Early prediction of malaria in forest hills of Bangladesh using AVHRR based satellite data

Mohammad Nizamuddin[1], Atiqur Rahman [1], Leonid Roytman[1], Felix Kogan[2], and Al Powell[2]
[1] Department of Electrical Engineering, The City College of the City University of New York, 138 street & convent Avenue, New York, NY-10003 nizam89@gmail.com, fmatiq@yahoo.com , roytman@ccny.cuny.edu
[2] National Oceanic and Atmospheric Administration (NOAA) Camp Spring, MD 20910 Felix.Kogan@noaa.gov

ABSTRACT: The control of epidemic malaria is a priority for the international health community and specific target for the early detection and effective control of epidemics. Malaria has reemerged as a major public health problem in the world during past few years. This study attempts to identify the potential factors for malaria epidemic in forest hills in Bangladesh. It estimates the correlation between various environmental factors that contribute to malaria transmission and shows the application of remote sensing data for improved predictions of malaria epidemics in Bandarban district of Bangladesh which has the highest frequency of malaria cases in the country. An algorithm uses the Vegetation Health (VH) Indices (Vegetation Condition Index (VCI) and Temperature Condition Index (TCI)) computed for each week over a period of 14 years (1992-2005) from Advance Very High Resolution Radiometer (AVHRR) data flown on NOAA afternoon polar orbiting satellite. The weekly VH indices were correlated with the epidemiological data. A good correlation was found between malaria cases and TCI characterizing thermal condition during the month of August and September. Following the results of correlation analysis the principal components regression (PCR) method was used to construct a model to predict malaria as a function of the TCI computed for this period. The simulated results were compared with observed malaria statistics showing that the error of the estimates of malaria is less than 10%. Remote sensing therefore is a valuable tool for estimating malaria well in advance thus preventive measures can be taken.

INTRODUCTION: Malaria is endemic to over 100 countries around the world and is responsible for over 300 to 500 million clinical cases and more than a million deaths each year (Faiz MA 2001, Nagpal B 1995) early approximately 40% of the worlds population mostly living in the poorest countries is at risk of malaria each year. Bangladesh is located in the north-eastern part of the South Asian sub- continent. It’s bordered by India on the west and the Myanmar on the south east. It has about 4180 km border with India in its western northern and eastern territory and a 190 km border with Myanmar (Burma) in its south-eastern side. Out of 32 bordering districts only 13 districts have malaria problem and some areas also belong to the epidemic prone areas. The forest and forest fringe areas particularly in the north eastern and south eastern border report more than 90% of total positive cases and more than 95% of total P.palciparum cases in the country (Parsel Alam 2008) Bangladesh has sub-tropical warm, wet and humid climate with 140 million population(Pampana E, 1969 Rosenberg R and Maheswary N, 1982) and a land area of about 147,570 sq .Km. Malaria parasite in Bangladesh is transmitted by female Anopheles mosquitoes(Elias M. and Rahman M, 1987, Rosenberg R and Maheswary N, 1982). His number increases considerably during warm and wet season (June – October). In the 90's the upsurge of diseases has created an alarming situation with reports of local P.falsiparum outbreaks in northeastern border areas and increase of P. falciparum and P.vivax cases in the endemic areas.

Remote sensing is playing an increasing role in understanding the natural environment and its inherent, physical and chemical processes. AVHRR based vegetation health indices were found to be very useful for early detection and for monitoring weather impacts on malaria. Therefore, this paper investigates the application of AVHRR based vegetable health indices as proxies for the characterization of weather conditions and their impacts on malaria.

METHODOLOGY: The study area is Bandarban district Bangladesh, a hilly area mostly covered with tropical evergreen forest that provide the considerable shade needed to breeding grounds, bounded between 22.5^0N - 22^0N latitude and 91.5^0E and 92.5^0 E longitude. The AVHRR–measured satellite data for solar energy reflected/emitted from the land surface (represented in 8–bit counts) were collected from NOAA Global Vegetation Index (GVI) data set from 1991 through 2005. GVI data set was developed by sampling 4 square km Global Area Coverage (GAC) data to 16 square km spatial resolution and daily observations to seven–day composite (Kidwell 1997). The GVI digital counts in the visible (VIS, 0.58–0.68µm, Ch1), near infrared (NIR, 0.72–1.00 µm, Ch2) and infrared (IR, 10.3–11.3 µm, Ch4) spectral regions were used in this study. Post–launch–calibrated VIS and NIR counts were converted to reflectance and used to calculate the Normalized Difference Vegetation Index (NDVI = (NIR – VIS)/ (NIR + VIS)). The Ch4 counts were converted to brightness (radiative) temperature (BT) (Kidwell 1997).

