\textsuperscript{20} to fit ARIMA models to monthly dengue incidence in 12 provinces, using the statistical forecast package in RStudio software (version 1.1) (RStudio, Boston, MA, USA).\textsuperscript{21} Dengue case definitions were based on WHO criteria\textsuperscript{22} and collected through the disease surveillance systems according to Viet Nam’s Ministry of Health regulations.\textsuperscript{23} The dengue cases reported from 2001 to 2015 were used for developing the time series model, and the cases during 2016 were used for validating the model. To avoid effects from the non-constant variance, we stabilized dengue counts by natural log transformation.

First, to confirm that ARIMA models were suitable for this analysis, we examined the data for seasonality and
Remote sensing based time–series modelling of dengue in the Mekong Delta region of Viet Nam

Nga et al

Fig. 1. Average monthly numbers of dengue cases for 13 provinces in the Mekong Delta region, divided into subregions I and II by a white line, 2000–2016

Fig. 2. Examples of monthly estimates for (a) GSMaP rain, (b) MODIS LSTd and (c) MODIS NDVI in the Mekong Delta region, November 2010

Interannual variations of dengue incidence and climate and environmental variables (rain, LSTd, LSTn, and NDVI) during the period 2000–2016 for each province. Then, the adequacy of each model for each province was verified by histogram, by autocorrelation of the standardized residuals, and by the Ljung-Box test, similar to previous studies. Next, the structure of the model followed the standard form for ARIMA, \((p,d,q)(P,D,Q)_s\), where \(p\) is the order of autoregression; \(d\), the degree of differencing; \(q\), the order of the moving average; \(P\), the seasonal autoregression; \(D\), the degree of seasonal differences; \(Q\), the seasonal moving average; and \(s\), the seasonal period.
Different ARIMA model forms (combinations of $p$, $d$, $q$, $P$, $D$ and $Q$) were tested to fit the log-transformed time series data without environmental covariates. The best ARIMA model was selected as that with the lowest Akaike information criterion, a measure of the relative goodness of fit of a model across the 12 provinces (Hau Giang province was excluded as it had politically separated from Can Tho province in 2004). Then, multivariate ARIMA models were fitted with log-transformed dengue cases in relation to all the climate variables with time lags that were chosen by their best correlation with dengue.

RESULTS

Seasonal variation of dengue and climate parameters

Fig. 3 presents an example of the time series of dengue, rain, $LST_d$ and NDVI during 2000-2016 for An Giang province. The plots showed strong seasonal and interannual variations of all variables. We found this seasonal pattern apparent and consistent for all provinces in the region with higher dengue cases coinciding with rainy seasons. This enabled us to apply the ARIMA model to the entire MDR.

Model in association with variables

Fig. 4a presents an example of an ARIMA fitting model plot for An Giang province and the comparison of fitted with reported dengue cases; Fig. 4b shows the regression function with its root mean square error value for the fitted period of 2000–2015. The final model for each province was confirmed by the Ljung-box test of the residual with no correlation for fitted data.

We generated 12 final models that closely fitted dengue incidence from every province in incorporating climate and environment variables as external predictors. After careful screening, those variables with the highest correlation with dengue incidence at specific time lags were kept in the final ARIMA models (Table 1) that show the correlation value, and the monthly time lag is shown in brackets. Correlations between these variables and dengue are spread over a range of time lags across provinces. Dengue had the strongest correlation with rain at a 1-month time lag across all provinces, reaching 0.60 at Long An province; dengue had a comparable correlation with $LST_d$ and $LST_n$ at 1-month to 4-month time lags. In addition, dengue had both positive and negative correlation with NDVI at 4-month to 5-month time lags. Notably, we found that the two subregions had different climate and environmental influences on dengue based on their correlations and time lags. In subregion I, dengue was found to have a higher correlation with rain with a 1-month time lag, and NDVI had a 5-month time lag; dengue was also associated with $LST_d$ generally with a 4-month time lag and $LST_n$ with a 1-month time lag. These findings are different than those for provinces in subregion II along the coast of the MDR where dengue was found to have a weaker correlation with rain and a negative correlation with NDVI with a 4-month time lag. For provinces in subregion II, dengue was correlated with a 1-month shorter time lag for $LST_d$ but a 1-month longer time lag for $LST_n$ compared to subregion I. For other provinces, including Vinh Long, Can Tho, and Kien Giang, dengue was found to have relatively low correlations with all variables and at mixed time lags. The variability in the association between dengue and climate and environmental factors across provinces in the study region emphasizes the need for a separate time-series model for each province.

