SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

JBEMi JOURNAL OF BIOMEDICAL ENGINEERING AND MEDICAL IMAGING VOLUME 1, ISSUE 1

Disease modeling – An alert system for informing environmental risk factor for TB infection

K. Ram Mohan Rao*, Yogesh Kant, Kapil Yadav**, Satya Chandra***

**National Remote Sensing Centre, Indian Space Research Organization, Hyderabad, India;*
***Indian Institute of Remote Sensing, Indian Space Research Organization, Dehradun, India;*
krmrao7@gmail.com

ABSTRACT

Tuberculosis (TB) is an infectious disease caused by the bacillus *Mycobacterium tuberculosis* and spreads through air by a person suffering from TB. A risk map is derived based on socio-economic, environmental, health facilities, and Biological characteristics for quantifying the infection risk of the locality. The inter relation between the environmental variables, meteorological parameters, and socio economic variables such as rainfall, temperature, family income and population density with respect to number of past cases have been studied using regression model. Risk is calculated by mean of the probability of occurrence of tuberculosis and vulnerability to the infection. The risk map is computed by using statistical techniques to form spatial maps. An alert system is developed by using GIS from the background of geospatial data and later published to the web using open source internet GIS technologies.

Keywords: Mycobacterium tuberculosis, Regression, vulnerability, GIS.

1. INTRODUCTION

The 1990 World Health Organization (WHO) report on the Global Burden of Disease ranked Tuberculosis (TB) as the seventh most morbidity causing disease in the world, and expected it to continue in the same position up to 2020 [1]. Each year, 8.74 million people develop TB and nearly 2 million die. This means that someone somewhere contracts TB every four seconds and one of them dies every 10 seconds [2]. Unless properly treated, an infectious pulmonary TB (i.e., the TB of lungs) patient can infect 10-15 people in a year [3]. TB is the most common opportunistic disease that affects people infected with HIV. As HIV debilitates the immune system, vulnerability of TB is increased many fold. It is estimated that without HIV, the lifetime risk of TB infected people developing tuberculosis is only 10%, compared to over 50% in the case of people co infected with Human immunodeficiency virus HIV and TB [4]. HIV is also the most powerful risk factor for the progression of TB infection to the disease. In a reciprocal manner, TB accelerates the progression of HIV in to Acquired Immune Deficiency Syndrome

DOI: 10.14738/jbemi.11.81
Publication Date: 15th February 2014
URL: http://dx.doi.org/10.14738/jbemi.11.81

(AIDS), thus shortening the survival of patients with HIV infection. Fortunately, TB is a curable disease even among the HIV infected people. The prevalence of TB and HIV co-infection worldwide is 0.18% and about 8% TB cases have HIV infection [5]. Currently, factors such as booming population, environmental pollution and rapid urbanization in many countries and global warning influence the conditions for disease outbreaks. Disease studies have revealed strong spatial aspects of disease diffusion. Thus, mapping spatial aspects of diseases could help people to understand some puzzles of disease outbreak. Unlike the raw disease data and disease maps offer a visual means of identifying cause and effect relationship existing between humans and their environment. Disease maps can enable health practitioners and the general public to visually communicate about disease distribution.

TB is generally considered to be linked to industrialization and urbanization. Peaking in the 1800 sand receding slowly after, the disease declined sharply in the West after World War II. TB has made a comeback in the last 20 years in developing countries such as China and India. Because socio-economic conditions alone cannot explain the connection between industrialization and TB. Historical statistics on coal consumption and TB disease in Canada, USA and China are correlated. A hypothesis linking TB and air pollution is developed in the context of industrialization. Historical statistics support a hypothesis linking tuberculosis and air pollution caused by coal. Trembla model is proposed whereby triggering of the interleukin-10 (IL-10) cascade by carbon monoxide in lung macrophages promotes the reactivation of Mycobacterium tuberculosis. Remotely sensed data can be used to identify, monitor and evaluate environmental factors between vector and environmental relationships. Recently, Geographic information system (GIS) and remotely sensed data are being used to evaluate and model the relationships between climatic and environmental factors with incidences of viral or bacterial borne diseases. Spatial analysis involves the use of Geographic Information System (GIS) for health that has been reviewed by several authors [6]. Both spatial and temporal changes in environmental condition may be important determinants of vector borne disease transmission. Remote sensing data can be used to provide information on spatial distribution of the vector vector-borne diseases and the physical environment [7]. Wong et al., 2006 presented on development of An Alert System for Informing Environmental Risk of Dengue Infections using Remote Sensing and GIS has discussed on incidence of dengue with the environment and climate [8]. In this study the factors used for analysis are interrelationship between ovitrap index (This is a measurement of mosquito eggs in specified geographic location, which in turn reflects the distribution of Aedine mosquitoes) and temperature. An alert system is created using risk level spatial map of locations. Use of Remote sensing and GIS helped to understand the behaviour of dengue vectors and its inevitable linkages with the environmental factors.

