Multi-disciplinary Development of an Early Warning and Automated Response System (EWARS) for Epidemic Prevention

Ajit N. Babu[a] and Engelbert Niehaus[b]

[a]Saint Louis University,
St. Louis, USA

[b]University of Koblenz-Landau
Pfalz, Germany

ajitnbabu@gmail.com; niehaus@uni-landau.de

ABSTRACT

It is becoming increasingly clear that changes in the environment can promote growth of disease vectors like mosquitoes and rodents as well as rise in water-borne illnesses like cholera and other enteric diseases. Our paper is based on the concept that a multidisciplinary informatics scaffold can serve as an Early Warning and Automated Response System (EWARS) for a variety of communicable diseases. It will utilize multiple information resources including Remote Sensing (RS), Global Positioning Systems (GPS) and Geographic Information Systems (GIS) for exposure assessment (John et al 2004). Early warning systems are not new (Witt et al 2009). However, what is novel in the proposal is the integration of an adaptive Fuzzy Logic Decision Support System to the early warning component being developed by the investigative team. It also markedly differs from other efforts to provide decision-support in the unique degree of detail and process guidance delivered as well as in the broad multidisciplinary nature of the study group. Use of such a system would enable prophylactic public health interventions to be rapidly deployed based on a real-time, scientifically-based assessment of the threat environment. Particular attention would be paid to environmental factors amenable to corrective interventions. The informatics component of the system would be based on open source software, and the software developed during this project will also be in the open source domain. The project is multinational and multidisciplinary.

Keywords: early warning systems; human health; environmental models; fuzzy logic

1. INTRODUCTION

In 1985, NASA initiated the Biospheric Monitoring and Disease Prediction Project, to determine if remotely-sensed data could be used to identify and monitor environmental factors that influence malaria vector populations. Initial studies used high resolution images from LANDSAT to monitor the development of canopy cover in Californian rice fields. Changes in rice canopy cover over the season were successfully used to predict fields with high or low mosquito densities.

Another long-term NASA project is the Malaria Modeling and Surveillance Project by the Armed Forces Research Institute of Medical Sciences in Thailand and the US Naval Medical Research Unit in Indonesia (Kiang et al 2006).

The European Space Agency's two-year Epidemio project, launched in 2004, combined data from Earth Observation satellites, such as ESA's Envisat or the French Space Agency's Spot, with field work to combat the spread of epidemics. Specific situations in which Epidemio assisted in the analysis of and response to epidemics include studies of Ebola hemorrhagic fever in Congo and Gabon. Combining ESA Envisat satellite data on water bodies, forest cover and digital elevation models (DEMs) with field results, scientists were able identify dryness and drought as key factors connected to the Ebola epidemics. High-resolution images through Epidemio also helped in controlling an outbreak of the Marburg virus in Angola in 2005.

In China, there have been recent studies on Highly Pathogenic Avian Influenza, identifying risk factors for transmission. A study published in 2008 compared confirmed HPAI H5N1 outbreaks in poultry and wild birds, as well as 21 human cases in mainland China during 2004-2006. These data, together with information on wild bird migration, poultry densities, and environmental variables (water bodies, wetlands, transportation routes, main cities, precipitation and elevation), were integrated into a Geographical Information System (GIS). A case-control design was followed by multivariate logistic regression analysis which revealed that minimal distance to the nearest national highway, annual precipitation and the interaction between minimal distance to the nearest lake and wetland, were important predictive environmental variables for the risk of HPAI (Fang et al 2008).

A variety of internationally connected efforts are active in various parts of Africa to use RS and GIS for control of Malaria, Leishmaniasis and other vector borne disorders, particularly focusing on weather anomalies and rainfall related indices. The major thrust has been on malaria given the high levels of mortality and morbidity associated with this disease on the African continent (Ceccato 2005).

There have been a number of studies in India with the assistance of the Indian Space Research Organization (ISRO) that have also evaluated the use of remote sensing with and without GIS, studying Japanese Encephalitis (JE) and Leishmaniasis in the North-East (Sudhakar et al 2006; Bhunia et al 2010), Filariasis in South India (Sabesan 2006), Cholera in West Bengal and Malaria near Delhi (Sharma and Srivastava 1997), all showing promising results.

The examples highlighted are credible attempts to further our ability to use the unique possibilities afforded by RS towards epidemic prevention. A desirable derivative of such projects is the cross-border cooperation that arises as a natural consequence, not only between states within a nation but across nations and continents. International cooperation fostered by such activities directed to the common good can not only help to protect global health, but also set new standards for political intercourse.

