Spatial Statistical Analysis of Japanese Encephalitis Occurrence and Identification of Disease Hot Spots - Case Studies in a JE Endemic District of North East India

Handique, B. K.,[1] Khan, S. A.,[2] Chakraborty, K.,[1] Goswami, J.,[1] and Sarma, K. K.,[1]
[1]North Eastern Space Applications Centre, Dept. of Space, Umiam 793103, Meghalaya, India
E-mail: bkhandique@rediffmail.com
[2]Regional Medical Research Centre, NER (ICMR), P.O Box: 105, Dibrugarh 786001 Assam, India
E-mail: sirajkhanicmr@gmail.com

Abstract
An application of spatial statistics analysis has been demonstrated for identifying Japanese Encephalitis (JE) incidence hotspots to prioritise interventions in a JE endemic district of Assam. Spatial order and association of JE reporting villages in the study area have been analysed using spatial statistics analytical techniques. Data on historical morbidity pattern of JE collected at village level provided the required stratification base for delineating the JE incidence hot spots. Spatial statistics parameters such as mean centre, standard deviational ellipse and spatial autocorrelation have been calculated for JE reporting villages. Strong spatial autocorrelation ($p<0.01$) among the JE reporting villages have been observed in terms of morbidity pattern as indicated by Moran's I index. General G statistics has been calculated to categorize JE prone villages and tested for statistical significance. Based on this G statistics, JE hot spots in the study district could be identified for taking timely intervention measures by the health authorities.

1. Introduction

Japanese Encephalitis (JE) is a dreaded vector (mosquito) borne viral disease mostly prevalent in Asian countries including India. Since the first record of JE case in India in 1955 in Tamil Nadu followed by isolation of JE virus from wild caught mosquitoes in 1956, in the last couple of decades, epidemics of JE have occurred in the states of West Bengal, Assam, Manipur, Nagaland, Uttar Pradesh, Bihar and Goa in addition to South India (Khan et al., 1996). JE cases have attained alarming proportions to pose as a major public health problem in India, more so due to unavailability of any cure for the disease and due to its very high case versus fatality ratio (Banerjee, 1996 and Kabilan et al., 2004). Recent advances in Geographic Information System (GIS) along with satellite remote sensing have added new dimensions to spatial statistics analysis in epidemiological studies (Back et al., 1994 and Hay et al., 1997). Many studies have applied these advanced tools in understanding the host-vector relationship and their spatial distribution (Barnes and Cibula, 1979, Glass et al., 1995, Hendrickxy et al., 1999, Dhiman, 2000, Abelardo et al., 2000, Jeganathan et al., 2001, Nageswara Rao et al., 2004 and Handique et al., 2005).

Spatial statistics analytical techniques help in analyzing the spatial order and association of a variable under study. In areas like ecology, epidemiology, geology and image processing, it is often not appropriate to randomize, block and replicate the data because of the spatial associations of attribute features associated with the study variable (Lawson, 2001). On the other hand, it is required to stratify and prioritise areas under a particular administrative unit for better planning and managing resources. Hence a sound technique has to be followed to prioritise these areas of importance or hot spots with sound statistical base. Use of predictive approaches demonstrated by different workers for study of mosquito vector borne diseases like malaria (Srivastava et al., 2001and Abeku et al., 2002). Occurrence status of spatial delimitation, forecasting and control of JE in India has been discussed in detail by Sabesan et al., (2008). In this study, we have employed different spatial statistics analytical tools in GIS domain to study the spatial distribution of JE cases and identify the disease hot spots at village level. This will help the District health authorities to mobilise man and materials for timely interventions.

2. Materials and Method

2.1 Study Area

The study was carried out in Dibrugarh district of Assam state located in the north eastern part of India considering the severity of impact of JE and its perennial occurrence. The district covering a geographical area of 7023.9 sq km lies between 27° 15'N - 28° N Latitude and 94° 45' E - 96° E Longitude (Figure 1). The district is divided into six blocks under six Primary Health Centres (PHCs) viz. Barbaruah, Lahowal, Panitola, Tengakhat, Khowang and Naharani for monitoring and providing health care services in the district.

