Optical Remote Sensing a Potential Tool for Forecasting Malaria in Orissa, India

Leonid Roytman[1], Mohammad Nizamuddin[1], Felix Kogan[2], and Mitch Goldberg[2]

[1]NOAA-CREST, City College, New York
[2]NOAA/NESDIS, Camp Springs, MD, USA

Abstract

Improved forecasting prevention and control of epidemics is a key technical element for malaria eradication program. The objective is to use NOAA\AVHRR environmental satellite data to produce weather seasonal forecasts as a proxy for predicting malaria epidemics in Orissa state, India which has one of the highest endemic of malaria cases in the country. We report an algorithm that uses Vegetation Health (VH) Indices (Vegetation Condition Index (VCI) and Temperature Condition Index (TCI)) computed from Advance Very High Resolution Radiometer (AVHRR) data flown on NOAA afternoon polar orbiting satellite. A significant relationship between satellite data and annual malaria incidences is found at least three months before the major malaria transmission period. Principal component regression (PCR) method was used to develop a model to predict malaria as a function of the TCI. The simulated results were compared with observed malaria statistics showing that the error of estimations of malaria is small. Optical remote sensing therefore is a valuable tool for estimating malaria well in advance thus preventive measures can be taken.

1. Introduction

A better understanding of the relationship between malaria epidemics, satellite data and the climatic anomalies could help mitigate the world-wide increase in incidence of the mosquito-transmitted diseases. Malaria is the most deadly, parasitic, human infection, accounting for an each year [1]. Human malaria also imposes drastic economic production losses [2]. Epidemic malaria causes almost one fifths of estimated annual worldwide deaths which include up to 50% of the estimated annual malaria mortality in persons less than 15 years of age [3]. This study attempts to identify the potential factors for malaria epidemic. 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 incidence in Orissa state India, a malaria endemic region in the country.

Malaria has become increasingly serious problem during the past 10 –15 years, and it is currently the number one public health problem in Orissa. The magnitude of the problem can be viewed in terms of 1998, when Orissa (with a share of less than 4 % of all-India population), accounted for 28.6 per cent of the detected cases of Malaria in all of India (two million), and 62.8 per cent of all Malarial deaths in the country, according to the National Malaria Eradication Programme Report for 1998 (Government of India 1999). Kandhamal, Keonjhar, Sundargarh, and Mayurbhanj districts of Orissa account for 50% of all Malarial instances and 40% of the Malarial deaths in the state (Orissa Voluntary Health Association 1995, p.25). Orissa and other forested areas in India primarily inhabited by ethnic tribes have experienced the highest incidence of mortality attributable to malaria [3].

Moisture and thermal condition are widely considered to be a major driver for inter annual variability of malaria incidence. There are casual relationship between rainfall and malaria (creation of breading sites favors malaria proliferation) and between humidity and malaria (increasing humidity favors vector survival). However the relationship is often nonlinear as because excessive rainfall may wash out breeding sites which even resulting in less malaria than expected [4, 5]. This algorithm uses the Vegetation Health (VH) Indices which are Vegetation Condition Index (VCI) characterizing moisture condition and Temperature Condition Index.

(TCI)) characterizing thermal condition, computed for each week over a period of 10 years (1997-2006) from Advance Very High Resolution Radiometer (AVHRR) data flown on NOAA afternoon polar orbiting satellite. The weekly indices were correlated with the epidemiological data of malaria.

Analysis of time series of malaria incidence for ten years and satellite data (Normalized Difference Vegetation Index and Brightness Temperature) for 23 years has been conducted. The relationships of variability in TCI and VCI to malaria incidence are assessed at a state level after removing the impact of non climatic and policy intervention. A significant relation was found between malaria cases and TCI during the month of March and April. 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%.

2. Study Area and Data set

There are three major seasons in Orissa- very hot summer (March-June), rainy season (July-September) and the winter (October-February). It is warm almost throughout the year in the Western districts with maximum temperature hovering between 40-46°C and in winter temperature goes down to 8°C . The climate is equable but highly humid and sticky. The average rainfall is 1500 mm, experienced as the result of south west monsoon during July-September. The month of July is the wettest and the major rivers often get flooded. The state also experiences small rainfall from the retreating monsoon in the months of October-November. January and February are dry.