2. EWARS CONCEPT

Perhaps as a consequence of the rapid technological developments occurring across the realm of healthcare and information technology, all too often the focus of programs and "research" becomes the technology itself rather than the outcome of interest. The burden of communicable disease has been a challenge to the public health machinery in the Third World. While the specter of communicable diseases has been receding, there remains the ever present danger of the resurgence of diseases such as malaria or new problems as exemplified by the recent H1N1 outbreak. The potential for problems can be amplified by the large number of individuals from a variety of Third World countries, who may enter into countries where communicable diseases are not a significant problem as visitors or immigrants, and facilitate the transmission of a disease which the local health care system is not well equipped to detect or tackle. Clearly, a multimodal surveillance system that uses sentinel, laboratory and integrated disease surveillance will be of value in ensuring the best use of available resources to successfully mount a targeted response to potential public health threats. Our proposal is built upon the concept that a multidisciplinary informatics scaffold can serve as an Early Warning and Automated Response System (EWARS) for communicable diseases, with a the potential for later extension to address non-communicable health issues of concern to the health system in question. It will utilize multiple information resources including remote sensing (RS), Global Positioning System (GPS) and Geographic Information Systems (GIS) with appropriate methodologies to store, process and visualize the products of these multilevel information sources. Early warning systems for communicable diseases are not new. However, what is novel in our proposal is the integration of an adaptive Fuzzy Logic Decision Support System to an early warning component. It also markedly differs from other efforts to provide "decision-support" (such as the Malaria DSS developed by IVCC in Africa) in its unique degree of detail and process guidance (Yang et al 2007). It would generate a dynamic risk map followed by the delivery of highly specific advice to the public health authorities on exactly which threat to address as the highest priority, and the step by step process that should be followed in terms of allocating human resources, sanitation measures, chemicals or biological agents and so on, tailored to the level of the actor receiving it as well as to a comprehensive database of existing resources within the public health system. Use of such a system would enable prophylactic public health interventions to be rapidly deployed based on a real-time, scientifically-based assessment of the threat environment to minimize the risk of even small outbreaks which could presage a full-blown epidemic. The informatics component of the system would be based entirely on open source software, and the software developed during this project will also be in the open source domain, extending the generalizability of the scaffold. The collaborative group includes physicians, mathematicians, logisticians, health economists, entomologists, remote sensing experts and social scientists drawn from India, Germany, United States, South Africa, Canada, Ethiopia, Australia and Sweden. A number of group members are also members of United Nations Action Team 6 which seeks ways to Improve Public Health by using space-based technology.

3. SOFTWARE COMPONENTS

The core software component is the GIS in which temporal spatial data is stored. The approach is necessary to utilize the tools visualization of GIS for disease prevention (Yang et al 2007). The capabilites of visualization and processing of spatial data is the first step in providing decision support. All data type definition of records related to temporal spatial modelling include the core record with attached time, longitude, latitude and height. Open-Source software is used because it allows to share implementations without the licensing costs associated with proprietary software. Especially when a successfully implemented pilot should be replicated the Total Cost of Ownership can be reduced. The technical parameters of the developed applications will be well documented, so that technical support is facilitated and others can build on the existing application for further customization or enhancement to meet their specific needs .

GRASS (Geographic Resource Analysis Support System, see http://grass.itc.it) is used because early warning of a possible epidemic should lead to a logistically optimized response utilizing available resources. Spatial analysis can be done with the statistical software R (http://www.r-project.org) for which GRASS provides an interface (http://grass.osgeo.org/statsgrass/). As computer algebra system MAXIMA (maxima.sourceforge.net) will be employed. In the final version of EWARS risk maps and resource supply maps according to risk will be delivered via web map server (http://mapserver.org).

4. MODELING METHODOLOGY

Modeling will focus on Spatial Fuzzy Logic (Petry et al. 2005) based on the concepts articulated by Zadeh (1965,1973,1979). We consider environmental factors that have an impact on the life cycle and propogation of a vector (e.g. mosquito). As a starting point we have temporal spatial data of an environmental factor (e.g. Temperature) available $(x, y, h, t, e_{1,} \dots e_n)$ where $(x, y) \in \mathbb{R}^2$ is the spatial location (i.e. longitude and latitude) *h* is the height above sea level, *t* is the time, when the data is collected and $(e_{1,\dots}, e_n)$ are the environmental variables collected at (x, y, h, t). A single environmental factor is stored in one layer of the GIS. Applying the Spatial Fuzzy Logic concept on a GIS environment, it is necessary to process the environmental data into fuzzy membership functions. Now we consider a simplified Fuzzy Logic membership function for the linguistic value

