CHAPTER 6

DISCUSSIONS

The present research aims at developing a statistical model which is helpful in predicting the anthrax outbreaks and studying the contributing factors of anthrax. Disease modelling can be used to evaluate the productiveness of control measures, besides identifying approaches to reduce economic losses due to the incidence of disease. Despite the availability of good vaccine, anthrax continues to occur and inflict economic losses to the livestock owners and requires serious measures to contain it (Shivachandra, et al., 2016). Descriptive epidemiological studies of anthrax provide valuable insights into the effects of outbreaks on population, duration and location of the outbreaks and percent mortality (Kellogg, et al., 1970). The grazing pattern and seasons will also play an important role in outbreak precipitation (Salb, et al., 2014). In the present study, descriptive epidemiological study was performed followed by development of predictive models in two scenarios, 1. Presence-only data and 2. Presence–absence data to understand the dynamics and epidemiology of the disease and to establish the control measures for prevention and control of disease.

An outbreak of anthrax in an enzootic area occurs usually after a prolonged hot dry spell, which in turn was preceded by heavy rains or with rain ending a period of drought (Islam, et al., 2013). The awareness of this fact amongst the farming community and the disease preventive and control measures taken up by the authorities might have resulted in reduced number of outbreaks. Though there is a decreasing trend in the occurrence of anthrax, it is still hyper endemic or hyper enzootic in many pockets of the country. The present study showed that Chikkaballapura taluk reported maximum number of anthrax outbreak, followed by neighbouring taluks viz., Doddaballapura and Devanahalli. The disease is endemic in Ballary and its neighbouring taluks viz., Siruguppa, Hospet etc. This can be attributed to the improper disposal of the carcasses of those animals died due to anthrax and lack of public awareness (Suma, et al., 2017). Anthrax outbreak data was reported using sentinel surveillance method, involving professional veterinarians to the department of Animal Husbandry and veterinary sciences, Karnataka. Further, monsoon months of August, September and October recorded the maximum outbreaks, peak being in the month of September. Seasonal pattern was observed, corroborating with previous studies.
Anthrax assumes highly significant place both in animal health and agricultural economy since it is peracute in nature and the farmer suffers a heavy financial loss due to the loss of valuable animals. Anthrax spores are highly resistant to extreme changes occurring in the environment. They are resistant to chemicals and disinfectants also. They can survive for decades and become active under favourable conditions for multiplication (Parker, et al., 2002). The present study on the basis of soil type revealed that more than 80% of outbreaks have occurred in Red (red loamy and red sandy) soil. Study on the basis of agro climatic zone showed that Eastern dry zone, close to central dry zone has the maximum risk of anthrax outbreak followed by part of Northern dry zone. In India, anthrax is being reported since many years both from large and small ruminants. It is one of the top ten livestock diseases reported from the country and also a major cause of deaths in livestock (Gajendragad and Uma, 2012). Anthrax in livestock is highly endemic in states like Maharashtra, Karnataka, Tamil Nadu, Andhra Pradesh, Kerala, Odisha, Jharkhand, Chattishgarh and West Bengal. The disease has also been reported from Madhya Pradesh, Gujarat, Rajasthan, Punjab and Bihar (Shivachandra, et al., 2016). Analysis of the occurrence of anthrax showed newer regions reporting anthrax. This can be attributed to either spread of the disease or improvement in reporting system adopted by the animal husbandry departments. It also gives an indication on the areas which need stringent control measures. There is a need for early diagnosis and forecasting of the disease. The present study aimed to provide the robust prediction models rationally utilizing the meteorological, remote sensing and anthropogenic variables.

Shankar et al. (2014) conducted retrospective study (2010 – 2012) on anthrax in Bangladesh. Their data analysis showed 5937 cases of anthrax, reported in almost all the districts in Bangladesh. Descriptive epidemiology in the present study on anthrax was conducted for the time period 2000 - 2016. There were 5225 cases of anthrax in a total of 402 outbreaks. Average number of cases per outbreak was 13. 20 out of 30 districts reported anthrax outbreaks during that period, whereas 56 taluks out of 177 taluks reported anthrax outbreak during the said period. More than 80% of outbreaks have been reported in a total of 17 taluks.