The study objective is to investigate the strength of the relationship between VH indices and malaria and determine if the strongest correlation coincides with malaria critical period of development, which is highly sensitive to weather conditions.

Following Brockwell and Davis (2000) malaria time series (Figure 1) was approximated by the linear Equation 1. 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 2)

$$Y_{trend} = 1059.95 - 0.506 * Year \quad (1)$$

$$DY = (Y / Y_{trend})*100 \quad (2)$$

where Y is actual percent of malaria cases, Y_{trend} is the malaria estimated from trend ; DY is deviation (%) from the trend. DY is a good relative metric of malaria cases. In 1998, DY was only 119% or 19% above the trend, whereas in 1992, DY was 87% or 13% below the trend. These estimations indicate that the 1992 (less % of malaria) was an unfavorable year for malaria whereas 1998 (higher % of malaria) was favorable (Rahman et al.; 2006).

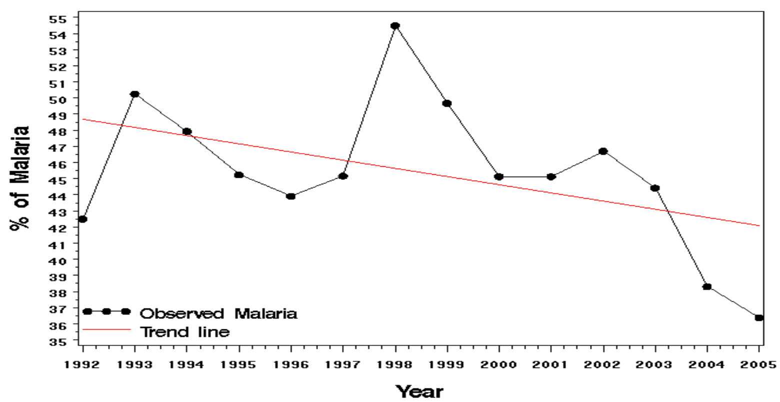

FIGURE1. Percent of malaria in Bandarban district, Bangladesh and trend line

The VH indices were calculated from NDVI and BT using methodology of Kogan (1997). The derivation steps include (a) complete elimination of high frequency noise from NDVI and BT annual time series, (b) approximation of annual cycle, (c) calculation of multi–year climatology and (d) estimation of medium–to–low frequency fluctuations during the seasonal cycle (departure from climatology) associated with weather variations. The Vegetation Condition Index (VCI) characterizing moisture and Temperature Condition Index (TCI) characterizing thermal conditions were calculated as:

$$VCI = 100(NDVI - NDVImin)/ (NDVImax - NDVImin) \quad (3)$$

$$TCI = 100(BTmax - BT)/(BTmax - BTmin) \quad (4)$$

where NDVI, NDVImax, NDVImin, BT, BTmax and BTmin are the smoothed weekly NDVI (BT) and their 1992–2005 absolute maximum and minimum (climatology), respectively. The range of VH indices is from 0 indicating severe vegetation stress to 100 indicating favorable conditions (Kogan, 2002). Average weekly values of VH were calculated for the area of Bandarban in Bangladesh.

RESULTS AND DISCUSSION: Since DY and VH indices were similarly expressed as a deviation from climatology (from trend for malaria and from max–min for VH), further investigation included correlation and regression analysis of these deviations to investigate the association between them for Bandarban district Figure 2 shows dynamics of correlation coefficients for DY versus VCI and TCI for Bangladesh.