" water temperature is optimal for the larvae of mosquitoes"

If water temperature would be optimal at 21ºC for larvae of mosquitoes we define the membership function as follows:

$$f : \mathbb{R} \to [0,1] \quad x \to \frac{1}{(1+(x-b)^2)^a}$$

The graph of the function with *b=21* is dependent on *a>0*. The following two graphs show the differences for *a=0.6* and *a=5.0:*

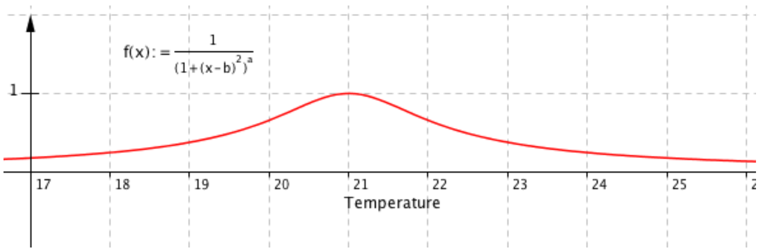

Figure 1: Graph of *f* with a=0.6

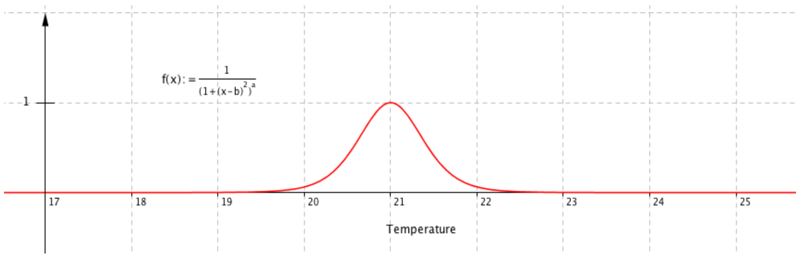

Figure 2: Graph of *f* with *a=5.0*

The parameter a defines, how tolerant the species is if the temperature deviates from the optimal value 21ºC . Let e_1 be the temperature variable. The membership function is applied on all records collected at location $(x, y, h) \in \mathbb{R}^3$ at time *t.* Doing this we get a spatial visualisation of optimal areas for the larvae of mosquitoes according to the variable temperature of water. This creates a membership f_t showing optimal areas at time *t.* We limit the domain of f_t to the subset of $\mathbb{R}^2$ with longitude and latitude $(x, y) \in \mathbb{R}^2$, so that the graph of f_t is a subset of $\mathbb{R}^3$. The height is dependent on the location $(x, y) \in \mathbb{R}^2$. The following figure is showing a membership function of f_t and the zaxis is the grad of validity of linguistic value *"temperature of water is optimal for the larvae of mosquitoes"*. At the coordinates *(4,7)* of the sample data $f_t(4,7)\approx 1$ which is indicating, that the temperature conditions for the larvae of mosquitoes are most optimal in the area with that peak. Let e_1 be the temperature at $(x, y)\in\mathbb{R}^2$, then we set:

$$f_t(x,y):=f(e_1)=\frac{1}{(1+(e_1-b)^2)^a}$$

This can be done for all $(x, y)\in\mathbb{R}^2$ for which the temperature is available.

The advantages of the transformation from degrees Celsius to a grad of validity become obvious if different membership functions are combined logically in the geographic space.