2.2 Collection of JE Case Data

Data pertaining to the JE cases during the period 1985-2010 were collected from the office of the Joint Director of Health Services in Dibrugarh district and from the office of the Directorate of Health Services, Guwahati, Assam. Average annual case load in Assam during the last 26 years since 1985 has been 298, the average annual incidence per million population being 12.5. Dibrugarh district alone shared a burden of 37 cases per million population. Average disease incidence per unit population was calculated based on the PHC wise census records collected from the respective PHCs compiled under National Vector Borne Disease Control Programme (NVBDCP). Data on socio-economic status viz., poultry, cattle and pig farming practice; flood proneness of the areas and personal protection measures adopted for prevention of mosquito bites, etc. were collected through a socio-economic survey. But the analysis of these socio-economic data is beyond the scope of the present study. The flow of information from the periphery to the JE incidence recording units mentioned above is shown in the following flow chart (Figure 2).

2.3 Spatial Statistics Analysis

Different analyses in GIS domain have been performed using ArcGIS 9.3 to study the pattern of spatial distribution of JE cases in the district. Following spatial statistics parameters have been calculated for the JE reporting villages.

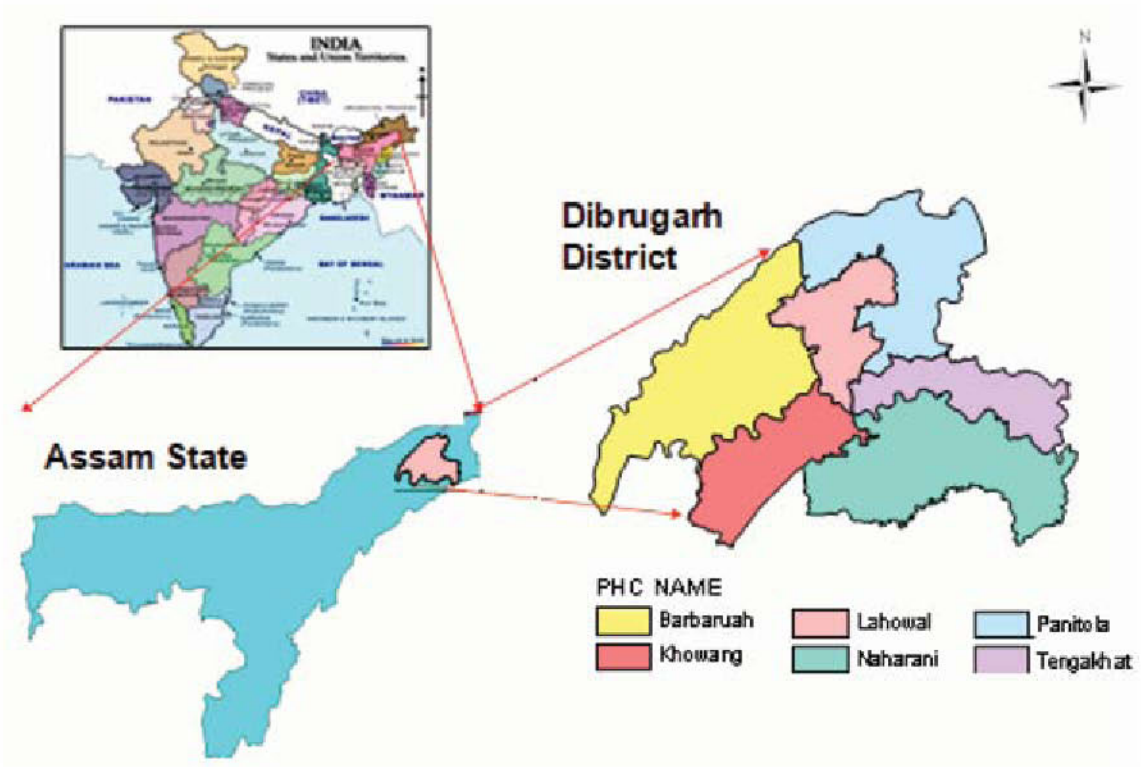

Figure 1: Location map of the study area

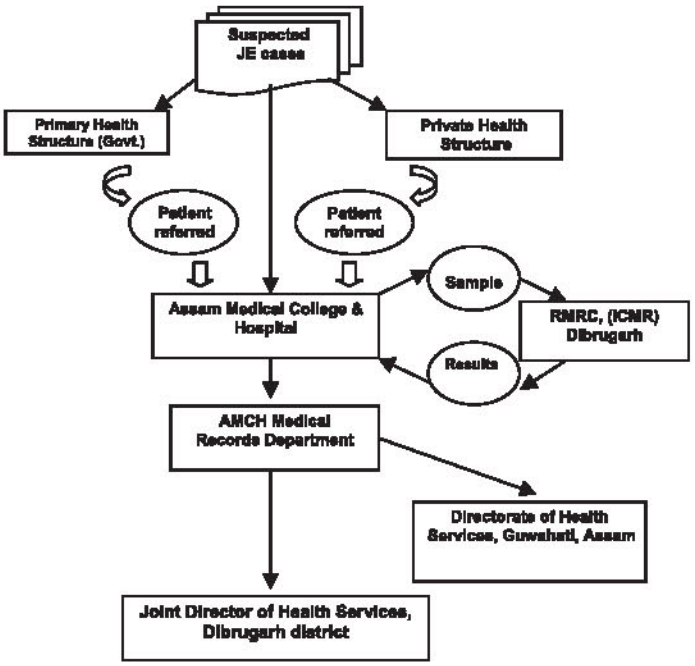

Figure 2: Information flow of JE cases

2.3.1 Spatial mean centre

The Mean Centre (MC) or spatial mean centre gives the average location of set of points. Here, locations of the villages having JE cases are considered for measuring the mean centre of their locations. This will indicate the area and extension around which JE cases are occurring. This information is critical for the health authorities to plan intervention measures for covering the geographical extent of the disease occurrence. Mean centre in terms of geographic latitude and longitude are measured as:

$$(\bar{x}_{mc}, \bar{y}_{mc}) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right)$$

Equation 1

Where, $\bar{x}_{mc}, \bar{y}_{mc}$ are coordinates of the mean centre, x_i, and y_i are coordinates of point i and n is the number of points/polygons (village locations in our study).

2.3.2 Directional distribution (Standard deviational ellipse)

Directional distribution (Standard Deviational Ellipse) measures whether a distribution of features exhibits a directional trend (whether features are farther from a specified point in one direction than in another direction). The standard distance circle shows the spatial spread of a set of point locations. The steps in deriving the standard deviational eclipse are as follows (Chou, 1997).

1. Calculate the coordinates of the mean centre (x_{mc}, y_{mc}). This will be used as the origin in the transformed coordinate system.
2. For each point, p_i in the distribution transform its coordinate by:

$$x'_i = x_i - x_{mc} \quad y'_i = y_i - y_{mc}$$

3. calculate the angle of rotation θ such that:

$$\tan\theta = \frac{\left(\sum_{i=1}^{n} x'^2_i - \sum_{i=1}^{n} y'^2_i\right) + \sqrt{\left(\sum_{i=1}^{n} x'^2_i - \sum_{i=1}^{n} y'^2_i\right)^2 + 4\left(\sum_{i=1}^{n} x'_i \sum_{i=1}^{n} y'_i\right)^2}}{2\sum_{i=1}^{n} x'_i \sum_{i=1}^{n} y'_i}$$

Equation 2

With θ from the step 3 we can calculate the deviation along x and y axes in the following manner:

$$\delta_y = \sqrt{\frac{\sum_{i=1}^{n} \left(x'_i \sin\theta - y'_i \cos\theta\right)^2}{n}} \quad \delta_x = \sqrt{\frac{\sum_{i=1}^{n} \left(x'_i \cos\theta - y'_i \sin\theta\right)^2}{n}}$$

Equation 3

Directional distribution of the disease cases is important to monitor the spread of the disease over time. Identification of areas in the direction of disease spread will help in taking early precaution measures.

