

**MATHEMATICAL MODELS FOR PREDICTION AND
OPTIMAL MITIGATION OF EPIDEMICS**

by

SOHINI ROY CHOWDHURY

B.E., Birla Institute of Technology, MESRA, India, 2007

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2010

Approved by:

Co-Major Professor
Caterina M. Scoglio

Approved by:

Co-Major Professor
William H. Hsu

Copyright

Sohini Roy Chowdhury

2010

Abstract

Early detection of livestock diseases and development of cost optimal mitigation strategies are becoming a global necessity. Foot and Mouth Disease (FMD) is considered one of the most serious livestock diseases owing to its high rate of transmission and extreme economic consequences. Thus, it is imperative to improve parameterized mathematical models for predictive and preventive purposes. In this work, a meta-population based stochastic model is implemented to assess the FMD infection dynamics and to curb economic losses in countries with underdeveloped livestock disease surveillance databases. Our model predicts the spatio-temporal evolution of FMD over a weighted contact network where the weights are characterized by the effect of wind and movement of animals and humans. FMD incidence data from countries such as Turkey, Iran and Thailand are used to calibrate and validate our model, and the predictive performance of our model is compared with that of baseline models as well. Additionally, learning-based prediction models can be utilized to detect the time of onset of an epidemic outbreak. Such models are computationally simple and they may be trained to predict infection in the absence of background data representing the dynamics of disease transmission, which is otherwise necessary for predictions using spatio-temporal models. Thus, we comparatively study the predictive performance of our spatio-temporal against neural networks and autoregressive models. Also, Bayesian networks combined with Monte-Carlo simulations are used to determine the gold standard by approximation.

Next, cost-effective mitigation strategies are simulated using the theoretical concept of infection network fragmentation. Based on the theoretical reduction in the total number of infected animals, several simulative mitigation strategies are proposed and their cost-effectiveness measures specified by the percentage reduction in the total number of infected animals per million US dollars, are also analyzed. We infer that the cost-effectiveness measures of mitigation strategies implemented using our spatio-temporal predictive model have a narrower range and higher granularity than those for mitigation strategies formulated using learning-based prediction models.

Finally, we coin optimal mitigation strategies using Fuzzy Dominance Genetic Algorithms (FDGA). We use the concept of hierarchical fuzzy dominance to minimize the total number of infected animals, the direct cost incurred due to the implementation of mitigation strategies, the number of animals culled, and the number of animals vaccinated to mitigate an epidemic. This method has the potential to aid in economic policy development for countries that have lost their FMD-free status.

Table of Contents

Table of Contents	v
List of Figures	viii
List of Tables	xv
Acknowledgements	xviii
Dedication	xix
1 INTRODUCTION	1
2 MATHEMATICAL MODELS FOR PREDICTING FOOD AND MOUTH DISEASE	5
3 PROPOSED NETWORK-BASED MODEL	10
3.1 Model Assessment	15
3.1.1 Model Calibration	16
3.1.2 Model validation	19
3.2 Simulation of Infection Networks	23
3.3 Ergodicity of Infection Networks	30
4 LEARNING-BASED PREDICTION MODELS	36
4.1 Experimental Data	39
4.2 Statistics for model parameterization	39
4.3 Model Description	42
4.3.1 Neural Network Models	42
4.3.2 Autoregressive Models	46

4.3.3	Bayesian Network Model	48
4.3.4	Monte-Carlo Simulations	52
4.4	Model Analysis	54
5	MITIGATION STRATEGIES	68
5.1	Theoretical Mitigation	69
5.1.1	Ergodicity for theoretical mitigation	71
5.1.2	Theoretical network fragmentation	71
5.2	Practical Mitigation	76
5.2.1	Learning-Based Models	81
5.2.2	Network Based Models	82
5.2.3	Impact of prediction errors on mitigation strategies	87
6	OPTIMAL MITIGATION USING GENETIC ALGORITHMS	95
6.1	Methodology for Genetic Algorithms	95
6.2	Optimal Mitigation using GA	97
6.2.1	Theoretical Framework	98
6.2.2	Fuzzy Dominance	98
6.2.3	Mediality	100
6.2.4	Evolutionary Algorithm	101
6.3	Results	102
7	CONCLUSIONS AND FUTURE WORK	106
7.1	Conclusions	106
7.2	Future Work	108
Bibliography		111
Bibliography		118