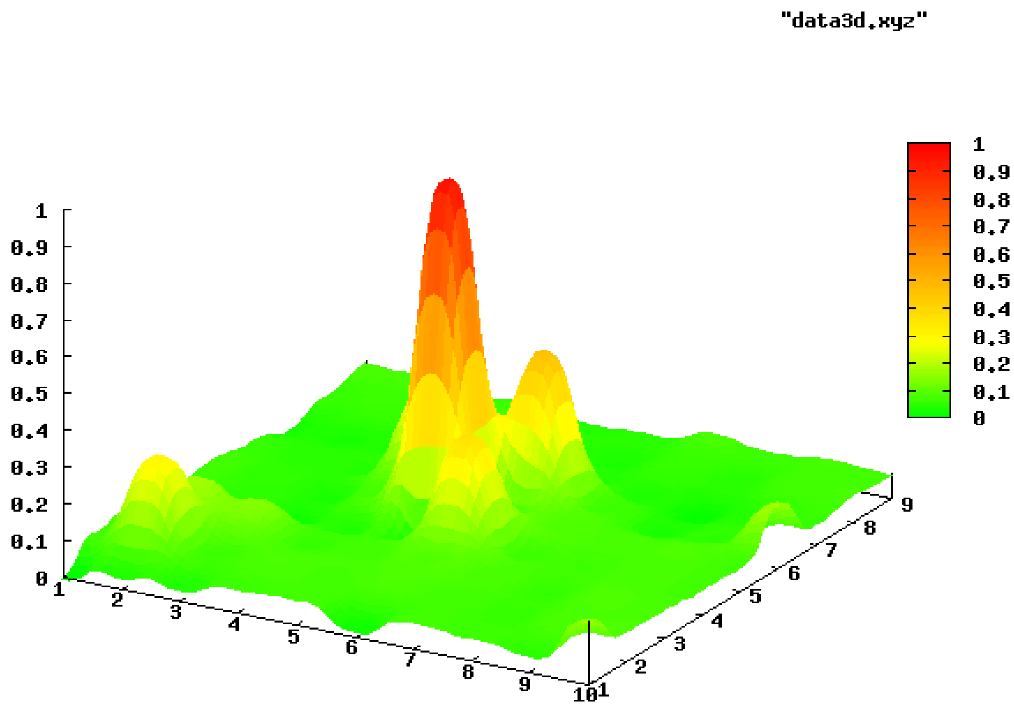

Figure 3: Spatial visualization of the optimality of temperature by membership function f_t

According to decision support system, we have only a small area where the membership function has a value close to 1 (red parts of the surface). For the decision support the validity of a linguistic value is normally dependent on more than on variable. Let q_t be another membership function that is indicating the quality of the water, so that the larvae can survive. Both properties have to be true (i.e. membership function close to 1 at $(x, y)\in\mathbb{R}^2$), so that even with perfect temperature only a few larvae can survive if there is bad water quality (e.g. larvicides). Applying a fuzzy-AND on the two membership function q_t and the membership function f_t provides a new spatial membership function indicating areas that fulfil both properties of temperature and quality of water. A fuzzy AND can be defined by the minimum.

$$\text{AND}(f_t,q_t):\mathbb{R}^2\rightarrow[0,1] \quad (x, y)\rightarrow min\{f_t(x, y), q_t(x, y)\}$$

The fuzzy OR and NOT can be defined as follows:

$$\text{OR}(f_t, q_t): \mathbb{R}^2 \rightarrow [0,1] \quad (x, y) \rightarrow max\{f_t(x, y), q_t(x, y)\}$$

$$\text{NOT}(f_t): \mathbb{R}^2 \rightarrow [0,1] \quad (x, y) \rightarrow 1 - f_t(x, y)$$

For decision support spatial rules must be carefully evaluated. We consider rules of the structure:

> IF <*environmental conditions*> THEN <*disease vector has optimal conditions*>
> IF <*vector has optimal conditions*> THEN <*application of resource R necessary*>

According to classical ("crisp") logic we have the following truth table for the implication (IF-THEN-statement " $\rightarrow$ ":

p	q	$p \rightarrow q$	$\neg p \vee q$
1	1	1	1
1	0	0	0
0	1	1	1
0	0	1	1

Table 1: Logical equivalence for IF-THEN statements

The equivalence of $p \rightarrow q$ to $p \rightarrow q$ can be transferred to fuzzy-implication applied to the membership functions p_t and q_t:

$$\text{IF-THEN}(p_t, q_t): \mathbb{R}^2 \rightarrow [0,1]$$
$$(x, y) \rightarrow max\{1 - p_t(x, y), q_t(x, y)\}$$

max is used for the Fuzzy-OR and $1 - p_t$ for Fuzzy-NOT of p_t.

With the spatial information of the validity of an IF-THEN-statement, the decision maker can see, where the rule can be applied spatially and provide the information $IF-THEN(p_t, q_t)(x, y)$ provide the grade of validity of the fuzzy-rule at location $(x, y) \in \mathbb{R}^2$.
It is obvious that we do have incomplete information about the membership function that are the basis for the Fuzzy-Rule (see Lovejoy 1991). The system is called adaptive, because permanent data collection changes the membership functions.

5. PROJECT PLAN AND IMPLEMENTATION

It is anticipated that a pilot study would be carried out in the state of Kerala this year in a semi-rural area with an area of 60 sq. km. and a population of roughly 200,000 individuals. GIS information on a variety of variables including residences, public buildings, roadways and waterbodies are already available from prior work collaborating members from the Achutha Menon Centre for Health Science Studies have performed in the same territory.

