Application of Advanced Very High Resolution Radiometer (AVHRR)-based Vegetation Health Indices for Estimation of Malaria Cases

Atiqur Rahman,* Nir Krakauer, Leonid Roytman, Mitch Goldberg, and Felix Kogan

National Oceanic and Atmospheric Administration Cooperative Remote Sensing Science and Technology (NOAA-CREST), The City College of New York, New York; National Oceanic and Atmospheric Administration/National Environmental Satellite, Data, and Information Service (NOAA/NESDIS), Camp Spring, Maryland

Abstract. Satellite data may be used to map climatic conditions conducive to malaria outbreaks, assisting in the targeting of public health interventions to mitigate the worldwide increase in incidence of the mosquito-transmitted disease. This work analyzes correlation between malaria cases and vegetation health (VH) indices derived from satellite remote sensing for each week over a period of 14 years for Bandarban, Bangladesh. Correlation analysis showed that years with a high summer temperature condition index (TCI) tended to be those with high malaria incidence. Principal components regression was performed on patterns of weekly TCI during each of the two annual malaria seasons to construct a model as a function of the TCI. These models reduced the malaria estimation error variance by 57% if first-peak (June–July) TCI was used as the estimator and 74% if second-peak (August–September) was used, compared with an estimation of average number of malaria cases for each year.

INTRODUCTION

Malaria has become one of the greatest health concerns in Bangladesh, in common with many other developing countries. The malaria endemic region of Bangladesh is located along the eastern and southern border, with about 35 million people at risk. Malaria constrains regional and nationwide economic development, causing deaths, morbidity, and economic losses. Economists believe that vector-borne diseases are responsible for an economic growth penalty of up to 1.3% per year. The risk of contracting malaria in endemic areas deters investment, which has negative impact on economic productivity. Efforts by the Government of Bangladesh to eradicate malaria have met with only limited success. To manage malaria epidemics effectively, especially in the years of their intensification, public health agencies can benefit from advance knowledge of when and where climate conditions favorable to vector proliferation could be expected, enabling epidemic containment and treatment efforts to be focused where they will be most needed.

Here, we look at remotely sensed climatic correlates to years of heavy malaria outbreaks in Bandarban district, southeastern border of Bangladesh, bounded between 22°N and 22.5°N latitude and 91.5°E and 92.5°E longitudes. This is a hilly area mostly covered with tropical evergreen forest.

MALARIA, MOSQUITO, AND ENVIRONMENT

Anopheles dirus (An. dirus) is the most widespread malaria vector mosquito in forested areas of Southeast Asia. Breeding habitats of An. dirus include puddles on footpaths and turbulence pits at the heads of drainage gullies, which hold water for some time without supplemental rainfall. In Bangladesh, An. dirus and the malaria parasite (Plasmodium spp.) are mostly restricted to the hilly districts in the east and south. The forested, green hills of Bandarban district are at the center of the malaria endemic area. The forest and forest fringe areas report more than 90% of total positive cases and more than 70% of total Plasmodium falciparum cases in all districts of Bangladesh. Because mosquitoes are sensitive to moisture and temperature conditions, weather monitoring can give an indication of their activity level and hence the risk of malaria transmission. Therefore, this work investigates the application of satellite remote sensing for the characterization of weather conditions in the Bandarban district and their impacts on interannual variation in malaria incidence.

Bandarban district is located in Chittagong division, the south-eastern part of Bangladesh. Total annual rainfall amount is around 3,000 mm per year, average temperature 21–22°C, and relative humidity around 80%. Monthly temperature and humidity are stable from year to year (standard deviations are on the order of 1°C and 1%, respectively), but interannual variation in total annual rainfall can be large (1,000 mm/year). In general, two seasons are defined in the annual cycle: wet and warm during April–October and cool and dry during December–February. During the wet season, the maximum temperature can reach 30°C or above. Variation in monthly rainfall between dry and wet months is the main influence on mosquito activity. In summer, the wet, warm, and humid climate is quite favorable for mosquito activity and produces an increase in the number of malaria cases.

Three principal environmental factors for mosquito activity and malaria epidemiology are important: temperature, humidity, and rainfall. The optimum temperatures for mosquitoes’ development and activity are 25–27°C. If daytime temperature exceeds 40°C, mosquitoes are less active and parasite transmission is limited. In general, a larger amount of rainfall stimulates mosquitoes’ development. However, frequent rainfall during the monsoon period might produce a reduction in malaria transmission, because it washes out eggs and reduces chances for development of adult mosquitoes. In the environment of Bandarban district, An. dirus females stay active during the period when precipitation exceeds 50 mm per month. However, a combination of large rainfall and hot weather during June–August might reduce mosquito activity. Mosquito activity and malaria transmission are also reduced if the humidity drops below 60%.