A APPENDIX: TEMPORAL PREDICTIONS OF H1N1 IN USA	119
A.1 Predictions regarding Influenza Like Illness	120
A.2 Temporal Prediction Models	122
A.2.1 Data and Statistics for ILI Predictions	122
A.2.2 Temporal Prediction Model Description	123
A.3 Influenza Like Illness Predictions	124
A.3.1 Time Series Model	124
A.3.2 Neural Network Model	127
A.3.3 Comparison of Time Series and Neural Network Predictions	127
A.4 Mortality Predictions	129
B Tables	132

List of Figures

3.1	Relative risk of infection represented using the transmission kernel as a function of Euclidean Distance.	15
3.2	Iterative minimization of Root Mean Squared Error by Nelder Mead Optimization	16
3.3	Receiver Operating Characteristic curve for determining the optimal parameter set for model W for the three data sets. The optimal parameter set is denoted by the parameter set corresponding to the position marked by the dark ring on the curve.	18
3.4	Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model W has the best predictions when compared to the others.	24
3.5	Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.	25
3.6	Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model W has the best predictions when compared to the others.	26
3.7	Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.	27

3.8 Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model <i>W</i> has the best predictions when compared to the others.	28
3.9 Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.	29
3.10 Foot and mouth disease infection network in Turkey after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [http://kinemage.biochem.duke.edu/software/king.php .]	31
3.11 Foot and mouth disease infection network in Iran after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [http://kinemage.biochem.duke.edu/software/king.php .]	32
3.12 Foot and mouth disease infection network in Thailand after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [http://kinemage.biochem.duke.edu/software/king.php .]	33

3.13 The growth of the size of the Giant connected Component for different values of R_0 using two methods of realization. In the first method infection networks are realized by using a single randomly selected value of p_{th} in [0.01, 0.2] over all possible time instants in the particular data set. In the second method, infection networks are realized by using a randomly selected value of p_{th} in [0.01, 0.2] over a particular randomly selected time instant. Both realizations provide similar estimates regarding the rate of growth of the giant connected component versus the secondary reproduction ratio. Since the slope of both realizations is almost similar, ergodicity of the dynamic infection networks may be assumed.	35
4.1 Map of Turkey depicting 79 administrative districts represented by circular nodes. The temporal evolution of FMDV in each node is studied separately.	40
4.2 Showing a Multi-Layered Neural Network.	43
4.3 Standard deviation and average RMSE for data sets I_t, I_v, S_t	57
4.4 Standard deviation and average RMSE for data sets S_v	58
4.5 Output versus input characteristic feature for Neural Network towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the neural network. The straight line is the ideal output curve.	58
4.6 Neural network predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.	59
4.7 Seasonality of Infection Incidence in Turkey	59
4.8 AIC and MSE for AR(1), AR(2) and AR(3) for all 79 nodes in Turkey.	60
4.9 AIC and MSE for AR(4) and AR(5) for all 79 nodes in Turkey.	61
4.10 AR predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.	62
4.11 The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the autoregressive model.	62