Descriptive epidemiology study on anthrax outbreaks gives an insight into variations in incidence of anthrax geographically and over time, but these studies are inadequate in meeting the increasing demand for controlling and prevention of the disease. Predicting the outbreak of diseases by modelling is very useful in saving the livestock population and thus averting the economic loss to the livestock farmer specifically and country in general. The
ability for early detection of the outbreaks is important to minimize the morbidity and mortality through timely implementation of disease prevention and control measures.

World Health Organization’s regional committee for the Eastern Mediterranean (1999) opined that a comprehensive and feasible strategic plan is needed to address the problem of communicable diseases and that future health scenario of communicable diseases that can predict likely, probable or even merely possible trends i.e., forecasting can play an important role in developing such a strategic plan. The committee recommended that the WHO member states with the support of WHO should continue to improve their disease surveillance systems by implementing the regional strategy for development and strengthening of epidemiological surveillance and introduce the concept of forecasting for priority communicable diseases.

Dubé et al. (2007) proposed that modelling would be most useful when applied to pre-outbreak, particularly in the areas of retrospective analysis of previous outbreaks, contingency planning, resource planning, risk assessment and training. Huppert et al. (2013) observed that an important role of modelling enterprises is that they can alert the researcher about the deficiencies in current understanding of the epidemiology of various infectious diseases, and raise crucial questions for extensive investigation and additional data that need to be collected. Siettos et al. (2013) categorized epidemiological models into three classes viz., statistical, mathematical-mechanistic state space and machine-learning based models. They suggested that public health organizations throughout the world to use such models to evaluate and develop intervention disease outbreak policies for ever-emerging epidemics.

Modelling of diseases helps to understand the biologic process of the diseases. Identification of risk factors plays a key role in the development of models. Understanding of these risk factors is essential in developing control measures for effective prevention of disease incidence. To identify the risk factors associated with anthrax, as there are limited studies on anthrax, risk factors of other diseases are also discussed. Singh and Prasad (2008) analysed the total annual average loss due to seven diseases in goats and also the economic losses due to diseases, estimated in terms of losses due to mortality, milk yield, body weight and opportunity cost through mathematical modelling of economic losses due to some important diseases in goats in India. Devi et al. (2003) used Multiple Linear Regression method (backward elimination method) to analyse the environmental factors related to malaria incidences such as vegetation cover, rainfall, water body, temperature and humidity in Salem
in Tamilnadu in India. They used remote sensing and geographic information system (GIS) for monitoring the environmental factors associated with vector-borne disease – malaria. Chatterjee et al. (2009) developed a non linear regression methodology in modelling and forecasting malaria incidence in Chennai city, India using multi-step polynomial regression model. They considered three types of data to develop the regression methodology i.e., a longer time series data of Slide Positivity Rates (SPR) of malaria, a smaller time series data (deaths due to Plasmodium vivax) of one year and spatial data (zonal distribution of Plasmodium vivax deaths) for the city along with the climatic factors viz., minimum temperature, maximum temperature, minimum humidity, maximum humidity, rainfall; population and previous incidence of the disease. Tuyishimire (2013) identified and mapped factors responsible for causing malaria in Ruhuha sector, Rwanda, through the use of remote sensing, GIS and spatial statistics. He considered a number of environmental factors such as altitude, land use and distance to anopheles mosquito breeding sites; demographic factors such as household size, age, gender and distance to household with infected people and economic factors such as house material and animal ownership as malaria causing factors to build the model. Nizamuddin et al. (2013) used NOAA/AVHRR (Advanced Very High Resolution Radiometer) environmental satellite data to produce weather seasonal forecasts for predicting malaria epidemics in Tripura state in India. They developed an algorithm that uses Vegetation Health Indices (Vegetation Condition Index (VCI) and Temperature Condition Index (TCI)) computed from AVHRR from National Oceanic and Atmospheric Administration (NOAA) afternoon polar orbiting satellite. Kumar et al. (2014) used covariates rainfall, relative humidity, mean maximum temperature and lag period of one month for all the variables for forecasting malaria cases using climatic factors in Delhi, India. Ramachandran et al. (2016) developed an empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity in East Delhi. Johansson et al. (2016) developed a structure to evaluate and compare dengue forecasts obtained from different types of models with and without climate variables for forecasting dengue incidence in Mexico. Song et al. (2015) performed Time Series analysis of Hand, Foot and Mouth disease (HFMD) integrating weather variables such as average temperature, maximum temperature, minimum temperature, Humidity, Visibility, Mean wind speed, Maximum sustained wind speed and Precipitation amount in the analysis. Soyirı et al. (2013) developed two related negative binomial models and compared them with a naive seasonal model to forecast asthma-related hospital admissions in London using weather related data such as ambient air temperature, vapour pressure (HPa) and humidity (%). In the present study, remote sensing variables viz,
LST, NDVI, LST lag, NDVI lag; Climate variables viz., temperature, relative humidity, amount of rainfall, excessive rainfall; soil parameters viz., soil pH, type of soil, soil nutrients - Phosphorus, Potassium, Zinc, Boron, Organic Carbon and Sulphur; anthropogenic variables viz., distance from city, distance from highway, distance from road, distance from railway track and distance from water body; elevation, sunshine hours and presence of forest nearby were considered to develop count models. Because of the wide range of risk factors chosen, the models so developed are robust enough to assess the influence of these factors on the outbreak of anthrax.