* Address correspondence to Atiqur Rahman, NOAA-CREST, The City College of New York, NY. E-mail: fmatiq@gmail.com
In our previous work, we analyzed 10 years of malaria data for Bangladesh as a whole.\textsuperscript{1} Correlation and regression analyses were used to build a predictive model using satellite remote sensing data. In this work, we extend the malaria and satellite series to 14 years, focus on a single district, investigate principal component regression for building a more robust estimator model, and quantify model skill using leave-one-out cross-validation.

**DATA AND METHODS**

Available public health data included district population; numbers of patients whose blood sample were tested for malaria and number of positive tests for malaria. Satellite remote sensing data was used in the form of vegetation indices derived from surface radiance in visible and infrared wavelengths.

**Public health data.** Malaria statistics were collected from Director General of Health, Bangladesh Ministry of Health. Malaria data were represented by annual number of people with fever or history of fever within the last 48 hours, an absence of signs of other disease, and inadequate antimalarials (or none) during the 4 weeks before present illness in local hospital. Standard blood slide examination was done by a medical officer trained in malaria microscopy, according to the World Health Organization (WHO) Guidelines and slide readers were blinded to the clinical diagnoses. Ten percent each of the positive and negative slides were reexamined blindly by the National Malaria Reference Laboratory to evaluate the accuracy of the study’s slide examination results. Tests could detect both \textit{P. falciparum} and \textit{P. vivax}. The hospital data were aggregated by local administrative unit health centers and further by administrative districts.\textsuperscript{7,14} Available malaria statistics are the number of persons who were tested for malaria in a hospital and the number of positive malaria cases. The dynamics of annual number of persons who were tested for malaria and number of positive malaria cases for the Bandarban district during the investigated period along with population size are shown in Figure 1. Although the district population increases, the number of persons who were tested for malaria and number of positive malaria cases are slightly decreasing, indicating some success of malaria prevention measures such as bed net distribution. If the number of positive malaria cases is expressed as a percent of the number of persons who were tested for malaria, as is normally done in malaria research as an indicator of malaria incidence, a decreasing trend in malaria incidence is also seen.

To determine the interannual difference between malaria rates, a linear trend was calculated from the time series of percent malaria cases to highlight interannual variation in malaria prevalence that is likely in part caused by interannual difference in weather patterns.\textsuperscript{13,15} The weather-related variations around the trend were expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 1):

\[
\text{DY} = \left( \frac{Y}{Y_{\text{trend}}} \right) \times 100, \quad (1)
\]

where \(Y\) is the actual percent of malaria cases, \(Y_{\text{trend}}\) is the malaria estimated from trend, and \(DY\) is the deviation (%) from the trend. The DY is a good relative metric of malaria cases. In 1998, DY was 119% or 19% above the trend, whereas in 1992, DY was 87% or 13% below the trend. Thus, malaria outbreaks were relatively light in 1992 (lower DY), whereas relatively severe in 1998 (higher DY).\textsuperscript{13,16}

**Satellite data: Vegetation health indices.** The Advanced Very High Resolution Radiometer (AVHRR) is a satellite-mounted instrument that measures radiances from the earth’s surface in several visible and infrared bands, achieving global coverage several times daily at a spatial resolution of 1 km\(^2\). The AVHRR-measured satellite data for solar energy reflected/emitted from the land surface were collected from the United States National Oceanic and Atmosphere Administration (NOAA) Global Vegetation Index (GVI) data set from 1992 through 2005. The AVHRR sensor counts in the visible (VIS, 0.58–0.68 µm, band1), near infrared (NIR, 0.72–1.00 µm, band2), and infrared (IR, 10.3–11.3 µm, band4 and 11.3–12.3 µm, band5) spectral regions were used to generate the GVI. Post-launch-calibrated VIS and NIR counts were converted to reflectance\textsuperscript{17} and used to calculate the Normalized Difference Vegetation Index (NDVI = [NIR – VIS]/[NIR + VIS]). The band 4 counts were converted to brightness (radiative) temperature (BT).