4.12 Showing a single layer Bayesian network model for predicting the probability of infection/susceptibility. The month type is classified into 2 types according to the peak and off-peak season for the virulence of FMD virus. The 'previous record level' input is the probability in the previous time step. The 'present record level' output is the probability in the current time instant that is predicted by the Bayesian network.	63
4.13 Bayesian network predicted probability of infection or susceptibility fitted against actual data on node ID 52 in Turkey.	63
4.14 Output versus input characteristic feature for Bayesian network towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the Bayesian network.	64
4.15 MCS predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.	64
4.16 Output versus input characteristic feature for MCS towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the Monte-Carlo Simulations.	65
4.17 The average of RMSE between data set S and data set I for 79 nodes for 24 monthly time instants.	66
4.18 Spatio-temporal versus temporal predictions for number of animals infected monthly.	67
 5.1 Ergodic behavior of the rate of growth of Giant connected component versus the highest node degree in the infection network.	72
5.2 Non-Ergodic behavior of the rate of growth of Giant connected component versus the highest betweenness centrality in the infection network.	73
5.3 Non-Ergodic behavior of the rate of growth of Giant connected component versus the lowest closeness centrality in the infection network.	74
5.4 Theoretical mitigation by node removal based on highest node degree.	75

5.5	Total number of infected animals presented against the total cost incurred for the 6 mitigation strategies using learning-based predictive models.	84
5.6	Effectiveness of 6 mitigation strategies using learning-based predictive models. .	85
5.7	Reduction in the total population of infected livestock due to implementation of mitigation strategy 1 adopted in Turkey, Iran and Thailand. The bounds on this reduction elucidate the impact of immediate culls, potent vaccines and strict movement restrictions as well as delayed culls and less strict movement bans. The implementation of the mitigation strategy is assumed from March 2005 and onwards each data set. Delayed implementation shall result in a lesser reduction of the total number of infected animals.	90
5.8	Total number of infected animals presented against the total cost incurred for the respective mitigation strategies.	91
5.9	Cost-effectiveness of 6 mitigation strategies in Turkey, Iran and Thailand	92
5.10	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Turkey. The prediction percentage error between the baseline models and model $W(\Delta sMPAE)$ is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.	93
5.11	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Turkey.	94
6.1	Reduction in the Total Number of Infected animals due to mitigation strategies without and with hierarchical dominance.	102
6.2	Optimization of f_{NI} , f_{DC} , f_{NC} and f_{VC} without hierarchical dominance	103
6.3	Optimization of f_{NI} , f_{DC} , f_{NC} and f_{VC} with hierarchical dominance	104

A.1	Fluctuations in the National ILI from week 40 in the year 2008 to week 46 in the year 2009. The national baseline of ILI% is 3.3. Increasing trend of ILI is noticed till week 42 followed by decreasing trends till week 46 in the year 2010.	126
A.2	National ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Neural Networks fitted against the real data. The national baseline of ILI% is 3.3. Unlike the Time-series prediction, the seasonality is not captured. Contrastingly, neural networks predict a steep decrease in the slope of ILI following week 42 in October 2009.	128
A.3	Regional ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Time Series predictions for Atlanta. The regional baseline of ILI% is 2.3. .	128
A.4	National ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Neural Networks fitted against the real data at Atlanta. The national baseline of ILI% is 2.3. This illustrates the dynamic fluctuations captured by neural networks.	129
A.5	Regional Mortality due to Pneumonia and Influenza from week 1 in the year 1997 to week 41 in the year 2009 taken from http://www.cdc.gov/mmwr/ , is collected used to train up a Bayesian Network. Predictions regarding the expected mortality from week 42 in 2009 till week 46 in 2009, and a 95% confidence interval regarding the number of fatal cases has been depicted for a sample case of Kansas City.	130
A.6	Regional Mortality due to Pneumonia and Influenza from week 1 in the year 1997 to week 41 in the year 2009 taken from http://www.cdc.gov/mmwr/ , is collected used to train up a Bayesian Network. Predictions regarding the expected mortality from week 42 in 2009 till week 46 in 2009, and a 95% confidence interval regarding the number of fatal cases has been depicted for a sample case of Denver.	131

B.1	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Iran. The prediction percentage error between the baseline models and model $W(\Delta sMPAE)$ is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.	145
B.2	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Iran.	146
B.3	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Thailand. The prediction percentage error between the baseline models and model $W(\Delta sMPAE)$ is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.	147
B.4	Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Thailand.	148