2.3.3 *Measure of spatial autocorrelation*

It is of interest to see the spatial distribution pattern of JE reporting villages in the district. If we observe the JE reporting villages to follow any kind of clustering pattern, we may like to relate the occurrence of the disease with underlying landscape and socio-economic features. In classifying spatial patterns as either clustered, dispersed or random, we can focus on how various points or polygons are arranged. We can measure the similarity or dissimilarity of any pair of neighbouring points or polygons. When these similarities and dissimilarities are summarised for spatial pattern, we have the spatial autocorrelation (Lee and Wong, 2001). Here, high autocorrelation would imply the occurrence of villages with higher number JE cases and the correlation is attributable to the geographic ordering of the villages. The most commonly used spatial auto-correlation statistic, Moran's I coefficient (Chou, 1997) is employed to measure the autocorrelation (Equation4-6). Moran's I can be defined as:

$$I = \frac{n\sum\sum w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W\sum(x_i - \bar{x})^2}$$

Equation 4

where, $W = \Sigma_i\Sigma_j w_{ij}$ Here, Euclidean distance is used to define the weights w_{ij}. Corresponding to each pair of sample points i and j, let d_{ij} represent the distance between them. The distance weight is applied in an inverse manner, since the intensity of spatial relationship diminishes when the distance increases. Hence $w_{ij} = 1 / d_{ij}$. When no spatial autocorrelation exists, the expected value of Moran's I is:

$$E(I) = -\frac{1}{(n-1)}$$

Equation 5

$$Var(I) = \frac{n^2 W_1 - nW_2 + 3W^2}{W^2(n^2 - 1)}$$

Equation 6

where, $W_1 = \left(\frac{1}{2}\right)\Sigma_i\Sigma_j\left(w_{ij} + w_{ji}\right)^2$

$W_2 = \Sigma_i\left(\Sigma_j w_{ij} + \Sigma_j w_{ji}\right)^2$

Here, n is the total number of geographic units (villages), x_i denotes number JE cases corresponding to i^{th} sample point. The value of the Moran's I coefficient ranges between -1 and 1. A larger positive value implies a clustered pattern, while a negative value significantly different from 0 is associated with scattered pattern. When the Moran's I coefficient is not significantly different from 0, there is no spatial autocorrelation and the spatial pattern is considered to be random. Spatial statistics tool in ArcGIS software used to measure spatial autocorrelation is based not only on locations of the village alone or on number of JE cases alone, but on both village locations and corresponding number of JE cases simultaneously. Given a set of village and associated JE cases, it evaluates whether the pattern expressed is clustered, dispersed or random. A 'Z' score is calculated to assess whether the observed clustering / dispersion is statistically significant or not. The Z score is calculated as:

$$Z = \frac{O(I) - E(I)}{SD(I)}$$

Equation 7

2.3.4 *Delineation of JE hot spots*

Moran's I has well-established statistical properties to describe spatial autocorrelation globally. However, it is not effective in identifying different type of clustering spatial patterns. These patterns are sometimes described as 'hot spots' and 'cold spots'. If high values are close to each other, Moran's I will indicate relatively high positive spatial autocorrelation. The clusters of high values may be labeled as a hot spot. But high positive spatial autocorrelation indicated by Moran's I could be created by low values close to each other. This type of clusters can be described as cold spot. Delineation of these hot spots and cold spots will help in optimising the use of resources for timely interventions. The *G statistics* (Getis and Ord, 1992) has the advantage of detecting the presence of hot spots or cold spots over the entire study area (Equation 8-11). A local measure of spatial autocorrelation is the local version of the General G statistics. The local G statistics is derived for each aerial unit to indicate how the value of aerial units of concern is associated with the values of surrounding aerial units defined by a distance threshold d. The Local G statistics is defined as:

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad ; \; i \neq j$$

Equation 8

This G statistics is defined by a distance d, within which the aerial units can be regarded as neighbours of i. The weight $w_{ij}(d)$ is 1 if aerial unit j is within d, or 0 otherwise. Thus the weight matrix is essentially a binary symmetrical matrix, but the neighbouring relationship is defined by distance, d. The sum of the weights is:

$$W_i = \sum_j w_{ij}(d) \quad \text{for } j \neq i$$

Equation 9

Basically, the numerator of (8) which indicates the magnitude of G_i (d) statistics will be large if neighbouring features (No of JE cases) are large and small if neighbouring values are small. A moderate level of $G_i(d)$ reflects spatial association of high and moderate values, and a low level of $G_i(d)$ indicates spatial association of low and below average values. Before calculating the G statistics we need to define a distance d, within which aerial units will be regarded as neighbours. In this exercise we have defined d as a distance of 500 meters based on the extent of the study area and spatial distribution of villages. So the Village points will be regarded as neighbours if they are within an aerial distance of 500 meters. For detail interpretation of the general G statistics we have to rely on its expected value and standardised score (Z score). To derive Z score and to test for the significance of the general G statistics, we have to know the expected value of $G_i(d)$ and its variance. The expected value of $G_i(d)$ is:

$$E(G_i) = \frac{W_i}{(n-1)}$$

Equation 10

where, $W_i = \sum_j w_{ij}(d)$

The expected value of $G_i(d)$ indicates the value of $G_i(d)$ if there is no significant spatial association or if the level of $G_i(d)$ is average. Then we need to derive the Z score of the observed statistics based on the variance. According to Getis and Ord (1992) the variance of $G_i(d)$ is:

$$Var(G_i) = E(G_i^2) - [E(G_i)]^2$$

Equation 11

Where:

$$E(G_i^2) = \frac{1}{\left(\sum_j x_j\right)^2}\left[\frac{W_i(n-1-W_i)\sum_j x_j^2}{(n-1)(n-2)}\right] + \frac{W_i(W_i-1)}{(n-1)(n-2)}$$

Where, n denotes the number of aerial units (villages) in the entire study area.

3. Results and Discussion

3.1 Spatial Pattern of Progression of JE Cases

Spatial autocorrelation among JE reporting villages have been measured with Global Moran's I index. Global Moran's I, O(I) calculated with all the villages reporting JE at least once during the study period is found to be 0.00683 with Expected value E(I) 0.00093. Z score is found to be 5.04, which is significant at 99% confidence level ($p < 0.01$). These results confirm that spatial distribution of JE occurrence is non-random (Figure 3) and hence calls for special attention in the clusters of high JE occurrence. Mean centres of JE case distribution within the Dibrugarh district has been computed for the years 1985-2005 at five yearly intervals. The observation shows location of mean centre at the border of Lahowal and Panitola PHCs in 1985, gradually shifting to a tri-junction of Lahowal, Barbaruah and Khowang PHCs during 1990, 1995 and 2000 and shifting further south near a quadri-junction of Tengakhat, Lahowal, Khowang and Naharani PHCs (Figure 4). Similarly, five standard deviational ellipses drawn at five yearly intervals show a gradual directional shift towards the south eastern part of the district. JE case distribution is also observed to spread to wider areas in successive years. Detail study in the district in terms of physiographic, socio-economic and climatic factors will be required to explain the shift of disease occurrence in the district.