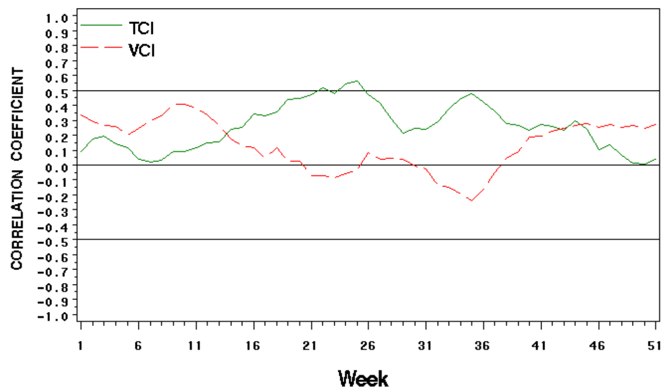

FIGURE 2. Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI

It is seen, during cooler month (spring and fall) when mesquites are less active; correlation is low for both indices. In spring (week 16) when mosquito activity season starts correlation start increases and reaching maximum (week 25 for TCI) at the end of June (Week 25-26), In week 32 correlation again increases reaching maximum at (.5 for TCI) at the end of September (week 35). This result is compatible with mosquito's activity. The result of fitting the ordinary least squares (OLS) model is approximated by equation 5 (Salazar et al.; 2008).

$$DY=b_0 + b_1TCI_{32}+b_2TCI_{33}+b_3TCI_{34}+b_4TCI_{35}+ b_5TCI_{36} \quad (5)$$

Table 1 show that the value of R^2 is large 0.88. A comparison of the relative degree of statistical significance of the model with those of the partial regression coefficients reveals a great degree of multicollinearity. The p value for the regression coefficient is high, which is not significant at $p<0.05$ level and very small tolerance and high variance inflation. The overall model may potentially fit the data well, but because several independent variables are measuring similar phenomena, it is difficult to determine which of the individual variables contribute significantly to the regression relationship. To resolve this problem, an alternative method of principal components regression (PCR) is used, which results in estimation and prediction better than OLS.

Table1. Results of multiple linear regressions (OLS) of DY on the equation (5)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	86.226	3.11859	27.65	<.0001	.	0
TCI32	1	-1.1168	0.59254	-1.88	0.0962	0.01203	83.12697
TCI33	1	-0.00702	1.19173	-0.01	0.9954	0.00307	326.03281
TCI34	1	0.22426	1.27245	0.18	0.8645	0.00254	393.64865
TCI35	1	2.33509	0.81987	2.85	0.0215	0.00617	162.00663
TCI36	1	-1.26114	0.61886	-2.04	0.0759	0.01051	95.18932

R^2=0.88, RMSE= 4.8

Using PCR methodology, the variables in model equation (5) were transformed into new orthogonal or uncorrelated variables called principal components (PCs) of the correlation matrix. The stepwise multiple regressions are eliminating some of the PCs to get a reduction in variance. Once the regression coefficients for the reduced set of orthogonal variables are calculated, they are transformed into a new set of coefficients that correspond to the original or initial correlated set of variables in model equation (5). These new coefficients are called principal component estimators (Salazar et al. 2007). Table 2 shows that model with $p<0.05$ for regression coefficient is significant. It has already been argued that the OLS estimates are unsatisfactory when multicollinearity is present. Hence, following PCR analysis the final set of coefficients for variables in model equation (5) are calculated and used to generate estimation model equation (6).

$$DY= 86.48- 0.98\ TCI_{32}-0.36\ TCI_{33}+0.61\ TCI_{34}+ 2.20TCI_{35}-1.31TCI_{36} \qquad (6)$$

Table2. Principal component results for Bandarban forest hill Bangladesh (R^2=0.84)

Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	99.9894	1.23302	81.09	<.0001
Prin1	1	1.64897	0.57579	2.86	0.0187
Prin2	1	-28.26484	6.11437	-4.62	0.0012
Prin3	1	-28.75508	11.9657	-2.4	0.0397
Prin4	1	43.6613	18.32133	2.38	0.041

Validation is the step in which the prediction with the chosen model is tested independently. At the beginning of the model development stage, the data is divided into two sets, the training and validation data sets. Figure 3 displays observed versus independently simulated percent of malaria time series, which shows excellent match for both time series, with RMSE value of 1.78 and R^2 value of 0.82 for simulated and observed time series.

CONCLUSION: Bandarban district is considered as one of the malaria endemic areas in Bangladesh. The two AVHRR-based VH indices characterizing moisture (VCI) and thermal (TCI) conditions were tested as predictors of malaria cases. It was found that malaria cases were more sensitive to thermal conditions. Correlation analysis between malaria cases deviations from trend (DY) with TCI during the period 1992 to 2005, showed correlation during the mosquito activities (weeks around 36, September-October). Therefore, this index was used for the statistical modeling of malaria cases. AVHRR data from NOAA polar orbiting satellites can provide valuable information about malaria cases to predict vector distribution in Bandarban district in Bangladesh and help the concerned authority to take preventive measures accordingly.