- Remote sensing and GIS data would be obtained from existing databases with the governments/science councils and academic research groups. This would be supplemented by field data (both existing and specifically collected for the project).

Standardized data formats will be adopted wherever feasible to ensure the usability of data in the open source system and in existing commercial IT infrastructure, and wherever feasible open-source GRASS would be used for geo-spatial data and R for statistical software.

- Risk mapping related to disorders of interest would be carried out using known variables as well as additional factors that may have weightage from a public health perspective. Given the current focus on mosquito-borne disorders, the group is currently considering a large number of potentially relevant variables, which are anticipated to be reduced and refined as work continues.

- Retrospective analysis of RS and GIS data from selected areas would initially be used for validating the risk mapping followed by prospective data collection.

- An automated risk score would then be developed such that the geographic area where the greatest public health threat is deemed present would be given the highest rating. The "threat" ranking would be generated taking into account not only variables compatible with high prevalence of the vector, but also proximity to human habitations, high traffic public areas and so on.

- Resources to deal with the identified problems would be automatically recommended for allocation based on the database of existing resources that would be compiled as part of the system with an adaptive "fuzzy-logic" methodology [Classical mathematical logic operates with statements and rules that can be evaluated as "true" or "false" (1 or 0 respectively). Rule based systems in the public health sector have to deal with statements that have a degree of truth ranging between 0 and 1. Fuzzy logic methodology as pointed out by Zadeh is a scientifically sound approach to implement decision support in complex environments]. The importance of spatial modeling and application of spatial rules in Geographic Information Systems are increasingly well established.

- The areas under surveillance would be reviewed using GIS and field data on a defined schedule to determine if the identified problems have been resolved or reduced in severity by intervention of the public health system. RS data would also be used where feasible to allow objective and timely verification of

- Risk scoring would change automatically and dynamically based on updated system inputs and resources would be automatically recommended for allocation in real-time based on the revised list of highest threats.

- The proposed system is a comprehensive Decision Support System. It supports health service administration in risk evaluation and allocation of resources based on risk scores. Feasibility and acceptance of EWARS at the level of administration, users and general populace will be systematically studied using both quantitative and qualitative methods incorporating elements of action research methodology, community-led development theory and constructivism theory along with more traditional approaches such as models and case studies.

6. FUTURE DIRECTIONS

The primary objectives over the next year would be to refine the project plan, implement the pilot project and pursue funding avenues as well as further strengthening collaborative connections with academic, governmental and public health authorities. The team is also exploring the possibility of extending the EWARS concept to dealing with public health problems outside the realm of communicable diseases – for example, motor vehicle accidents. Periodic updates have been presented to Action Team 6 of the United Nations, and at meetings of the UN SPIDER forum for disaster management. The group looks forward to continuing and expanding on the work commenced under this international project.