List of Tables

3.1	Model Parameters	17
3.2	Statistical Analysis of Spatial and Temporal Prediction	22
4.1	Input Probability Level Classification	50
4.2	Output Probability Level Classification	51
4.3	The Conditional Probability Table	51
4.4	Average output probability corresponding to each output level	52
4.5	Comparative analysis of spatio-temporal versus temporal predictions	55
5.1	Sequence of Mitigation Tasks in different Mitigation Strategies.	79
5.2	Cost-effectiveness of mitigation strategies based on learning-based predictive models on data set of Turkey.	83
5.3	Cost-effectiveness of network-based mitigation strategies on data set of Turkey, Iran and Thailand.	86
5.4	Impact of prediction errors using learning-based models	89
6.1	Optimization of 4 objectives	104
B.1	Input Level Classification	132
B.2	Output Level Classification	132
B.3	The Conditional Probability Table with 10 input states	133
B.4	Neural Network Results for Data set I	134
B.5	Neural Network Results for Data set I continued	135
B.6	Neural Network Results for Data set S	136
B.7	Neural Network Results for Data set S continued	137

B.8	AR Model Results for Data set I	138
B.9	AR Model Results for Data set I continued	139
B.10	AR Model Results for Data set S	140
B.11	AR Model Results for Data set S continued	141
B.12	RMSE analysis for best Neural network structure.	142
B.13	RMSE analysis for best Neural network structure continued.	143
B.14	RMSE analysis for best Neural network structure continued.	144

Acknowledgments

Today, as I end my venture towards my Masters degree, I look back to see so many people who stood by me and helped me throughout this journey. I take this opportunity to thank them all with my whole heart for it would not have been a possible journey without each and everyone of them.

I begin with thanking my advisor Dr. Caterina Scoglio for her incessant support, help, and concern. In all, she was one person who made me realize my capabilities as a researcher. I would like to thank my advisor Dr. William Hsu for his constant support and his guidance that was my drive to improve my research capabilities. It would not have been possible to bring his high level of quality into my work without his reassuring inputs. I also wish to thank Dr. Sanjoy Das for believing in me and helping me out with my academic difficulties whenever I hit a wall. It is his confidence in me and his immensely diverse ideas that have built the foundation for my research.

I wish to thank all the group members in the Sunflower Networking group for their help and support. Most importantly, I wish to thank Ali Sydney for his belief in me and his constant help, support and overall his friendship. I want to thank my roommate Lutfa Akter for being more than a friend, for being a guardian, and most importantly for being my academic advisor. Thanks also to my friend Farhana for infusing mirth, joy and faith in me at times I needed the most. Thank you all so much for being there!

I must thank my best friend Shailendra Agarwal for just being there by my side always. I would't be anything without his support. Also, I must thank my constant source of inspiration and support, my brother Shounak Roy Chowdhury, for being such an awesome friend and guide, and for helping me make good decisions in life. I thank my friend Arka Chattopadhyay and Parangiri Nivas for their friendship. I thank the people at Cafe Q for their excellent coffee that was the most important resource for my research! Like Paul Erdos, I aimed at converting coffee to research papers, and I think I succeeded quite a bit!

I thank all people I met in Manhattan ever since Fall 2008 cause they have all made me grow

as a person. At the end of this exciting journey, above all I thank God for being with me always!

Dedication

This thesis is dedicated to my parents, my pillars of success, or should I call them my wishing pillars, cause every time I hug them and wish for something, my wishes come true!

My Father, has been my constant inspiration throughout my academic life. His perseverance, his thirst for knowledge, his zest for life and his unconditional love is the primary reason I have been able to achieve whatever little I have achieved in my life.

My mother is the epitome of faith, honesty, integrity, hard work and overall of unending love and care. She is my best friend, my guide, my best critic and my biggest supporter in life.