GIS tools have been applied to investigate the spatial relation between malaria risk and distance from breeding sites [9]. However, no attempt has been made to explain, on a larger scale, the existing disease patterns by linking disease incidence data with environmental, population, socioeconomic and entomological features on a GIS platform. Smith 2008 have carried out study on linking the tuberculosis occurrence and environmental pollution due to use of coal and he found that overall coal consumption when compared to TB notification rates, shows an apparent close relationship between global increases in coal consumption and TB disease notification rates [10]. TB incidence and coal use have increased in a similar fashion in the last 20 years, with simultaneous peaks around 1986, 1990 and 1996. However no attempt is made to relate the existing tuberculosis incidence with environmental factors like (temperature, rainfall) population, socioeconomic feature on GIS platform. Mandy Tang and Cheong-wai Tsoi done a study on GIS initiatives in improving the dengue vector control they reviewed the methodology adopted in Hongkong for control of the disease and proposed a GIS based approach to enhance the management and monitoring of the disease, by considering temporal, environmental and climatic factors. For this they used the correlation between ovitrap index and the various meteorological factors such as temperature and rainfall using conventional correlation and regression techniques [11]. This study helped them to understand that the GIS approach can be used to analyze spatial patterns of vector borne diseases. Spatial analysis techniques in a GIS can help determine the most likely areas of mosquito infestation. The GIS techniques can also help understand correlations between climatic factors and vector surveillance data (e.g. ovitrap indices). The use of GIS in disease modeling like tuberculosis requires the relationship between the environmental and meteorological factors with the incidence of disease cases in specific. This study provide social level risk map to improve prevention measures using GIS technologies. The use of RDBMS technologies to link this information with environmental factors and establish a model, and to publish this risk spatial map to web using open source internet GIS technologies.

Geographical Information Systems (GIS) has strong capabilities in mapping and analyzing not only spatial data, but also disease data, and can integrate many kinds of data to greatly enhance disease surveillance. It can render disease data along with other kinds of data like environmental data, representing distribution of contagious disease with various cartographical styles. Meanwhile, the rapid development of the internet technologies influences the popularity of web-based GIS, which itself shows great potential for sharing of disease information through distributed networks. The overall objective of this study is to generate an alert system using GIS modeling for the tuberculosis prone areas based on socio-economic environmental health facilities and individual biological characteristics. The study focuses on generation of the social risk spatial maps for tuberculosis incidences using socio-economic, environmental factors and incidences of Tuberculosis in GIS environment. Finally these spatial risk maps are published on web using open source Internet GIS technologies for the sharing of the risk indices to the general public and governmental agencies.

2. STUDY AREA

Uttarakhand is located in the northern part of India and has a total geographic area of 51,125 sq kms (figure 1). The co-ordinates of Uttarakhand are 28^{o} C 43' N to 31^{o} C 27' N (Latitude) and 77^{o} C 34' E to 81^{o} C 02' E (Longitude). Almost the entire region of Uttarakhand is covered by mountains (approximately 93%) and forests show up on about 64% of the mountains**.** The primary reason for selecting the study area is, in Uttarakhand state there are several cases of tuberculosis, especially in districts like Dheradun and Haridwar according to the performance report of the Revised National Tuberculosis Control Programme (RNTCP) there are more than 1000 cases in Dehradun in the year of 2009 and same in the other districts as well hence, it is meaningful to study about the Uttarakhand area. The spatial data is prepared up to the districts level using administrative data. Individual districts are provided with districts name, total population, income, and number of TB cases in different year.