3.6 JE Hot Spots and High Risk Areas

The general *G statistics* has been calculated to delineate the areas based on whether large number of cases tends to cluster in the area. In other words it will identify the JE hot spots in the study district. Highest value of G_i is calculated to be 8.374 and the lowest is -1.908. Z Scores have been calculated for testing the statistical significance. Villages with Z score more than 2.56 has been considered to significant at 99% confidence level ($p<0.01$) and put in the hot spot category. Three prominent clusters have been categorised as JE hot spots. The biggest hot spot with 54 villages is observed around the district head quarter on the bank of Brahmaputra river. The second hot spot with 4 villages has been observed near Namrup tea garden. Another hot spot with two villages has been observed near Ekoratoli tea garden. Since the location of major clustering of the disease cases (hot spots) are identified in the eastern and western part of the district, mean centre of spatial distribution of disease cases are found to be located around central part of the district. These clusters of villages need immediate attention in terms taking long term intervention measures. There are 422 villages in the district which are found to have Z value more than 1.65 (significant at 90% confidence level). These villages are categorised as High JE risk villages (Figure 5).

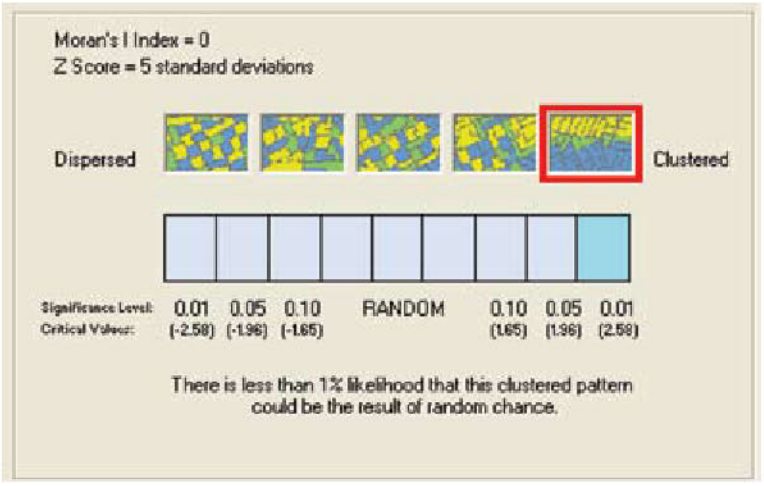

Figure 3: Autocorrelation among JE reporting villages

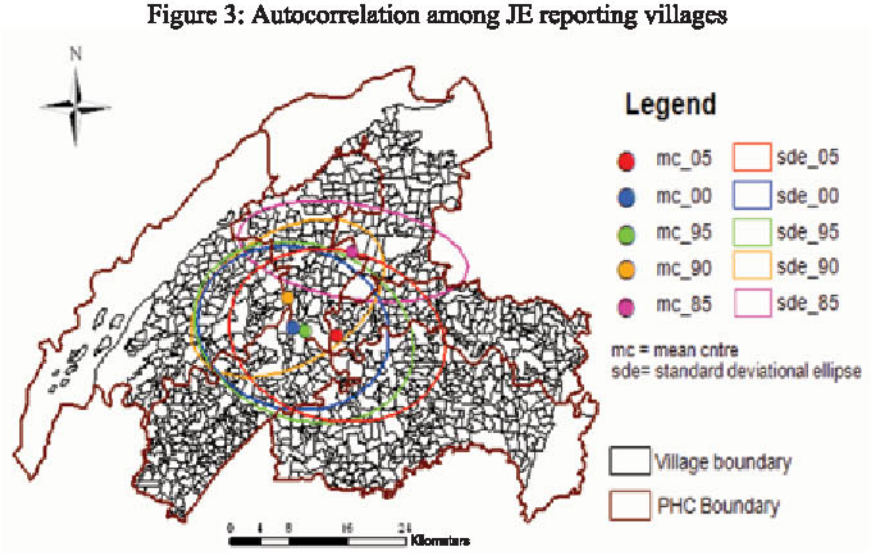

Figure 4: Spatial mean centre and standard deviational ellipses of JE reporting villages