These two precious people make me who I am. If god is the creator of life, for me my parents are my god. I love them, worship them and honor them with my whole heart. If only I could dedicate my life to them, I would!

Chapter 1

INTRODUCTION

Foot and Mouth Disease (FMD) is a viral infection that spreads rapidly among cloven-hoofed animals. It causes livestock to get sick, pregnant livestock to abort, dairy livestock to dry up, and even death of the infected animals. This disease poses a global threat due to the rapid rate of spread of the FMD virus that results in massive unethical culls as the common measure to impede the infection spread. Although vaccines have been developed to increase the immunity of livestock against FMD, administration of vaccines is not preferred primarily since the meat of a vaccinated livestock responds similarly to the test of infection as an infected livestock. Loss of meat, dairy and cattle trading privileges are the primary economic losses associated with a country reporting an FMD outbreak besides many other secondary losses involved. Hence, adequate pre-emptive modeling of disease dynamics is necessary to curb economic losses due to FMD.

Research indicates that 80% of the livestock slaughtered during the FMD outbreak in the UK in 2001 on precautionary grounds, were actually not infected [1] [2]. Also, economic losses are magnified by instances such as the FMD outbreak in Taiwan in 1997 that spawned situations wherein farmers intentionally introduced the FMD virus into their flocks since the compensation offered to culled swine was higher than the market value of swine [3], [4], [5]. Thus, mitigation strategies must be optimized to reduce the total number of infected animals and the cost incurred for the implementation of mitigation strategies. Although vaccination against the appropriate strain of FMD virus may be effective in reducing the spread of infection, optimal vaccination strategies will help not only to retard the spread of infection but also to help reduce the cost of

mitigation.

Consequently, several network-based models have been devised to study the underlying dynamics in the spread of FMD, and to suggest optimized mitigation policies for reduced unethical culls and losses [6], [7], [8]. Several spatio-temporal models have been designed to optimize the impact of FMD outbreaks both ethically and economically [9], [10], [11]. Also, significant work has been done in simulating experimental instances of FMD outbreaks, and analyzing the cost of implementing optimal mitigation strategies [12]. Among the models that were developed based on case studies of FMD outbreaks, a few well known models developed during the 2001 FMD outbreak in the UK are the following:

1. Imperial College model from Imperial College, UK: This is a deterministic mathematical model in which the disease is assumed to flow through the population [13], [14], [15],
2. Cambridge Edinburgh combined model from the Cambridge and Edinburgh University: This is a nondeterministic model that simulates the progress of the disease from farm to farm daily. Using the Monte Carlo technique it determines the likelihood that a farm will be infected or not [16], [17], and
3. Veterinary Laboratories Agency (VLA) model from the state Veterinary Labs Agency with colleagues from Massey University, New Zealand: This is a nondeterministic model based on InterSpread, a computer program that can be used to model FMD. It relies on heavily parameterized Monte Carlo method for simulations [18].

Several countries like Turkey, Iran, Afghanistan and Thailand for instance have reported reoccurring instances of FMD over the past decade. However, such countries do not maintain a well developed database regarding the outbreak instances. The primary reason for reoccurring instance of FMD in such countries is the lack of surveillance and inadequate modeling of disease dynamics. In this work, we implement a meta-population based stochastic model to predict the disease dynamics in such countries, and we devise mitigation strategies based on the model to curb ethical and economic losses thus incurred.

We use an S-I-R compartmental model in which individuals are either in a Susceptible ‘S’ state, an Infected ‘I’ state, or a Recovered/Removed ‘R’ state to study the evolution of FMD over a weighted contact network [19]. The weight ages on the links of the contact network are characterized by the human intervention between locations containing infected livestock, wind directions, grazing patterns, heterogeneity in livestock populations and density of farms, feedlots, dairy and meat markets. The rate of transmission of infection between locations (nodes) in time has been modeled stochastically to predict the probability of infection at locations in future time steps. The model is verified using a data set developed by the OIE Incidence Reports regarding the incidence of the FMD virus in farms in Turkey from January 2005 through December 2006 and in Iran and Thailand from January 2005 through December 2005. Our study invokes the contribution of wind in the spread of FMD using network based predictive models. We analyze spatio-temporal predictions with and without the impact of wind as a contributor to the spread of FMD virus.