Model validation

The best time-series ARIMA model with final independent variables found for each province was applied to predict dengue for the period from January to December 2016. In most cases, rain and $LST_d$ remained in the final model; $LST_d$ and NDVI were occasionally removed when correlations were less than 0.25. We compared predicted dengue with reported cases for the whole MDR, as shown in Fig. 5. The results showed that predicted dengue in every province closely followed the trend of reported data (Fig. 5a) and that these data are in good linear regression with the square of correlation of 0.65 for the whole region (Fig. 5b).

We also evaluated the association between monthly predicted and reported dengue incidence from January to December 2016 by calculating the Pearson correlation coefficient for each province (the numbers presented correspond to the provinces in Fig. 5a). The correlation varies significantly, from approximately 0.22 to 0.85 (with $P < 0.05$) for different provinces; nine provinces had a correlation greater than 0.50, and only three provinces had a correlation below 0.50. We found that the predicted dengue during the period January to December
Remote sensing based time–series modelling of dengue in the Mekong Delta region of Viet Nam

Nga et al

Fig. 3. Time series of (a) dengue, (b) rain, (c) LSTn, and (d) NDVI for An Giang province, 2000–2015

Fig. 4. Example of ARIMA model for An Giang province: (a) model fitting and (b) scatter plot of fitted and reported dengue cases

Table 1. Final ARIMA models with correlation of dengue with climate and environmental variables for each province in time lags

LSTn = land surface temperature, night
NDVI = normalized difference vegetation index
Nga et al. Remote sensing based time-series modelling of dengue in the Mekong Delta region of Viet Nam

In subregion II and lower in subregion I, controverting the distribution of dengue among these two subregions (as in Fig. 1). This is a limitation of ARIMA modelled with only climate factors relating to vector abundance; transmission of dengue is also affected by other factors such as population density and activities, relating human susceptibility to the disease. Therefore, an assessment of prevailing vulnerabilities to dengue could be independent of its spatial distribution, but closely related to a combination of climate, environment and sociodemographic conditions.

Possible reasons of inconsistent effects of NDVI on dengue incidence include subprovincial variations in precipitation and a variety of land covers. Climatologically, the onset of the rainy season in the MDR usually appears first along the coast (subregion II) and moves gradually inland (toward subregion I), resulting in a different temporal time lag in relationships between rain and dengue transmission over the region. Also, vegetation type and growth stage may play important roles in determining vector abundance, irrespective of their association with rain.

An analysis of NDVI distribution in relation to land cover data over the whole region indicated that lower values of NDVI in subregion II corresponded to more water bodies, shrubs and mixed horticulture land cover types. Higher NDVIs were found in subregion I, which corresponded to more rice paddy land cover.

Several investigators have examined the associations among climate variables, demography and dengue incidence in the southern provinces of Viet Nam using a wavelet time series analysis, a standard multiple