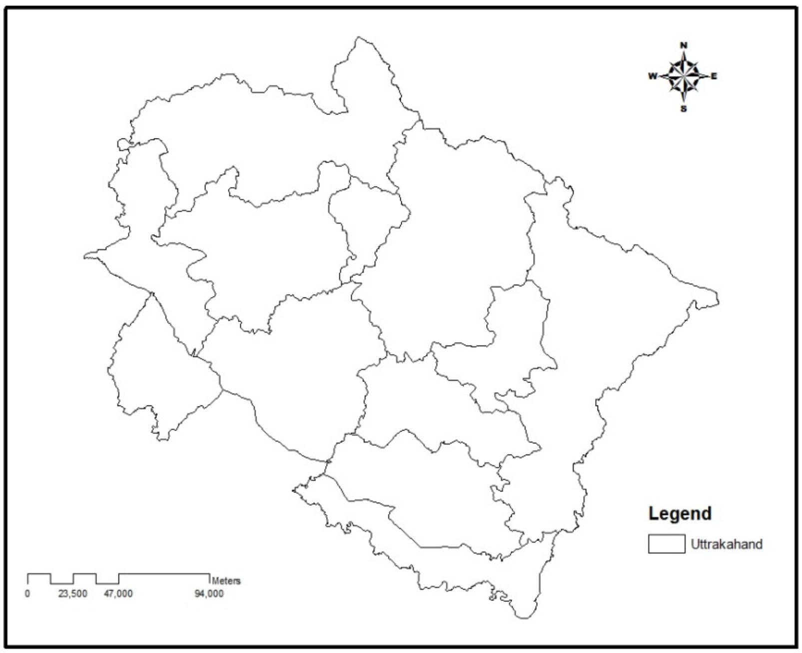

Figure 1: Location of the study area

3. METHODOLOGY

This study is carried out to identify the significant parameters namely socioeconomic indicator of Tuberculosis. Uttarakhand state map is categorized up to the district level along with district name, area, population, past TB cases, meteorological data. Tuberculosis incidence data are collected from the RNTCP (Revised National Tuberculosis Control Programme) performance report of Uttarakhand. This data include information about all suspected and confirmed tuberculosis cases reported during the year 2008 and 2009 for each quarter of the year according to each district in Uttarakhand. Data on average rainfall and temperature is obtained from the websites Indian Meteorological Department. To determine the correlation between socio economic factors and tuberculosis incidences, correlation and regression method is used. Using these parameters and tuberculosis data social risk map is generated. The risk map is then published on web using internet GIS technologies [12]. A general scheme of work flow is given in figure 2.