The Vegetation Condition Index (VCI), characterizing moisture, Temperature Condition Index (TCI), characterizing thermal conditions, and Vegetation Health Index (VHI), were calculated from NDVI and BT as\textsuperscript{18}

\[
\text{VCI} = 100(\text{NDVI} – \text{NDVImin})/(\text{NDVImax} – \text{NDVImin}), \quad (2)
\]

\[
\text{TCI} = 100(\text{BTmax} – \text{BT})/(\text{BTmax} – \text{BTmin}), \quad (3)
\]

\[
\text{VHI} = a \times \text{VCI} + (1 – a) \times \text{TCI}, \quad (4)
\]

where NDVI, NDVImax, NDVImin, BT, BTmax, and BTmin are smoothed weekly NDVI (BT) and their 1992–2005 absolute maximum and minimum (climatology), respectively; \(a\) is a coefficient quantifying the share of VCI versus TCI contribution in overall vegetation health. Because this share is not known for specific location it is assumed that the shares are equal, \(a = 0.5\).\textsuperscript{19} All three indices are scaled to range from 0 (severe vegetation stress - low NDVI, hot BT) to 100 (exceptionally favorable conditions - high NDVI, cool BT).\textsuperscript{13,120} The GVI data set processed at 16 km\(^2\) spatial resolution and weekly time resolution was averaged over land pixels in Bandarban, Bangladesh.\textsuperscript{17}
RESULTS

Correlation analysis. We begin by examining the correlation of yearly malaria deviation from trend DY with the VH indices for each week of the year, where DY is the detrended index of malaria incidence intended to capture interannual variability in intensity of malaria outbreaks. Figure 2 shows the dynamics of correlation coefficients for DY versus VCI, TCI, and VHI for the Bandarban district. During cooler months, when mosquitoes are less active, correlation with DY is low. The beginning of the rainy season (Week 16), when mosquito activity starts, the correlation of TCI with annual malaria case index starts increasing, reaching a maximum of +0.56 at the end of June (Weeks 25–26) during first peak malaria season (May–June). After first peak malaria season, correlation is gradually decreasing to low indicating that TCI has low predictive ability. From Week 32 the correlation again increases reaching a secondary maximum of +0.53 at the end of September (Week 35) during second peak malaria season (August–September). After this second maximum correlation decreases to low level by the beginning of the next cool season in November. Correlations of malaria cases with VCI and VHI are lower (|\( r \)| < 0.50).

It is important to emphasize that the correlation of DY with TCI is positive, indicating that a larger number of malaria cases (DY is above the trend) is associated with TCI greater than 50, which indicates cooler weather. A smaller number of malaria cases (DY is below trend) is associated with lower TCI (below 50, hotter weather). The differences in TCI dynamics between years with relatively high and low malaria incidence can be seen by comparing TCI for individual years. For example, 1992 (year with low DY) and 1998 (high DY) TCI time series shown in Figure 3, indicate that the observed positive correlation of malaria cases with TCI indicate that above average malaria cases are associated with higher TCI (cooler thermal condition) and below average malaria cases are associated with lower TCI (hotter) during Weeks 22–27 and 33–37 and in the weeks before and after.

Regression analysis. Given correlation analysis results for regression of weekly TCI individually onto malaria cases, one possible strategy for optimally estimating malaria cases would be multiple regression analysis, where DY would be regressed on the linear combination of weekly TCI values for the two peak malaria seasons, Weeks 22–27 and 33–37, respectively. Following the correlation analysis the results of fitting the ordinary least squares models are approximated by Equations 5 and 6 for two different peak periods and are shown in Table 1.

\[
DY_{w22-27} = b_0 + b_1TCI_{22} + b_2TCI_{23} + b_3TCI_{24} + b_4TCI_{25} + b_5TCI_{26} + b_6TCI_{27},
\]

\[
DY_{w33-37} = b_0 + b_1TCI_{33} + b_2TCI_{34} + b_3TCI_{35} + b_4TCI_{36} + b_5TCI_{37},
\]

A comparison of the relative degree of statistical significance of the model with those of the partial regression coefficients reveals multicollinearity. The overall model for Equation 6 is highly significant with \( F \) values of 6.09 and \( P \) values much smaller than 0.05 but model for Equation 5 is not significant at \( P < 0.05 \). The largest \( P \) value for a partial regression coefficient for both models are not significant at \( P < 0.05 \) level. This type of result is a natural consequence of multicollinearity: the overall model may fit the data quite well, but because several independent variables measure similar phenomena (vegetation

![Figure 2. Correlation coefficient of the annual malaria case index DY vs. weekly Temperature Condition Index (TCI), Vegetation Condition Index (VCI), and Vegetation Health Index (VHI).](image)