DISCUSSION

In an effort to overcome the limitations of spatial and time scales in climate data collected from meteorological stations, we created a high-quality data set of satellite remote sensing data for climate and environmental factors, i.e. rain data from GSMaP and temperature and vegetation data from MODIS for the entire MDR for dengue modelling using a time-series approach. Different combinations of components were evaluated to construct the best predictive ARIMA models for 12 provinces across the study region. We decided to use the ARIMA model because it can cope with a stochastic dependence of consecutive data and to account for autocorrelations in time series as well as seasonality, long-term trends and time lags. The selection of external variables was based on the best correlation of dengue with rain, LST_n, LST_d, and NDVI at different time lags for the 2000 to 2015 data. The results of the validation showed different performances of the ARIMA model over the region. The correlation of predicted and reported dengue during the period January to December 2016 was found higher in subregion II and lower in subregion I, controverting the distribution of dengue among these two subregions (as in Fig. 1). This is a limitation of ARIMA modelled with only climate factors relating to vector abundance; transmission of dengue is also affected by other factors such as population density and activities, relating human susceptibility to the disease. Therefore, an assessment of prevailing vulnerabilities to dengue could be independent of its spatial distribution, but closely related to a combination of climate, environment and sociodemographic conditions.

Possible reasons of inconsistent effects of NDVI on dengue incidence include subprovincial variations in precipitation and a variety of land covers. Climatologically, the onset of the rainy season in the MDR usually appears first along the coast (subregion II) and moves gradually inland (toward subregion I), resulting in a different temporal time lag in relationships between rain and dengue transmission over the region. Also, vegetation type and growth stage may play important roles in determining vector abundance, irrespective of their association with rain. An analysis of NDVI distribution in relation to land cover data over the whole region indicated that lower values of NDVI in subregion II corresponded to more water bodies, shrubs and mixed horticulture land cover types. Higher NDVIs were found in subregion I, which corresponded to more rice paddy land cover.

Several investigators have examined the associations among climate variables, demography and dengue incidence in the southern provinces of Viet Nam using a wavelet time series analysis, a standard multiple
regression model, ARIMA and a Poisson distributed lag model.\textsuperscript{33} We found that time-series analysis was useful in establishing the relationship between the change of weather parameters, environmental factors and the incidence of dengue diseases for the entire MDR. Input variables were selected based on the best correlation of dengue with precipitation, LST\textsubscript{d}, LST\textsubscript{n}, and NDVI data, for the period 2000 to 2015 at varying time lags as in previous studies.\textsuperscript{19,33}

**Limitations**

One challenge to the existing statistical approaches in modelling dengue is the difficulty of quantifying the influences of myriad human activities on vector-borne transmission. An integrated approach might be to build an index framework for underlying socioeconomic and demographic factors.\textsuperscript{29} Disease transmission is also closely linked with mosquito behaviour and population dynamics that are largely influenced by climate factors.\textsuperscript{5} Therefore, direct modelling by using climate data could be useful in informing health and sanitation sectors of potential increases in mosquito activity and subsequent disease risk, especially if mosquito population data can be sampled and integrated.

**Conclusions**

This study focused on presenting climate and environmental factors from remote sensing data in modelling and predicting dengue fever in the MDR of Viet Nam. Our results indicate that this approach may be effective for predicting regional dengue incidence and trends. The results also revealed that the higher correlation of dengue using a single variable does not improve the model performance, and validation of the model is crucial for assessing its accuracy. Our findings support previously made conclusions that dengue prediction models vary due to their complexity and methodology and are dependent on the type of data collected and the nature of the variables.\textsuperscript{14} No universal models exist for global analysis and prediction,\textsuperscript{14} even if limited to a climatologically homogeneous area such as the MDR.

We describe the first effort to apply remote sensing data to perform time-series modelling of a vector-borne disease in Viet Nam to enhance the dengue early warning system. Similar approaches have been used throughout the world.\textsuperscript{34–36} Integration of remote sensing and modelling to provide early warning of vector-borne disease outbreaks has been successfully demonstrated for malaria throughout Africa\textsuperscript{35} and for dengue epidemics in Brazil.\textsuperscript{14} To determine our model’s usefulness as an early warning tool, the results of our study have been presented on the Internet not only for the MDR, but also for other regions of Viet Nam and the Philippines. However, the system should be evaluated by end users for its effectiveness for dengue predictions for two countries.