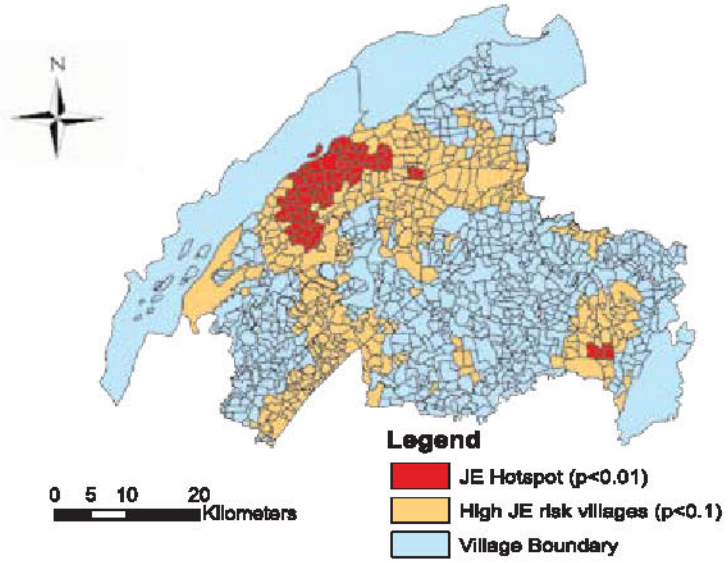

Figure 5: JE incidence hot spots and high JE risk villages in Dibrugarh

4. Conclusion

The study elucidates the dynamics of JE in the study area and shows the shift of disease occurrences over time. It also shows the potential application of spatial statistics analysis to delineate the disease incidence hot spots at village level. Out of 1048 villages in the district, 60 villages have been found to be within the disease hot spots. This accounts for about 6% areas in the district with the risk of very high JE incidence. Maximum attention should be given to these villages by the health department authorities to minimise fatalities. This may be followed by taking precaution measures in the 422 villages identified as high risk areas. Information generated in this exercise will serve as baseline information and will help in future monitoring of the disease in the district. Detail study in the hot spots in terms of physiographic, socio-economic and climatic factors may be carried to understand the critical factors responsible for disease transmission and outbreak.

Acknowledgements

The authors are thankful to Director, Regional Medical Research Centre, NE Region and Director, North Eastern Space Applications Centre for their guidance and encouragements. Sincere thanks also due to Joint Director of Health Services, Govt. of Assam and Directorate of National Vector Borne Disease Control Programme (NVBDCP), Govt. of India for providing historical data on JE cases in the study area.

References

Abeku, T. A., de Tarekegn, A., Vlas, S. J., Borsboom, G., Teklehaimanot, A., Kebede, A., Olana, D., van Oortmarsen, G. J., and Habbema, J. D. F., 2002, Forecasting Malaria Incidence from Historical Morbidity Patterns in Epidemic-Prone Areas of Ethiopia: A Simple Seasonal Adjustment Method Performs Best. *Tropical Medicine and International Health,* 7(10), 851-857.

Abelardo, C., Moncayo, J., Edman, D., and John, T. F., 2000, Application of Geographic Information Technology in Determining Risk at Eastern Equine Encephalomyelitis Virus Transmission. *Journal of American Mosquito Control Association,* 16(1), 28-25.

Banerjee, K. 1996, Emerging Viral Infections with Special References to India. *Indian Journal of Medical Research,* 103, 177-200.

Barnes, C. M., and Cibula, W. G., 1979, Some Implications of Remote Sensing Technology in Insect Control Programmes Including Mosquitoes. *Mosquito News*, 39, 271-282.