![Figure 3. Temperature Condition Index (TCI) dynamics for the years with the lowest and largest annual malaria cases.](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>Pr &gt;</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCI_{22}</td>
<td>0.04359</td>
<td>0.33019</td>
<td>0.13</td>
<td>0.8987</td>
<td></td>
</tr>
<tr>
<td>TCI_{23}</td>
<td>0.27024</td>
<td>0.50766</td>
<td>0.53</td>
<td>0.6110</td>
<td></td>
</tr>
<tr>
<td>TCI_{24}</td>
<td>−1.75771</td>
<td>0.83040</td>
<td>−2.12</td>
<td>0.0721</td>
<td></td>
</tr>
<tr>
<td>TCI_{25}</td>
<td>2.01674</td>
<td>0.83730</td>
<td>2.41</td>
<td>0.0469</td>
<td></td>
</tr>
<tr>
<td>TCI_{26}</td>
<td>1.58026</td>
<td>0.57741</td>
<td>2.65</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>TCI_{27}</td>
<td>−1.67617</td>
<td>0.61073</td>
<td>−2.74</td>
<td>0.0287</td>
<td></td>
</tr>
</tbody>
</table>

Equation 5: \( R^2 = 0.73, \) RMSE = 6.40, \( F = 3.11, P < 0.08 \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>Pr &gt;</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCI_{33}</td>
<td>−1.66775</td>
<td>0.62139</td>
<td>−2.68</td>
<td>0.0278</td>
<td></td>
</tr>
<tr>
<td>TCI_{34}</td>
<td>1.15251</td>
<td>0.82263</td>
<td>1.40</td>
<td>0.1988</td>
<td></td>
</tr>
<tr>
<td>TCI_{35}</td>
<td>1.53209</td>
<td>0.77741</td>
<td>2.65</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>TCI_{36}</td>
<td>−0.6309</td>
<td>0.77083</td>
<td>−0.82</td>
<td>0.4263</td>
<td></td>
</tr>
<tr>
<td>TCI_{37}</td>
<td>−0.19507</td>
<td>0.43840</td>
<td>−0.44</td>
<td>0.6681</td>
<td></td>
</tr>
</tbody>
</table>

Equation 6: \( R^2 = 0.79, \) RMSE = 5.25, \( F = 6.09, P < 0.01 \).
RMSE = root mean-square error.
The first part of Table 2A shows the eigenvalues of TCI for Weeks 22–27. The “Eigenvalue” column shows the first principal component results for first malaria peak (Weeks 22–27). The second part of Table 2A (“Eigenvectors”) shows the eigenvectors for each of the PCs. These coefficients relate the components to the original variables listed on the first column and are scaled so that their sum of squares is unity. The first PC approximates an average of the weekly TCIs, with slightly larger weights to TCI_{24} (0.416) and TCI_{25} (0.418). Note that while the first PC accounts for most of the interannual variation in TCI, this does not necessarily mean that it encapsulates the relationship between TCI and malaria incidence (DY); this has to be determined by regression of DY against the PCs.

The first part of Table 2B shows the eigenvalues of TCI for Weeks 33–37. The “Eigenvalue” column shows the first PC accounts for most of the total variance in interannual TCI over those weeks (4.95), and the others account for much smaller fractions yet of the variance. The first component accounts for 98% of the total variation, a result that is typical when the original variables are highly correlated.

The second part of Table 2B (“Eigenvectors”) shows that the first PC approximates an average of the weekly TCIs, with slightly larger weights to TCI_{36} (0.448) and TCI_{37} (0.448).

Table 3 shows a summary of stepwise selection of PCs for TCI for both peak malaria periods. For Weeks 22–27 TCI (Equation 5) the first and sixth PCs were significant estimators, whereas for Weeks 33–37 (Equation 6) the first, second, and third PCs were significant estimators. In each case, the first PC corresponds approximately to the average value of TCI over Weeks 22–27 and 33–37, respectively, whereas the other PCs describe trends in TCI over the period.