Next, we analyze the spatio-temporal evolution of FMD using learning-based predictive models and compare the predictive performance to the network based predictive models. These models include non linear predictive models such as neural networks, estimators of infection probability distribution function such as autoregressive models, and Bayesian network models that serve as good predictors in the absence of any consistent probability distribution function. Also, Monte-Carlo simulations help to simulatively evaluate the predictive performance of Bayesian networks, and to estimate the 95% confidence intervals for the learning-based predictions thus produced.

Finally, we assess mitigation strategies using simulative approaches for all the network-based and learning-based prediction models. The cost-effectiveness of these mitigation strategies along with the level of impedance offered to the spread of FMD by culling, vaccination and movement ban strategies is studied. Our analysis leads to the conclusions that vaccination strategy is the most cost-effective and sensitive mitigation strategy. Thus implementation of vaccinations and movement ban strategies in the pre-outbreak period is a cost-effective mitigation strategy; however premise culling is imperative to impede the rapid spread of infection after the onset of an epidemic outbreak. We have also resorted to fuzzy-dominance based genetic algorithms to optimize the

mitigation strategies at monthly time instants so as to reduce the economic cost impacts, as well to reduce the pre-emptive unethical culls on precautionary grounds.

Chapter 2 describes the importance of predictive epidemiology and existing models and studies for FMD. This is followed up by chapter 3 that describes the spatio-temporal model that we propose. Following this, chapter 4 describes learning-based models that can be used to predict occurrences of FMD with time and chapter 5 which assesses mitigation strategies based on all spatio-temporal predictive models discussed so far. Chapter 6 explains the use of genetic algorithms to optimize the impact of the mitigation strategies and final conclusions and future work is summarized in chapter 7.

Chapter 2

MATHEMATICAL MODELS FOR PREDICTING FOOD AND MOUTH DISEASE

Livestock diseases have become a global concern owing to their welfare and economic consequences. Notable examples of livestock diseases include classical swine-fever and swine vesicular disease in pigs, bovine spongiform encephalopathy (BSE), and lumpy skin disease to name a few. Additionally, some diseases such as avian influenza have the potential to cause infection in humans, a typical phenomenon also known as zoonosis, thus increasing concern about their prevalence. However, Foot-and-Mouth Disease (FMD) is considered by far the most serious of all these infections owing to its rapid transmission between a wide range of cloven-hoofed livestock species [20]. Its epi-zootic effect is caused by Foot-and-Mouth Disease Virus (FMDV), which is picornavirus, the prototypic member of the Aphthovirus genus in the Picornaviridae family. This virus is a highly variable and transmissible in nature. There are seven FMD serotypes: O, A, C, SAT-1, SAT-2, SAT-3, and Asia-1. These serotypes are mostly specific to a particular region, while the O serotype is the most common.

In general, FMD has a mortality rate of 40-94% in different species of livestock, however there is a morbidity rate of almost 100%, causing blisters on the mouth and feet (hence the name) and a deterioration of livestock health, often leading to a dramatic decline in milk production in dairy cattle and very slow weight gain in other livestock [21]. Additionally, the economic consequences

of an FMD epidemic outbreak within a country are severe, owing to massive reduction in the export of meat and milk to many regions of the world, thus eliminating a vital source of revenue. Therefore, following the detection of an FMD outbreak, the goal of any control policy is to obtain the disease-free status as quickly as possible, with the minimum of impact on the livestock community. Unfortunately, these two objectives of minimizing economic impacts and epidemic eradication are often in conflict. Striking a balance between these two objectives is a critical decision that must be taken by the appropriate administrative authorities and policy makers since, without the economic guidance, it is meaningless to talk about optimal control strategies.