Not much work has been done on developing a forecasting model on anthrax. So, to study various modelling approaches adopted by researchers, modelling methods on other diseases are discussed. Niyikora (2015) used multiple logistic regression to model diabetes with age, gender, occupation status, smoking, alcohol consumption, cholesterol level, hypertension and family history of diabetes considered as risk factors of diabetes. He observed that the best fitting model showed accuracy of more than 80% as ROC value. Ahmad et al. (2015) examined the factors that are associated directly or indirectly in pneumonia patients among the children between the age group 0 – 12 years through a study on applications of zero inflated models for health sciences data. They considered diagnosis of Pneumonia as the response variable and diabetes, influenza, septicemia, diarrhoea, asthma, anemia, tuberculosis, health services, acute tonsilitis, streptococcus pneumonia and heart disease as independent variables to model Pneumonia. They fitted Poisson, Negative Binomial (NB), Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) regression models to pneumonia data. They evaluated the model using AIC. In the present study, Poisson, Negative Binomial, Truncated Poisson and Truncated Negative Binomial were fitted to the data on outbreak of anthrax. The models were initially tested for Goodness of fit using Chi Square test statistic and further evaluated using Fit statistics AIC, AICC and BIC. AIC values were 249.18, 251.18, 231.4 and 209.2; AICC values were 279.85, 285.68, 262.1 and 236.4; BIC values were 293.74, 297.76, 275.9 and 251.7 for Poisson, Negative Binomial, Truncated Negative Binomial and Truncated Poisson respectively. The fit statistics were least for Zero Truncated Poisson distribution. So, Zero Truncated Poisson distribution was considered as best fitting model for anthrax outbreak in Karnataka at taluk level. Gunaseelan et al. (2011) used Multiple logistic regression to assess the association between the three effector variables viz., population of cattle per district, percent of villages with alkaline soil and presence of a tannery with the occurrence of greater than 25 outbreaks of anthrax in Tamilnadu in India.
The logistic regression gives dichotomy prediction. To assess the quantitative prediction of disease outbreaks in Karnataka, count models as mentioned earlier were employed. Through the best fitting Zero truncated model, it was observed that remote sensing variables LST and its one month lag, red loamy soil, sufficient amount of Phosphorus and Boron, anthropogenic variable-distance from road, presence of forest, rainfall, excessive rainfall, temperature and relative humidity are significant at 1% level whereas NDVI is significant at 5% level. The interactions of LST with its one month lag; temperature with relative humidity; one month lag of NDVI with Zinc, Sulphur and Boron and soil pH with red loamy soil showed significant influence at 1% level on the outbreak of anthrax. Additionally, these models help us to understand disease dynamics and interpretation of modelling provides the directives for taking appropriate decisions at the right time, thereby controlling the major losses due to morbidity and mortality related to disease incidence.