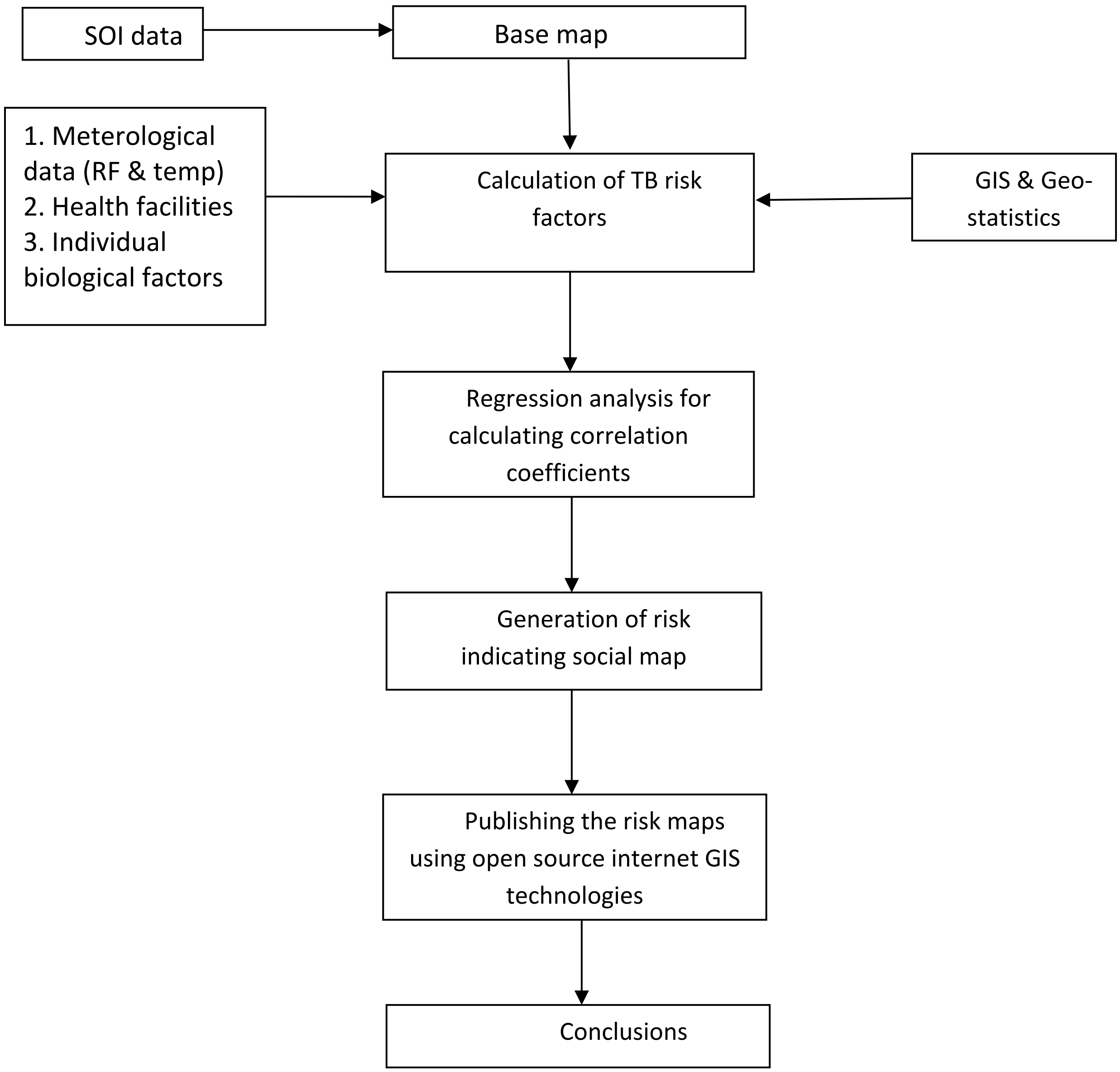

Figure 2: Methodology

Climate data is compared with tuberculosis incidences data to evaluate tuberculosis favorable conditions, because of the reason that Tuberculosis is a disease whose occurrence depends on several environmental and socio-economic variables like temperature, rainfall, population. Surrounding environment, income details based on the previous studies, and knowledge about the occurrence of tuberculosis disease factors used in this study. These factors are related with the occurrence of tuberculosis, and their individual correlation is found to show how much each of them is correlated with disease occurrence.

Now, the risk R is calculated by the formula given by Mohammed et. al given location[13].

$$R= P\times V \text{ .... } (1)$$

Where, P is probability of occurrence of the disease at given location and V is expression of certain context, condition conductive. The probability P is expressed as:

$$P=k_0+k_1*(x_1)+k_2*(x_2)+k_3*(x_3)+k_4*(x_4) \text{ .... } (2).$$

x_1=Temperature data

x_2= Rainfall data

x_3= Population

x_4= Family income

$k_{0,}k_1,k_2$ k_3 and k_4 are correlation coefficients. Similarly vulnerability V represents:

$$V= (Y_1\times Y_2\times Y_3)/Y_4 \text{ .... } (3)$$

Where,

Y_1=Individual biological factors (immuno-deficiency).
Y_2=Social and economic circumstances (crowding, poor nutrition).
Y_3=Environmental and institutional factors (silica dust, pollution).
Y_4=Availability of health facilities.

Correlation and statistical analysis are the major tools used in the study for investing and testing the relation between various parameters and tuberculosis incidences. Various variables like temperature, rainfall, population, and family income are taken in account for analysis. From the various variables the four variables namely temperature, rainfall, income and population that are significantly correlated to tuberculosis incidence are submitted to the multiple regression analysis. In the multivariate case there is more than one independent variables, therefore the regression line cannot be visualized in two dimensional spaces. However, it is possible to construct a linear equation containing all the variables in general multiple regression using the equation.

$$Y=k0+k1*(x1)+k2*(x2)+k3*(x3)+k4*(x4) \text{ ...... } (4)$$

Y=incidence of tuberculosis (dependent variable)

x1=Temperature data

x2= Rainfall data

x3= Population

x4= Family income and $k_{0,}k_1,k_2$ k_3 and k_4 are correlation coefficients.