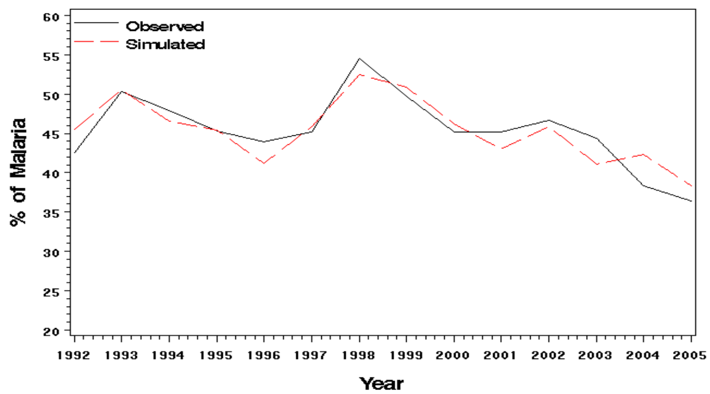

FIGURE 3. Simulated and observed percent of malaria for Bandarban District, Bangladesh

Acknowledgement

This study is supported by National Environmental Satellite Data and Information Service (NESDIS) award no. NA07NES4280009.

Disclaimer

The manuscript contents are solely the opinions of the authors and don't constitute a statement of policy, decision, or position on behalf of NOAA or U.S government

REFERENCES:

BROCKWELL, P. J. & DAVIS, R. A., (2000), **Introduction to time series and forecasting**, New York, Springer.

ELIAS M. AND RAHMAN M., (1987), **the ecology of malaria carrying mosquito Anopheles Philippinensis Ludlow and it's relation to malaria in Bangladesh**. *Medical Research Council Bulletin, Bangladesh 13*: 15-28

FAIZ MA., YUNUS EB, RAHMAN MR, HOSSAIN MA, PANG LW, RAHMAN ME AND BHUIYA S.N, 2001. **Failure of national guidelines to diagnose uncomplicated malaria in Bangladesh**. *Am J Trop Med Hyg 67***:** 2002,396 -399

GUNST, R.F. and MASON, R.L., 1980, Regression Analysis and its Application: a Data oriented Approach (New York: m. Dekker).

KIDWEL K. B., (1997), **Global Vegetation Index user's guide**, Camp Springs, Md., U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite Data and Information Service, National Climatic Data Center, Satellite Data Services Division.

KOGAN, F. (1997), **Global droughts watch from space**. Bulletin American Meteorological Society, 78**,** 621-636.

KOGAN, F. (2002), **World droughts in the new millennium from AVHRR-based Vegetation Health Indices**

KOGAN, F., BANGJIE, Y., GUO, W., PEI, Z. & JIAO, X., (2005), **Modeling corn production in China using AVHRR-based vegetation health indices.** International Journal of Remote Sensing, 26, 2325-2336.

NAGPAL B AND SHARMA (1995) I**NDIAN ANOPHELINES, NEW DELHI**, 416-423

ROSENBERG R AND MAHESWARY N, 1982, **Forest Malaria in Bangladesh, I.Parasitology,** *Am J Trop Med Hyg, 31***:** 175 –182

ROSENBERG R AND MAHESWARY N, 1982, **Forest Malaria in Bangladesh, II.Transmission,** *Am J Trop Med Hyg 31***:** 183–191

PARSEL ALAM (2008), **Malaria country report**, Bangladesh

Rahman A, Kogan F, Roytman L (2006) **Analysis of malaria cases in Bangladesh with remote sensing data**. Am J Trop Med Hyg 74 (1): 17–19

ROSENBERG R., (1982), **Forest Malaria in Bangladesh, III .Breeding Habitats of Anopheles Dirus**. *Am J Trop Med Hyg*: 31**,**192- 201

Salazar L, Kogan F, Roytman L (2008). **Using vegetation health indices and partial least squares method for estimation of corn yield**. International Journal of Remote Sensing, Volume 29, Issue 1, 2008 Pages 175 - 189

Salazar Kogan F, Roytman L (2007)**. Use of remote sensing data for estimation of winter wheat yield in the United States**. International Journal of Remote Sensing Vol.28, Nos.17-18, 2007, 3795-3811