Many cases of anthrax go unreported since it occurs as an isolated incidence and under reporting of the cases by the field staff. Hence, mere absence of reported outbreaks is no proof of absence of the disease. Thus, the reported outbreaks only provide an index of the magnitude of the disease in India and could be an underestimate of the extent of the problem (Suma, et al., 2017). Count models developed may not be realistic as they consider disease-presence only data, as a result, prediction is restricted to area where outbreak has occurred. As the environmental factors play vital role in the disease precipitation, it is essential to consider control data such as locations where outbreak has not occurred. Hence, modelling the presence-absence scenario may be helpful in predicting the various levels of risk for entire study area.

Species distribution models (SDMs) are numerical aids that combine observations of species occurrence with environmental estimates. They are used to acquire ecological and evolutionary insights and to predict distributions over locations, sometimes requiring extrapolation in space and time (Elith et al. 2009). Many algorithms have been used in species distribution modelling for both presence only and presence-absence data.

Blackburn (2006) explored the spatial ecology and potential pathways of infection of anthrax in North America. He used ecological niche modeling with the Genetic Algorithm for Rule-set Production (GARP) to predict the geographic distribution of anthrax in the continental United States of America, considering temperature, precipitation, elevation, soil moisture, soil pH and NDVI as risk factors. He opined that there is need to study how other
environmental factors, such as wildlife densities or land use can improve model accuracy. Joyner (2010) developed Ecological niche model with the Genetic Algorithm for Rule-set Prediction (GARP) in the model-building process, which can predict the current and future potential distributions of *Bacillus anthracis* in Kazakhstan using annual mean temperature, annual temperature range, annual precipitation, precipitation of wettest month, precipitation of driest month, elevation and normalized difference vegetation index (NDVI) as predictors. He suggested that other modelling approaches also identify environmental ranges and should be explored extensively in future studies. Mullins *et al.* (2013) modelled the ecological niche of *Bacillus anthracis* in United States of America, Italy and Kazakhstan to understand the geographic distribution of anthrax and its potential associations between regional populations and ecology. They developed country-specific ecological-niche models with the Genetic Algorithm for Rule-set Production (GARP) using environmental variables i.e., elevation, annual temperature range, annual mean temperature, precipitation of the driest month, precipitation of wettest month, annual precipitation and remote sensing variable - normalized difference vegetation index (NDVI). They suggested that further studies should evaluate the impact of variable selection of models. In the present study, for the data on outbreak of anthrax at both taluk level and village level, twelve Species Distribution Models viz., Generalized Linear Model (GLM), Generalized Additive Model (GAM), Maxent, Random Forest, Gradient Boosting Machine (GBM), Artificial Neural Network (ANN), Multiple Adaptive Regression Splines (MARS), Flexible Discriminant Analysis (FDA), Classification Tree Analysis (CTA), Support Vector Machine (SVM), Naïve Bayes (NB) and Adaptive Boosting (ADA) were developed by taking anthrax presence-absence data. At taluk level, models were developed using elevation, LST, NDVI and livestock population as risk factors. At village level, models were developed using remote sensing variables viz., LST, NDVI; climate variables viz., cloud cover, potential evapotranspiration, precipitation, relative humidity, mean temperature, minimum temperature, maximum temperature, vapour pressure, wet day frequency, wind speed; and livestock population.

Steenkamp (2013) developed Species Distribution model for anthrax in the Kruger national park, South Africa. He used maximum entropy as a statistical model, with environmental variables viz., NDVI, elevation, distance of positive locations from dams, pans, rivers, springs, troughs, bore holes and water holes; soil data and vegetation data as predictors. He recommended that future studies should compare the results of the Maxent model with the results of other modelling techniques. His model showed area under the ROC as 0.9372. In
this study, Maxent was compared with other eleven models mentioned earlier. The Maxent model showed area under ROC as 0.82 at both taluk and village levels. Infact, all the twelve models were evaluated using Kappa, ROC and True Skill Statistic (TSS).