The final set of coefficients for the original weekly TCI variables used in Equations 5 and 6 are calculated and used to generate PC-based estimation models expressed by Equations 7 and 8.
Selection of principal components for prediction based on stepwise regression

<table>
<thead>
<tr>
<th>Model</th>
<th>Principal components</th>
<th>$R^2$</th>
<th>$F$ value</th>
<th>$Pr &gt; F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCI$_{22-27}$ (Equation 5)</td>
<td>Prin1 Prin6</td>
<td>0.57</td>
<td>7.24</td>
<td>0.009</td>
</tr>
<tr>
<td>TCI$_{33-37}$ (Equation 6)</td>
<td>Prin1 Prin2 Prin3</td>
<td>0.74</td>
<td>9.60</td>
<td>0.003</td>
</tr>
</tbody>
</table>

\[
DY_{\text{w22-27}} = 92.76 - 0.08\text{TCI}_{22} - 0.77\text{TCI}_{13} - 1.88\text{TCI}_{34} + 1.12\text{TCI}_{35} + 1.78\text{TCI}_{26} - 1.47\text{TCI}_{27}, \tag{7}
\]

\[
DY_{\text{w33-37}} = 87.79 - 1.22\text{TCI}_{33} + 0.19\text{TCI}_{14} + 1.69\text{TCI}_{35} + 0.17\text{TCI}_{36} - 0.62\text{TCI}_{37}, \tag{8}
\]

VALIDATION OF MODEL

Validation is the step in which the ability of the chosen model to adequately describe the phenomenon under study, in this case interannual variation in malaria incidence, is tested using independent evidence. Because the training data is short, leave-one-out cross-validation\(^{19}\) (the jackknife technique) was used to verify the predictive value of vegetation indices derived from satellite imaging for malaria cases later in the season. A model ($DY = f(\text{TCI})$) was developed with 1 year out and this model was applied to the removed year to estimate the number of malaria cases deviation from trend ($DY$) based on satellite data of the eliminated year. The eliminated year was then returned to the data set and the next year was removed for model development and testing. Each year data were removed one at a time and the candidate model was fit 13 times to the eliminated year. As a result of this procedure, 13 independent estimations were obtained. Each year of data was successively removed and a principal components regression model using TCI from Weeks 33–37 was fit to reduced data set employing the same criteria as those used above for fitting the entire set. Finally, an estimation of malaria cases for the eliminated year was made from the regression equation derived using data from the other years. As a result of this procedure, 14 independent estimations were obtained.

Figure 4A displays observed versus independently estimated time series of percent positive malaria cases, which shows that years of both high and low malaria incidence is generally estimated well, with $R^2$ between estimated and observed percent of malaria cases of 0.78. The root-mean-square estimation error is 2.08 as compared with 3.28 for a naive forecast of the mean incidence of malaria cases for each year, corresponding to a 74% reduction in estimation error variance.

We also performed similar validation for the Weeks 22–27. Figure 4B displays observed versus estimated time series of percent of malaria cases with $R^2$ between estimated and observed percent of malaria cases of 0.65.

DISCUSSION

We tested indices derived from satellite remote sensing characterizing moisture (VCI) and thermal (TCI) conditions derived from satellite remote sensing as estimators for estimation of malaria cases in the Bandarban district, Bangladesh, over 1992–2005. Correlation between the number of malaria cases deviations from trend (DY) with TCI was strong during June and September (main malaria season) and weak during November through May. Beginning in April (warm season starts), the correlation with TCI rapidly increases, and in June (Weeks 22–27) the correlation reaches a maximum, corresponding to the first of the two seasonal peaks in malaria transmission (before the heaviest summer monsoon rains). A second correlation maximum is reached in September (Weeks 33–37), during the second peak in malaria transmission. We found that interannual variation of malaria incidence was more sensitive to thermal (TCI) than to moisture (VCI) conditions or the composite vegetation health (VHI) index. The correlation between TCI and malaria incidence was large enough to enable statistical models to be developed that skillfully estimated interannual variation in malaria incidence based on TCI. Although this correlation is qualitatively consistent with established understanding of the meteorology and mosquito ecology in this region of Bangladesh, comparing the VHI with available surface and satellite data for traditional meteorological variables, such as temperature, precipitation, and cloudiness, should offer further insight into exactly how TCI reflects variables relevant to malaria transmission. Further work will also study the distribution of malaria cases within each year, and whether remotely sensed VHI from early in the year can provide advance warning of several weeks or months of conditions conducive to severe malaria outbreaks. This would enable better deployment of public health resources to prevent and treat malaria.

CONCLUSION

Correlation and regression analysis shows that evaluation of malaria epidemic risk can be estimated with VHI. Similar models based on VHI might be calibrated to estimate malaria and other vector-borne diseases in other regions as well. The Vegetation Health Indices used in this study are available in real time (weekly) at http://orbit.nesdis.noaa.gov/smcd/emcb. Further study might include other high-resolution data such as sea surface temperature and soil moisture from satellite instruments, such as MODIS, AMSR–E, and RADARSAT, and combining satellite sensor data with ground-based weather data.
REFERENCES