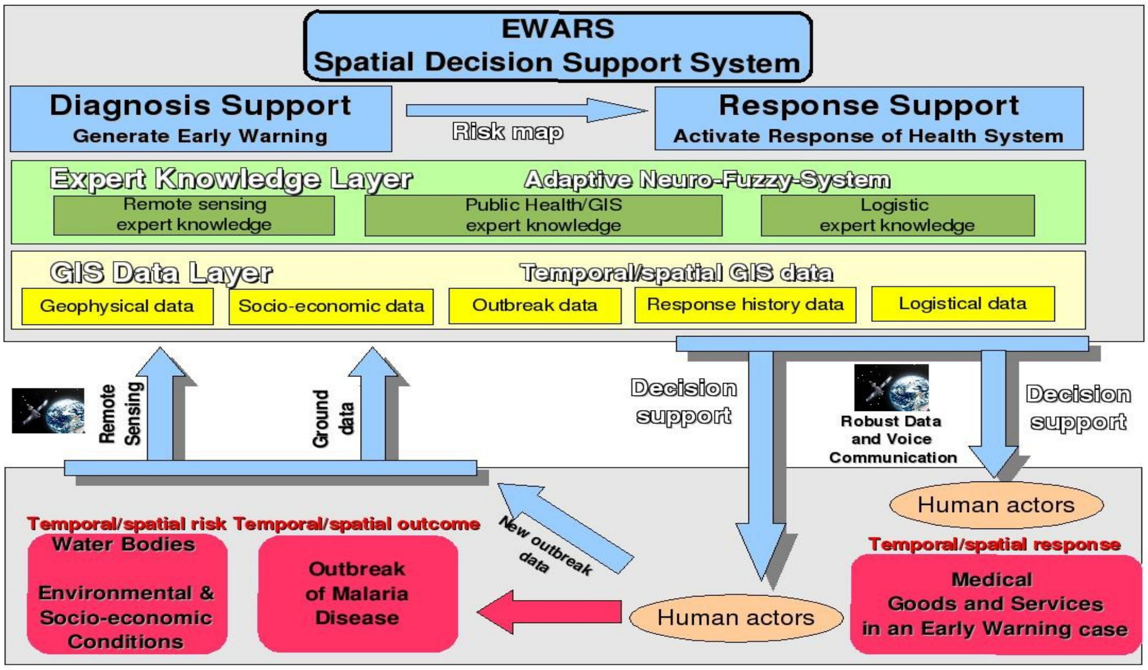

Figure 4: Schematic showing an EWARS system for Malaria

7. CONCLUSIONS

Environmental variables play a key role in the propagation of vector-borne illnesses. The use of a comprehensive, data-driven early warning and response system which can lead to the optimal use of existing public health resources offers great promise. A fuzzy logic based dynamic risk mapping approach is under development by an international group. The planned pilot project should shed light on the success of the proposed model, and its scope for adoption by public health authorities.

8. REFERENCES

Bhunia GS, Kumar V, Kumar AJ, Das P, Kesari S. The use of remote sensing in the identification of the eco-environmental factors associated with the risk of human visceral leishmaniasis (kala-azar) on the Gangetic plain, in north-eastern India. Ann Trop Med Parasitol. 2010 Jan;104(1):35-53

Ceccato P, Connor SJ, Jeanne I, Thomson MC, 2005. Application of Geographical Information System and Remote Sensing Technologies for Assessing and Monitoring Malaria Risk. Parassitologia 47, 81-96.

Fang L.-Q, de Vlas S.J., Liang S, Looman CWN, Gong P, Xu B et al. Environmental factors contributing to the spread of H5N1 avian influenza in mainland China. *PLoS ONE.* 2008;3:e2268.

John R, Nuckols W, Mary H. and Jarup L, Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiologiy Studies, in: Environmental Health Perspectives, volume 112, number 9, pages 1007-1015, 2004.

Kiang R, Adimi F, Soika V, Nigro J, Singhasivanon P, Sirichaisinthop J, Leemingsawat S, Apiwathnasorn C, Looareesuwan S. Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand. Geospat Health. 2006 Nov;1(1):71-84.

Lovejoy WS. Suboptimal policies, with bounds, for parameter adaptive decision processes, in: Operations Research, volume 41, number 3, pages 583-599, 1993.

Petry FE, Robinson VB and Cobb MA. Fuzzy Modeling with Spatial Information for Geographic Problems, Springer, Berlin, ISBN 978-3540237136, 2005.

Sabesan S, Raju HK, Srividya A, Das PK. Delimitation of lymphatic filariasis transmission risk areas: a geo-environmental approach. Filaria J. 2006 Nov 9;5:12.

Sharma VP, Srivastava A. Role of geographic information system in malaria control. Indian J Med Res. 1997 Aug;106:198-204.

Sudhakar S, Srinivas T, Palit A, Kar SK, Battacharya SK. Mapping of risk prone areas of kala-azar (Visceral leishmaniasis) in parts of Bihar State, India: an RS and GIS approach. J Vector Borne Dis. 2006 Sep;43(3):115-22.

Witt C, Anyamba, A, Linthicum K, Grieco J, Rueda L, Foley D, Schnabel D, Sang R, Richardson J and Lewis S. Predictive Vector-Borne Disease Surveillance: A Multidisciplinary Approach for the Early Detection and Response to Outbreaks, in: International Meeting on Emerging Diseases and Surveillance, 2009.

Yang K, Peng, S, Xu Q and Cao Y. A study on Spatial Decision Support Systems for Epidemic Disease Prevention Based on ArcGIS, Lecture Notes in Geoinformation and Cartography, pages 30-43, chapter 1, Springer, 2007.

Zadeh LA. Fuzzy Sets, in: Information and Control, volume 8, pages 338-353, 1965.

Zadeh LA. A Theory of Approximate Reasoning, in: Machine Intelligence, pages 149-196, Ellis Horwood/John Wiley, 1979.

Zadeh LA. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, in: IEEE Trans. on System Man and Cybernetics, volume 3, number 1, pages 28-44, 1973.