As there is dearth of modelling methods used to develop species distribution models on anthrax, SDMs developed on other diseases are discussed. Here the intention is to study various modelling methods and not the risk factors as risk factors are different for different diseases. Penna (2004) tried to evaluate recurrent neural networks as a predictive technique for time-series in the health field through a study on use of an artificial neural network for detecting excess deaths due to cholera in Ceará, Brazil. He opined that ANN displayed good predictability. ANN model in the present study showed more than 80% accuracy at village level, but at taluk level ROC was 0.53. Senthilkumar et al. (2013) provided an introduction to the theory and the importance of Multivariate Adaptive Regression Splines (MARS) in the disease diagnosis through the collected data for diabetes to develop an intelligent decision support system. In this study, MARS provided prediction accuracy of 92% through ROC at taluk level and at village level ROC was 94%. Vadicherla et al. (2013) analysed the application of Support Vector Machine (SVM) and Artificial Neural Network (ANN) for classification and prediction of heart disease. The present study showed ROC as 0.91 at taluk level and 0.92 at village level for SVM model. Liu (2014) investigated the validity of different selection strategies when using presence only and pseudo absence data while modelling the spatial distribution of dengue fever using Boosted regression tree (BRT) model. He evaluated the model using ROC. The BRT model in this study provided prediction accuracy of ROC value as 0.92 at taluk level and 0.97 at village level. Bargaje et al. (2017) proposed a Decision tree with adaptive boosting technique to classify the brain Magnetic Resonance Imaging (MRI) image as normal or abnormal. Adaptive Boosting provided prediction accuracy of 0.79 at taluk level and 0.87 at village level as ROC value in this study. Almayyan (2016) developed a model to enhance lymphatic diseases diagnosis by the use of random forest technique. The Random Forest model in this study displayed high prediction accuracy of 0.96 at taluk level and 0.99 at village level as ROC value. Further, other two evaluation statistics viz., Kappa and TSS also provided high prediction accuracy at both taluk and village levels compared to other modelling methods.

Blackburn (2006), Joyner (2010), Mullins et al. (2013) and Steenkamp (2013) used Area under the Receiver Operating Characteristics (ROC) curve to evaluate the performance of
their Ecological Niche models on anthrax. The model developed by Blackburn (2006) showed model accuracy of more than 80% as ROC value. The models developed by Joyner (2010) predicted accuracy of more than 60% as ROC value. Allouche et al. (2006) compared the responses of Kappa statistic and True Skill Statistic to prevalence using empirical data. In the present study, the models were evaluated using evaluation statistics Kappa, Receiver Operating Characteristic curve (ROC) and True Skill Statistics. At taluk level, ROC was more than 90% for Random Forest, Gradient Boosted Machine, Multiple Adaptive Regression Splines, Support Vector Machine and Classification Tree Analysis whereas ROC was more than 80% for Generalized Linear Models, Generalized Additive Models, Maxent and Naïve Bayes. TSS statistic showed more than 70% accuracy for Random Forest whereas TSS was more than 60% for Generalized Additive Models, Gradient Boosting Machine, Multiple Adaptive Regression Splines, Support Vector Machine and Classification Tree Analysis. Kappa statistic showed more than 60% accuracy only for Random Forest. At village level, ROC was more than 90% for Random Forest, Gradient Boosted Machine, Generalized Additive Models, Multiple Adaptive Regression Splines and Classification Tree Analysis whereas ROC was more than 80% for Generalized Linear Model, Maxent, Artificial Neural Network, and Adaptive Boosting. TSS statistic showed more than 70% accuracy for Random Forest, Generalized Additive Models, Multiple Adaptive Regression Splines, Classification Tree Analysis, Support Vector Machine and Adaptive Boosting whereas TSS was more than 60% for Generalized Linear Model. Kappa statistic showed more than 70% accuracy for Random Forest, Classification Tree Analysis, and Adaptive Boosting, whereas Kappa was more than 60% for Multiple Adaptive Regression Splines. For Random Forest model at taluk level, ROC and True Skill Statistics showed more than 70% accuracy. At village level, the three statistics viz., Kappa, ROC and True Skill statistics showed model accuracy of more than 70%. Infact, ROC and True Skill Statistics showed model accuracy of more than 90%. So, Random Forest was considered as the best fitting model at both taluk and village levels. Instead of depending on one single model as the best model, combination of models (Random forest and Gradient Boosting Machine), with the criteria, ROC >0.9, TSS>0.8 and Kappa > 0.5 was considered to develop average score prediction model at village level. Average score prediction model showed high risk of anthrax in Chikkaballapura and its neighbouring taluks, followed by Ballary and its neighbouring taluks.