An empirical studies have demonstrated existence of unpredictability in malaria endemicity in forested and broken forested villages [6]. In the beginning of August, An. fluviatilis becomes thickly dense and reaches to the peak in December. The intensity of malaria declines in January and reaches to its lowest level between May and July. In contrast to the behavior of An. fluviatilis, the concentration of An. culicifacies is high between March and September with its peak falling between March and April. The incidence of An. culicifacies is generally low during November to February. Various relevant studies have shown that these vectors of malaria pulsate alternatively and maintain recurrent spread in forested villages.

Malaria statistics and satellite data have been used in this research. The first data set Malaria statistics were represented by annual total clinical malaria cases for 1997-2006. The data were collected from the Ministry of Health India. These data provided the number of malaria cases from all patients with fever, who came to the hospitals from effected area. The data were aggregated to local administrative unit health centers and to district level. These data included the number of persons tested and the number of positive malaria cases. In this study, the numbers of malaria cases were calculated as slide positive rate (% of malaria) expressed as malaria positive rate in percent of number of people tested.

We have used the weekly NDVI and BT data set from the Advanced Very High Resolution Radiometers (AVHRR), provided by the NOAA Global Vegetation Index (GVI) from 1996 through 2005. For reduction of cloud effect the GVI data set was developed by sampling 4 km² Global Area Coverage (GAC) data to 16 km² spatial resolutions and from daily observations to seven-day composite data [7]. These satellite data were collected from most populated regions which have the largest percentage of malaria cases (Figure 1). Kogan has promoted the use of additional indices in order to describe more comprehensively the vegetation activity over a given area. The TCI is based on the thermal band (channel 4) of the AVHRR and is used to assess temperature-related vegetation stress and The VCI is based on visible (channel 1) and near infra red (channel 2) of the AVHRR quantifies the vegetation greenness component is defined as

$$VCI = 100*(NDVI - NDVI\ min)/ (NDVI\ max - NDVI\ min) \quad (1)$$

$$TCI = 100*(BT\ max - BT)/ (BT\ max - BT\ min) \quad (2)$$

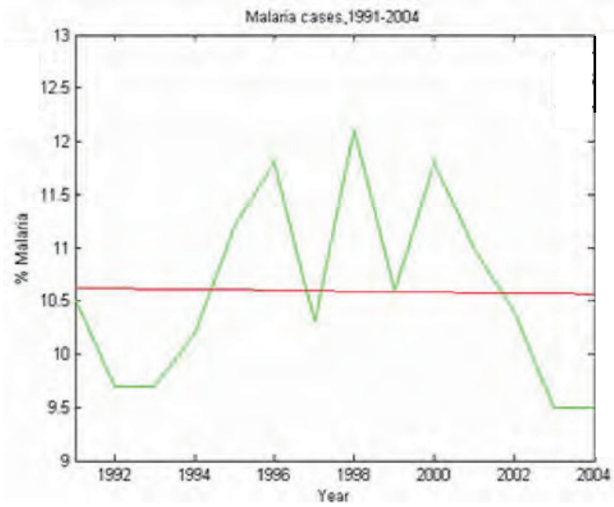
Where NDVI, NDVI max, and NDVI min (BT, BT max, and BT min) are smoothed weekly NDVI (BT), their multi year absolute maximum and minimum respectively. VCI and TCI algorithm have been developed from NDVI which separates weather component and ecosystem component by Max-Min criteria [8, 9, 10]. VCI could assume values between 0 and 100, corresponding to variations from stressed to favorable and TCI could assume values between 100 and 0, corresponding to variations from stressed to favorable vegetation conditions. Finally, VHI is defined as a combination between the two previous indices:

$$VHI = a^* VCI + (1 - a)^* TCI\ (a = .5) \quad (3)$$

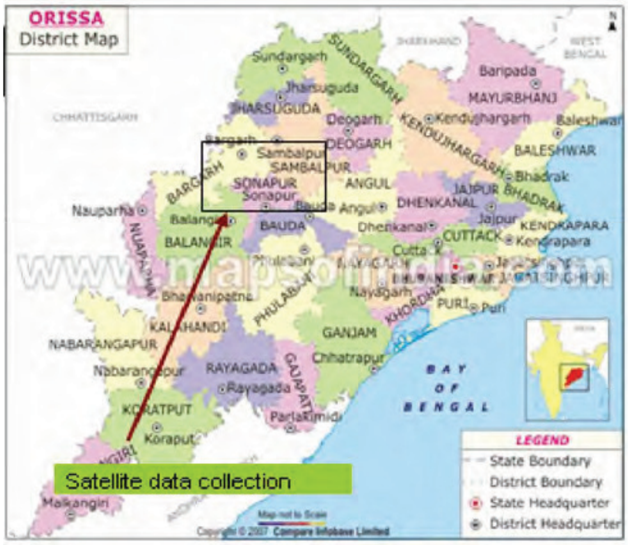
The vegetation health indices were designed with the idea of extracting the weather component from NDVI and BT value [11]. The long-term changes in vegetation driven by climate, soil, vegetation type, topography etc. and short term weather fluctuation such as moisture and thermal condition are represented by NDVI and BT.

3. Methodology

Trend Analysis - The malaria time series for Orissa (Figure 1.(A) was approximated [12] by the linear equation (4). The weather-related variations around the trend are expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 5).



(A)



(B)

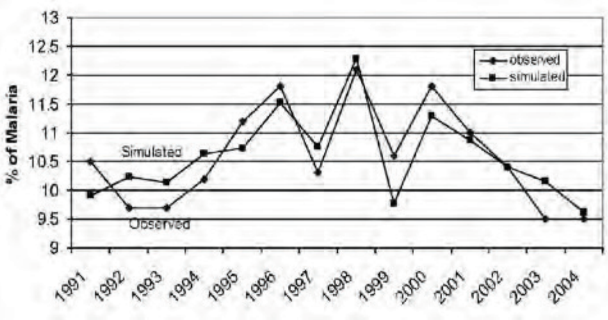
Fig. 1. (A) Percent of malaria in Orissa state, India and trend line (1991-2004); (B) District Map of Orissa, the rectangular box where Satellite data collection has been performed.

$$Y_{trend} = 19 - 0.004 * \text{Year} \quad (4)$$

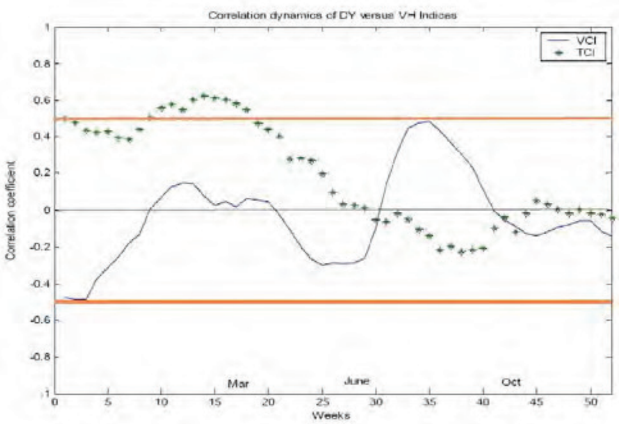
$$DY = (Y / Y_{trend})*100 \quad (5)$$

4. Correlation Dynamics

In 1998, DY was 114 % or 14% above the trend, whereas in 2004, DY was 90% or 10% below the trend. These estimations indicate that the 2004 (less % of malaria) was an unfavorable year for malaria whereas 1998 (higher % of malaria) was favorable.



(A)



(B)

Fig. 2: (A) Detrended value (DY) and Temperature Condition Index (TCI week 13), moisture condition index (VCI) from year 1991-2004, (B) Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI for Orissa state.

Figure 2 (A) shows detrended value (DY) and Temperature Condition Index (TCI) for moisture condition index (VCI) from years 1991-2004, which has clear indication that TCI (week 12) is positively correlated to DY. Investigation included correlation analysis of trend malaria cases (DY) versus VCI and TCI, shown in Figure 2 (B). Since Orissa has monsoon, humid, subtropical warm summer, during cooler months (November – December) when mosquitoes are less active; correlation is low for both indices. In spring (pre monsoon, March and April) when mosquito activity season starts, correlation (positive) start increasing to reach maximum (0.6 for TCI and 0.2 for VCI) and high correlation persist until at the end of March (week 10-15). These phenomena can be explained as follows, if the temperature during this time reduces, malaria activities increase. In other words if more rainfall (more cloud cover and less temperature) occurs during this time malaria activities will increase. Higher amount of precipitation causes more breeding places for mosquito and malaria transmission increases.

5. Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 10 to 15 and to VCI for week 3 to 6 therefore in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 10 to 15). Table 1 presents the results of fitting the ordinary least squares (OLS) regression model approximated by equation (6) for Orissa

$$DY = b_0 + b_1\,VCI_3 + b_2\,VCI_4 + b_3\,VCI_5 + b_4\,VCI_6 + b_5\,TCI_{10} + b_6\,TCI_{11} + b_7\,TCI_{12} + b_8\,TCI1_3 + b_9\,TCI_{14} + b_{10}\,TCI_{15} \quad (6)$$

Table 1. Results of multiple linear regression (OLS) of DY on the equation (6) Orissa state, $R^2 = 0.90$

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	99.83	16.84	5.93	0.01	.	0.00
VCI_3	0.92	0.54	1.70	0.19	0.02	45.57
VCI_4	-3.16	1.95	-1.62	0.20	0.00	545.50
VCI_5	1.81	3.36	0.54	0.63	0.00	1372.3
VCI_6	0.44	1.75	0.25	0.82	0.00	339.16
TCI_{10}	-0.09	0.69	-0.13	0.90	0.00	245.20
TCI_{11}	0.87	1.26	0.69	0.54	0.00	805.41
TCI_{12}	-1.92	0.71	-2.71	0.07	0.00	274.15
TCI_{13}	2.50	1.10	2.27	0.11	0.00	768.53
TCI_{14}	-2.13	1.29	-1.65	0.20	0.00	830.12
TCI_{15}	0.82	0.56	1.48	0.23	0.01	124.03

Table 1 shows coefficients of equation (6) and that the value of R^2 is large 0.90. However p-values for the regression coefficients are high, which are not significant at $p < 0.05$ level and very small tolerance and high variance inflation. R^2 for the equation (6) is 0.90 with largest p-value of 0.9 which clearly indicate collinearity among the predictor variables. In addition, VH indices of neighboring weeks are highly correlated as seen in Table 2.

Table 2. Correlation Matrix among weekly TCI and VCI, Orissa, India

	TCI_{15}	TCI_{17}	TCI_{18}	TCI_{19}	TCI_{20}
TCI_{15}	1	0.8487	0.7537	0.5674	0.3756
TCI_{16}	0.9501	0.9488	0.8697	0.7025	0.5477
TCI_{17}	0.8487	1	0.9501	0.8238	0.6921
TCI_{18}	0.7537	0.9501	1	0.9525	0.8598
TCI_{19}	0.5674	0.8238	0.9525	1	0.9627
TCI_{20}	0.3756	0.6921	0.8598	0.9627	1

In order to avoid this multicollinearity problem, we used an alternative method of estimation, principal component regression, [13, 14] (PCR), which results in better estimation and prediction than OLS.

6. Principal Component Regression

Using PCR methodology, principal component estimators are shown in Table (3). We found simulated percent of malaria time series using predictor variables VCI_3 - VCI_3, TCI_{10} through TCI_{15} for four principal components (Prin1, Prin4, Prin6, Prin8). Figure 3 shows observed and simulated malaria in the state of Orissa using PCR model.

$$DY = 97.69 + 0.18\ VCI_3 - 1.56\ VCI_4 + 1.12\ VCI_5 - 0.005\ VCI_6 - 0.78\ TCI_{10} + 1.85\ TCI_{11} + 0.3\ TCI_{12} + 0.03\ TCI_{13} + 0.30\ TCI_{14} + 0.03\ TCI_{15} \quad (7)$$

Table 3: Results of principal component regression for equation (7) Orissa $R^2 = 0.70$, RMSE = 0.49.