The regression coefficient represents the independent contribution of each independent variable for the prediction of the dependent variable. Multiple regression analysis was done using the MATLAB tools. Using raster surfaces for each of the independent variables probability surface is generated. Interpolation techniques are used for generation of raster surfaces using correlation parameters for probability surface using raster calculator in Arc GIS. Avaibility of health facilities and individual biological factors like (cases of HIV) is used for generation of vulnerability surface. Since risk is the combination of probability and vulnerability is already defined risk surface is also generated using spatial analyst techniques. The final risk map is published on intranet using Geoserver tool, which is the reference implementation of the Open Geospatial Consortium (OGC) based Internet GIS tool having Web Feature Service (WFS) and Web Coverage Service (WCS) standards, as well as a high performance certified compliant Web Map Service (WMS) [12].

4. RESULTS AND DISCUSSIONS

Using socio-economic, environmental, and individual biological factors probability and vulnerability to the occurrence of tuberculosis maps are generated using GIS technologies. Based on the other factors like temperature, rainfall, population density and family income along with the correlation coefficients for each of the parameter probability of occurrence of tuberculosis is calculated. Raster surface for temperature is created using the krigging interpolation techniques, similarly raster surface is created for the average rainfall values using thiessen polygon interpolation algorithm. Using these surfaces regression coefficients are derived for probability to the occurrence as per the following equation.

$P=2830.8 + (30.30) \times RAIN - (114.6) \times Temperature + (.025) \times Population\ density + (.02) \times Family\ income$……. (5)

Using multiple regression analysis coefficients of multiple determination value for different parameters are calculated (R square) which comes out to be 0.7289 which means that 72.89 % of change in tuberculosis incidences can be explained by the change of independent variables, therefore this analysis is significant at 95% confidence interval. Table 1 shows the coefficients of multiple determination value of the different parameters.

Table 1. Coefficient of multiple determination values

Parameter	R value
Temperature	0.58
Rainfall	0.41
Population	0.09
Income	0.05
HIV	0.42
Overall	0.72

R square for temperature is 0.5681 means 56.81% of change in tuberculosis incidence can be explained by the change in temperature values. Similarly 40.12 % of change in tuberculosis can be explained by the independent change in rainfall values. As it is clear from the R square value of the regression that the independent variables are significantly correlated to the incidences of tuberculosis. Using these correlation coefficients and the raster surface created for all independent variables surface for probability is evaluated using the in Arc GIS software tools. As from the multiple regressions analysis there is significant relationship between actual and predicted values of the tuberculosis incidences. The correlation graph is as shown in the figure 3.

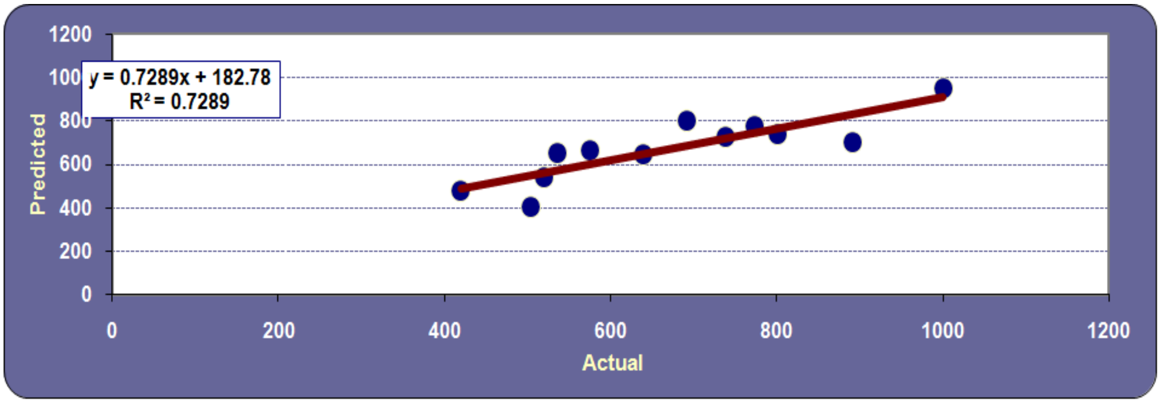

Figure 3. Actual Vs Predicted values of tuberculosis incidences

The probability surface is generated with the inputs of rainfall, population density and family income in a spatial environment. Figure 4 shows the spatial distribution of probability surface generated in GIS environment. The spatial distribution of probability of disease in the state is classified into five categories namely very high, high, moderate, low and very low.