**Acknowledgements**

We appreciate the constructive comments from Dr Amelia Kasper for improving the manuscript.

**Funding**

This study was funded by the Asia Pacific Network for Global Change Research (APN) - CAF2016-RR11-CMY-Pham.

**Conflicts of interest**

No potential conflicts of interest were reported by the authors.

**References**

1. World Health Organization. Global strategy for dengue prevention and control 2012–2020. Geneva: WHO; 2012. Available from: https://apps.who.int/iris/bitstream/handle/10665/78303/9789241504034_eng.pdf

2. Vu HH, Okumura J, Hashizume M, Tran DN, Yamamoto T. Regional differences in the growing incidence of dengue fever in Vietnam explained by weather variability. Trop Med Health. 2014 Mar;42(1):25–33. doi:10.2149/tmh.2013-24 pmid:24808744

3. Cuong HQ, Vu NT, Cazelles B, Boni MF, Thai KT, Rabaa MA, et al. Spatiotemporal dynamics of dengue epidemics, southern Vietnam. Emerg Infect Dis. 2013 Jun;19(6):945–53. doi:10.3201/eid1906.121323 pmid:23735713

4. Lee HS, Nguyen-Viet H, Nam VS, Lee M, Won S, Duc PP, et al. Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. BMC Infect Dis. 2017 Mar 20;17(1):218. doi:10.1186/s12879-017-2326-8 pmid:28320341

5. Banu S, Hu W, Hurst C, Tong S. Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors. Trop Med Int Health. 2011 May;16(5):598–607. doi:10.1111/j.1365-3156.2011.02734.x pmid:21320241

6. Higa Y, Yen NT, Kawada H, Son TH, Hoa NT, Takagi M. Geographic distribution of Aedes aegypti and Aedes albopictus collected from used tires in Vietnam. J Am Mosq Control Assoc. 2010 Mar;26(1):1–9. doi:10.2987/09-5945.1 pmid:20402344
Eastin MD, Delmelle E, Casas I, Wexler J, Self C. Intra- and inter-seasonal periodicity, synchronicity and climate variability. PLoS Negl Trop Dis. 2011 Jul-Aug;44(4):436–40. doi:10.1590/S0037-89162011000400007

Hyndman RJ, Khandakar Y. Automatic time series for forecasting: the forecast package for R (No. 6/07). Monash University, Department of Econometrics and Business Statistics; 2007. Available from: https://www.monash.edu/business/econometrics-and-business-statistics/research/publications/eb/06-07.pdf

World Health Organization. Dengue: guidelines for diagnosis, treatment, prevention and control. Geneva: WHO; 2009. Available from: https://apps.who.int/iris/bitstream/handle/10665/164135/9789241547871_eng.pdf

Nga et al. Remote sensing based time–series modelling of dengue in the Mekong Delta region of Viet Nam.
33. Phung D, Huang C, Rutherford S, Chu C, Wang X, Nguyen M, et al. Identification of the prediction model for dengue incidence in Can Tho city, a Mekong Delta area in Vietnam. Acta Trop. 2015 Jan;141 Pt A:88–96. doi:10.1016/j.actatropica.2014.10.005 pmid:25447266

34. Chuang TW, Henebry GM, Kimball JS, Vanroekel-Patton DL, Hildreth MB, Wimberly MC. Satellite microwave remote sensing for environmental modeling of mosquito population dynamics. Remote Sens Environ. 2012 Oct;125:147–56. doi:10.1016/j.rse.2012.07.018 pmid:23049143

35. Ceccato P, Vancutsem C, Klaver R, Rowland J, Connor SJ. A vectorial capacity product to monitor changing malaria transmission potential in epidemic regions of Africa. J Trop Med. 2012;2012:doi:10.1155/2012/595948 pmid:22363350

36. Health vulnerability to climate change. Hanoi: Vietnam National Space Center; updated 12 February 2019. Available from: http://apn-climateandhealth.com/