Past outbreaks of FMD show that unplanned culling may result in massive unethical culling actions being taken as a precautionary measure. Recent studies estimate that 80% of the livestock slaughtered during the FMD outbreak in the UK in 2001 on precautionary grounds, were in fact not infected [22]. Furthermore, reported economic losses are exaggerated in some instances, such as the FMD outbreak in Taiwan in 1997 that spawned situations wherein farmers intentionally introduced the FMD virus into their flocks since the compensation offered to culled swine was higher than the market value of swine [22]. Thus, to combat the severe consequences of FMD, accurate data collection and analysis is imperative for an economically-effective decision making process. For all the three aspects of data collection, statistical analysis and computer-based simulations, close collaborations are necessary to facilitate the best scientific and economic approach to eliminate FMD. Well developed surveillance and infection incidence databases can thus aid the implementation of optimal mitigation strategies [23].

After accurate data collection, mathematical modeling and computer simulated analysis play a very crucial role in disease prediction and prevention as they help to trace several unknown entities that contribute to the rapid spread of the FMDV. For example the entities responsible for the 2001 FMD outbreak in UK are: unknown origin and date of viral introduction, undefined viral characteristics, weather conditions that favored virus survival, potential spread prior to identification of the index case by, animals, people, equipment and vehicles, possible airborne transmission, and involvement of numerous animals of different species and susceptibilities in widely varying

regions of the UK [23].

Contrarily, it is argued by Taylor that “tactical decision making should be based more on real veterinary intelligence than on predictive modeling [24]. Thus, mathematical models are neither the final verdict for economic decision making, nor are they the best perfect cure to epidemics. The shortfalls of mathematical modeling can be assessed with a example of the Contiguous Premise (CP) /Dangerous Contact (DC) culling policy in UK in 2001 wherein livestock at dangerous contiguous contacts were pre-emptively culled within 48 hours of infection reports. This policy was later questioned due to the absence of “precise rationale for the target of 48 hours” [23]. Delayed reporting, incompliance with mitigation strategies and economic concerns expedite the spread of infection. In spite of all these shortcomings, mathematical analysis retains its novelty and importance for coining mitigation policies with an example of the 3-km pre-emptive slaughter, also known as ring culling and ring vaccination strategies that were developed to mitigate the impact of the FMD outbreak in UK in 2001.

Consequently, several spatio-temporal models have been designed to study the underlying dynamics in the spread of FMD and to optimize the impact of FMD outbreaks both ethically and economically [10] [7]. Also, significant work has been done in simulating experimental instances of FMD outbreaks, and analyzing the cost of implementing optimal mitigation strategies [12]. The primary concerns of a few well established FMD prediction models are the following.

1. Interspread Model: Founded in the early 1990s, this model is a large, complex and very flexible stochastic simulation model that can predict infection incidence by considering as many as 50 different global parameters. It is intuitive that complex models are frequently considered better and more accurate. However, it is noteworthy that any model is only as good as the data that is used for parameterization, and since complex models require more parameters, the impact of errors would complicate the prediction process. Hence, this model suffered from over-parameterization in spite of providing qualitatively agreeable simulative predictions.
2. Cambridge Edinburgh model: This is a explicit spatial model, which is initialized with the

location of all farms in the UK and their livestock, recorded at the latest census. Transmission is modeled as a simple function of the disease between infected and susceptible farm, compounded by the number and species of animals on each farm. However, this model supported DC/CP culls, methods which were later deemed as questionable in their impact on impeding the spread of the FMDV. Additionally, this model is a simplified version of Interspread model and contains only a few parameters.

3. Imperial Model: It was based on differential equation based S-I-R models. The initial versions of this model were very crude and under parameterized as they did not consider the heterogeneities between farms containing different species of livestock. However, several modifications have been made to this model till date. This model also sacrifices many of the details present in InterSpread for the ability to rapidly and robustly to parameterize the equations from the epidemic incidence data.
4. Simulative economic models: Several mathematical models have been developed after the 2001 outbreak in UK to account for the economic impacts of FMD outbreaks [9], [12], [25]. Since economic costs of an epidemic are multi-faceted, more research is required before all the elements are well understood.