Variable	Parameter Estimate	t Value	Pr > |t|
Intercept	100.00	69.34	<.0001
Prin1	1.98	3.42	0.01
Prin4	-10.24	-1.66	0.13
Prin6	24.96	1.65	0.13
Prin8	-84.76	-2.09	0.07

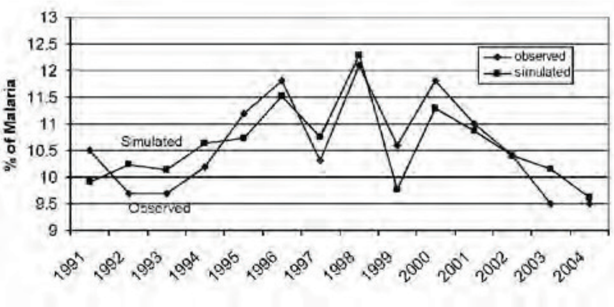


Fig. 3: Independently tested simulated and observed percent of malaria Orissa, India ($R^2 = .70$).

7. Conclusions

The correlation dynamics of malaria and TCI and VCI demonstrates seasonal malaria transmission in Orissa. Malaria is sensitive to thermal condition as well as moisture condition in this region which can be measured using AVHRR based satellite data. The correlation between malaria and TCI is positive and high during week 10- 18 which signifies temperature needs to decrease for malaria proliferation. The possible explanation is that the average temperature is very high in Orissa during April and May which is not good for malaria transmission. So decrease of temperature in this area has a positive affect on malaria transmission. At the same time more rainfall during July, August shall prepare more breeding places which means that high VCI (better moisture condition, more rainfall) has positive relation with malaria transmission. This can be seen from correlation dynamics for VCI during month of July. This model can estimate malaria incidence in this area two months earlier than the main transmission season (post monsoon). Thus AVHRR-based vegetation health indices (VCI and TCI) can be used as a proxy for numerical estimation of malaria cases in Orissa state.

8. References

[1] Thomson, M. C., Doblas-Reyes, F. J., Mason S. J., Hagedorn, R. S., Connor, J., Phindela, T., Morse, A. P., and Palmer, T. N., 2006, Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature, 439, 576-579.

[2] Sachs, J. and Malaney, P., 2002, The economic and social burden of malaria, Nature, 415, 680-685.

[3] Kumar, A.,Valecha, N., Jain, T., and Dash, A. P., 2007, Burden of Malaria in India: Retrospective and Prospective View; American Journal of Tropical Medicine and Hygiene, 77, 69 -78.

[4] Lindsay, S. W., Bodker, R., Malima, R., Msangeni, H. A., Kisinza, W., 2000, Effect of 1997-1998 El Nino on highland malaria in Tanjania, Lancet 355: 989-990.

[5] Najera, J. A., Kouznetzsov, R. L., Delacollette, C., 1998. Malaria Epidemic: Detection and control, Forecasting and Prevention. Geneva: World Health Organization, WHO/MAL/98.1084

[6] Nanda, N., Yadav, R. S., Subbarao, Sarala K., Joshi, Hema and Sharma, V. P., 2000, Studies on Anopheles fluviatilis and Anopheles culicifacies in relation with malaria in forest and deforested riverine ecosystems in northern Orissa, India. J. Am. Mosq Control Assoc. 16(3), 199.

[7] Kidwell, K. B., 1997, Global Vegetation Index user's guide (Camp Springs MD: US Department of Commerce, NOAA, National Environmental Satellite Data and Information Service, National Climatic Data Center, Satellite Data Services Division).

[8] Kogan, F. N., 1995, Droughts of the late 1980s in the United State as derived from NOAA polar orbiting satellite data, Bulletin of the American Meteorological Society, 76: 655-668.

[9] Kogan, F. N., 1997, Global drought watch from space, Bulletin of the American Meteorological Society, 78, 621-636.

[10] Kogan, F. N., 2000, Global drought detection and impact assessment from space. In Drought: A Global Assessment (D. A. Wilhite, Ed.) Vol. I, Hazard and Disaster Series, Routledge, London and New York, 196-210.

[11] Kogan, F. N., 2001, Operational space technology for global vegetation assessment, Bull Am. Meteorol. Soc. 82: 1949-1964.

[12] Brockwell, P. J. and Davis, R. A., 2000, Introduction to time series and forecasting (New York: Springer).

[13] Gunst, R. F. and Mason, R. L., 1980, Regression Analysis and its Application: a Data oriented Approach (New York: M. Dekker).

[14] Ryan, T. P., 1997, Modern Regression Method, John Wiley & Sons, Inc.