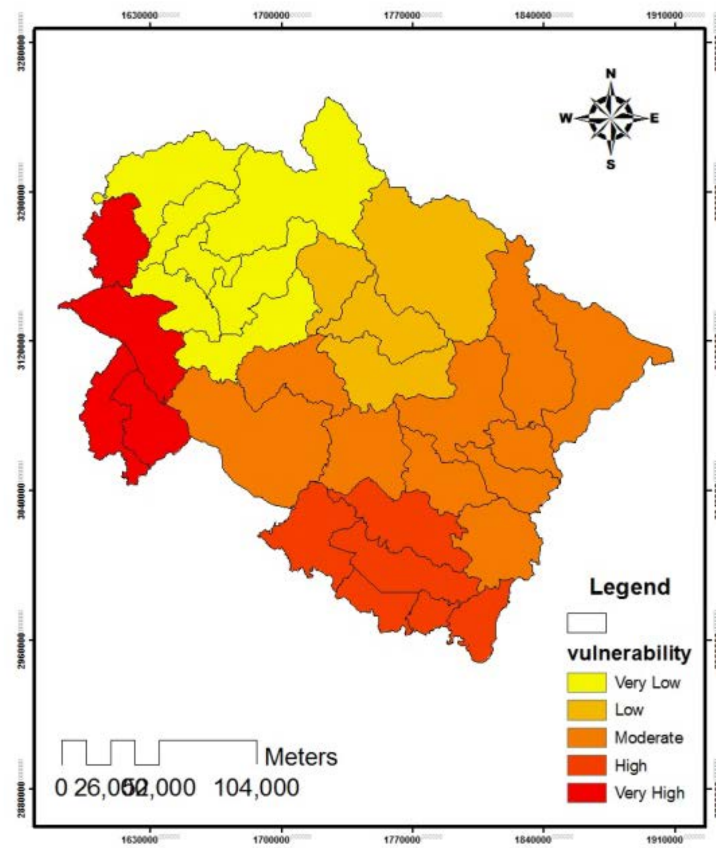

Figure 4: Probability distribution surface.