Back, E. L., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Rejmankova, E., Ulloa, A., Meza, R. A., Robert, D. R., Paris, J. F., Spanner, M. A., Washino, R. K., Hacker, C., and Legters, L., 1994, Remote Sensing as a Landscape Epidemiological Tool to Identify Village at High Risk for Malaria Transmission. *The American Journal of Tropical Medicine and Hygiene.* 51(3), 271-280.

Chou, Y. H., 1997, *Exploring Spatial Analysis in Geographic Information Systems*. (Onwards Press).

Dhiman, R. C., 2000, Remote Sensing: A Visionary Tool in Malaria Epidemiology. *ICMR bulletin.* 30(11), 1-2.

Getis, A., and Ord, J.K., 1992, The Analysis of Spatial Association by use of Distance Statistics. *Geographical Analysis*, 24 (3), 189-207.

Glass, G. E., Schwartz, B. S., Morgan, J. M. III, Johnson, D. T., Noy, P. M., and Israel, E.,1995, Environmental Risk Factor for Lyme Disease Identified with Geographic Information System. *American Journal of Public Health.* 85, 944-948.

Handique, B. K., Khan, S. A., Goswami, J., Chakraborty, K., Bora, R., and Sarma, K. K., 2005, Characterization of JE Prone Areas of Dibrugarh District of Assam using Remote Sensing and GIS. *Proceeding of the 25th ISRS Annual Convention and National Symposium,* 6-9 Dec, 2005, Ranchi, Jharkhand, India.

Hay, S. I., Packer, M. J. and Rogers, D. J., 1997. The Impact of Remote Sensing on the Study and Control of Invertebrate Intermediate Hosts and vectors for disease. *International Journal of Remote Sensing,* 18(14), 2899-2930.

Hendrickxy, G., Napala, A., Dao, B., Batawuli, D., DeDeken, R., Vermilien, A., and Stingenberg, J. H.,W., 1999, A Systematic Approach to Area Wide Tsetse Distribution and Abundance Maps. *Bulletin of Entomological Research*. 89, 231-244.

Jeganathan, C., Khan, S. A., Ramesh, C., Singh, H., Srivastava, V., and Raju, P. L. N., 2001, Characterisation of Malaria Vector Habitats using Remote Sensing and GIS. *Journal of the Indian Society of Remote Sensing*. 29(2), 31-36.

Kabilan, L., Rajendran, R., Arunachalam, N., Ramesh, S., Srinivasan, S., Philip, S., and Dash, A. P., 2004, Japanese Encephalitis in India: An Overview. *The Indian Journal of Paediatrics,* 71(7), 609-615.

Khan, S. A., Narain, K., Handique, R., Dutta, P., Mahanta, J., Satyanarayana, K., and Srivastava, V. K., 1996, Role of some Environmental Factors in Modulating Seasonal Abundance of Potential Japanese Encephalitis Vectors in Assam. *Journal of Tropical Medicine and public Health*. 27 (2), 382-391.

Lawson, A. B., 2001. *Statistical Methods in Spatial Epidemiology* (Singapore: John Wiley and Sons Ltd.)

Lee, J., and Wong, D. W. S., 2001, *Statistical Analysis with ARCVIEW GIS*. (Singapore: John Wiley & Sons Ltd.)

Nageswara Rao, P. P., Elango, H. R., Rajmohan, H. R., and Krishna Murthy, V., 2004, Satellite-Based Assessment of Physiographic Disposition to Health Problems in Kasargod District, Kerala, India. *Journal of the Indian Society of Remote Sensing*, 32(1), 75-79.

Sabesan, S., Konuganti, H. K. R., and Perumal, V., 2008, Spatial Delimitation, Forecasting and Control of Japanese Encephalitis: India –a case study. *The Open Parasitology Journal*, 2, 59-63.

Srivastava, A., Nagpal, B.N., Saxena, R. and Subbarao, S.K., 2001, Predictive Habitat Modelling for Forest Malaria Vector Species. *An. dirus* in India – a GIS based Approach. *Current Science*. 80(9), 1129-1134.