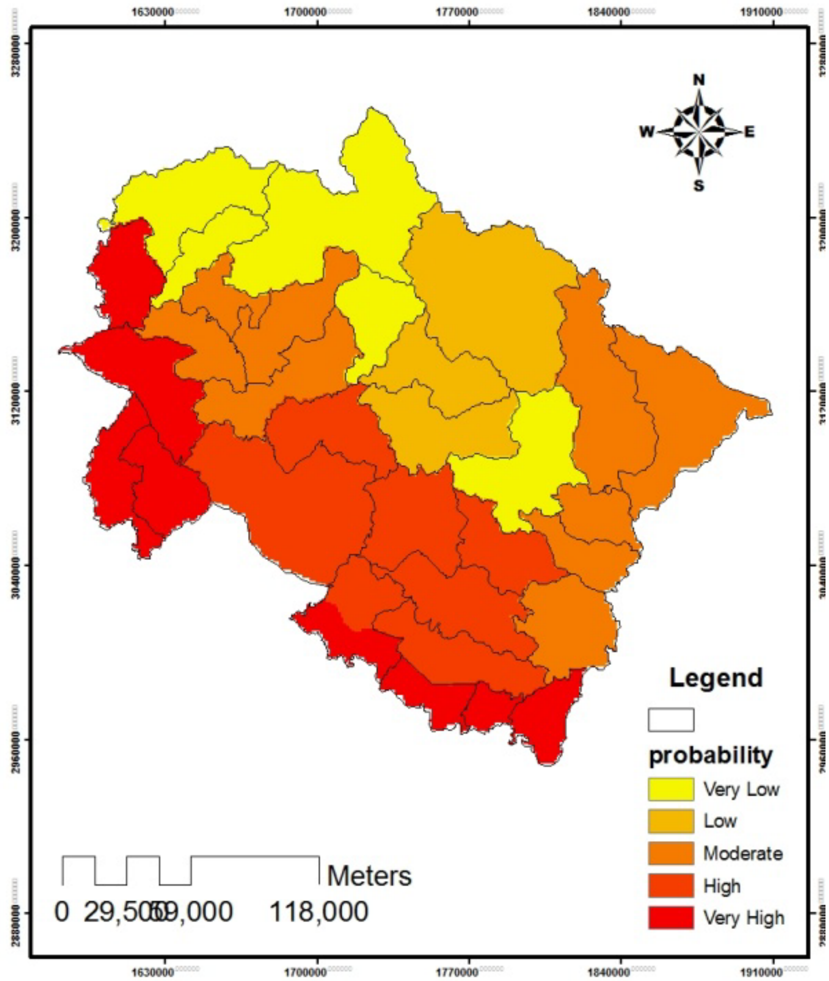

Figure 5: Vulnerability distribution surface

After evaluating the probability surface, vulnerability to the occurrence of tuberculosis is calculated, based on the vulnerability criteria as per the equation 3. Figure 5 shows the vulnerability representation of the Uttarakhand state.

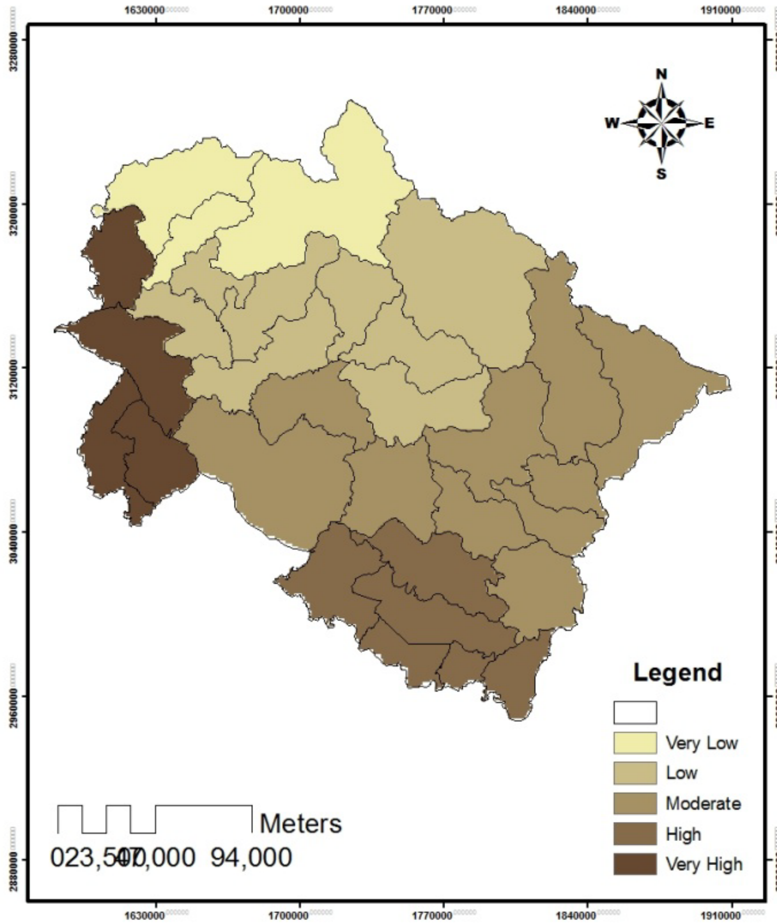

Figure 6. Risk Map for occurrence of tuberculosis in Uttrakhand

Finally the risk values for the tuberculosis are calculated on the basis of equation 1. Like probability and vulnerability surfaces the final risk is also classified in five different classes. Figure 6 shows the final risk values for the tuberculosis distribution on the state of Uttarakhand. According to the risk map obtained, the districts namely Dheradun, Haridwar, Nanital and U.S Nagar are at higher risk to the tuberculosis infection. This is because of the reason that this part of region is having more number of past cases of tuberculosis and highest cases of HIV in the state as compared to other districts, along with other dominant parameters. The temperature value is relatively high in these areas supported by high population density. Districts namely Uttrakashi, Chamoli, Ghrhwal are at low risk to the tuberculosis infections because of low temperature and low population density less number of past cases of tuberculosis. The risk map along with vulnerability and probability parameters are published on the web using open source Geoserver.

5. CONCLUSION

This study demonstrates the use of GIS technologies for disease surveillance which plays a major role in public health and epidemiology. Study also illustrate the spatial statistical analysis can help to identify and visualize potential risk areas of the disease. The criteria to visualize the spatial distribution of the risk to the tuberculosis provide the information on incidences of infections and provide a more objective way of defining risk levels with the aid of GIS unlike the conventional practice of posting statistics about the disease occurrence in tabular formats. This allows visualization of the various variables dependence on occurrence of tuberculosis in association with five risk levels for easy interpretation by the decision makers.

This study illustrates a method for expressing alert warnings in the spatial scale. The use of spatial graphics provide a clearer picture and an easy way to understand the severity or spread of tuberculosis infections which is considered more superior to the data presented in statistical and textual formats. In addition statistical analysis allows the user to explore contribution of the various factors in occurrence of tuberculosis. The study therefore serves as both objective platform for informing risks and a tool for evaluating the influence of various factors on the risk levels.

It is anticipated that the risk alert system can contribute to the prevention of tuberculosis infections in real situations. This can be effective means for of raising awareness of the public in tuberculosis risk. However an effective prevention will require active participations from all the sectors of community and individual family aiming at awareness about the causes of tuberculosis and its prevention.

REFERENCES

[1]. Murray CJL, and Lopez AD: The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. World Health Organization Document 1996, W 74 96GL-1/1996.

[2]. Narain JP (ed.): Tuberculosis-epidemiology and control. World Health Organization, Regional Office for South East Asia, New Delhi, India 2002, SEA/TB/2002. 248:15-18.

[3]. India 2005. RNTC Status Report. Central TB Division, Directorate General of Health Services, New Delhi. [http://www.tbcindia.org].

[4]. Cauthen GM, Pio A, and ten Dam HG. Annual risk of infection. World Health Organization Document 1988, WHO/TB/88.154: 1-34.

[5]. Dye C, Scheele S, Dolin P, Pathania V and Raviglione MC.1999. Global burden of disease: estimated incidence, prevalence, and mortality by country. J Am Med Assoc 1999, 282: 677-686.

[6]. Mayer JD. 1986.The role of spatial analysis and geographic data in detection of disease casusation.soc sci med.17:1213-1221.doi:10.1016/0277-9563 (83090014-X.

[7]. Hay SI, Randolph SE, and Rogers DJ. 2000. An overview of remote sensing and Geodesy for Epidemiology and Public Health application..oxford: academic press: .pp 1-35

[8]. Telzak EE. 1997.Tuberculosis and Human Immunodeficiency Virus infection. Med Clin North Am, 81: 345-360.

[9]. Gunawardena,D.M.,wickeremasinghe,A.R.;Muthuwatte, L.; Weerasingha,.1998.Malaria risk factors in an academic region of Srilanka, impact and cost implications of risk-factors based interventions. American Journal of Tropical Medicine and Hygiene 58:533-542

[10]. Smith JM, Miron M, Tremblay T, Ellis E.2008 Burden of Latent Tuberculosis Infection Among Federal Inmates 1998 to 2005. 6th Tuberculosis Conference 2008—Tuberculosis: It's a Small World, Edmonton, Alberta, March, 2008.

[11]. Mandy Tang and Cheong wai Tsoi. 2007. GIS intitiatives in improving the Dengue Vector Control. Eds. Poh. C and Ann S.H. GIS for health and management. Development in the Asia and pacific region. Springer publication.

[12]. Peng ZR, Tsou MH, 2003, Internet GIS: Distributed Geographic Information Services for the internet and wireless networks. ISBN 0-471-35923-8m.

[13]. Mohemmad Zouiten, Mostafa Harti, Chakib Nejjari, 2010. An architecture and an ontology-based context model for GIS health monitoring and alerting: Case of tuberculosis in Morocco. International Journal of Computer Science and Network Security VOL 10. No.11 November 2010;