Thus, the implementation of minimally parameterized prediction models that are geographically independent and yet considerably accurate in their prediction capabilities is desired. This is the challenge that we aimed to address in our work.

The first step towards optimal parameterization of a prediction model is the identification of the prime causes to the spread of the disease virus. Movement of infected animals is one of the most important method of spread of FMD from one property to another [13]. However, on occasions movement of airborne virus particles by wind has been responsible for infecting properties some distance downwind [26]. Under favorable climatic conditions wind-borne spread can be an important factor in FMD epidemics [27]. Analysis of weather data has shows that long-distance spread of virus particles requires stable atmospheric conditions and low wind speeds, factors which con-

tributed to long-distance disease propagation in Australia [27]. However, for infection to occur downwind, animals must be exposed to sufficient virus particles. This depends on the amount of virus produced and the volume of air breathed by exposed animals. Thus, every species of livestock has a different responsiveness to disease transmission and susceptibility.

Additionally, the risk of spread is proportional to the density of livestock downwind, with large concentrations of animals such as sale yards and feedlots being particularly vulnerable. Cattle are more likely to be infected than are sheep or pigs because of their higher respiratory volume sheep have one quarter, and pigs one twelfth, the risk of cattle. Once one animal in a herd has become infected, the disease will spread rapidly through the herd by close contact. All these factors help us in identifying the primarily important parameters for the mathematical modeling of FMD that has been presented in Chapter 3.

Finally we acknowledge that economic risks are higher for countries with un-developed databases regarding the impact of FMD, since such countries report recurrence of FMD in absence of adequate epidemic analysis or economic mitigation policies. Our work aims at studying in the subsequent chapters, the underlying transmission parameters in the spread of FMD to devise mitigation strategies for such countries with an example of Turkey, Iran and Thailand.

Chapter 3

PROPOSED NETWORK-BASED MODEL

We implement a stochastic compartmental model based on a weighted contact network to map the dynamic spread of FMD independent of any geographic location [19]. This flexibility of epidemic prediction in space and time is achieved by the weights on the links of the weighted contact network that represent the heterogeneities in space and the parameterization of the proposed model characterizes the dynamic spread of infection spatio-temporally.

The initial realization of a weighted contact network is achieved by defining nodes as geographical locations defined using the latitude and longitude of regions housing livestock that can be potentially infected or that can contribute to the spread of the FMD virus. For our case studies of FMD in Turkey from January 2005 through December 2006, Iran and Thailand from January 2005 through December 2005, we define a node as a single administration division in the respective countries. Due to the lack of detailed data, we assume each node to be equally susceptible to impending infection and every node has an equal responsiveness to mitigation strategies. The weighted links connecting the nodes of our meta-population based model are characterized using the following factors:

- Human interactions between the various locations based on the rate of thoroughfare between locations and the regional population densities [28], and
- Regional wind speed and direction [29].

However, it is important to note that our meta-population based stochastic model assumes that all

individuals in the same node have the same state at a certain time instant.

Our model assesses the risk of wind borne spread of the FMDV over distances greater than 10 km. The human interactions between locations are modeled using the gravity model for meta-population models. The impact of the wind vector projected onto the distance vector between nodes is an additional term to fine tune the weightages on links. Long distance movements of livestock due to their relocation from farms, to feedlots, to slaughter houses and to markets are a vital parameter that heavily impacts the spread of FMD. However, we do not consider this parameter for the construction of the weighted contact network owing to the lack of data bases regarding long distance movements.

Next, we characterize parameters that impact the spread of FMD infection. For example, different species of animals, such as cattle, sheep, and swine, react differently to the FMD virus, and thus the rate of infectivity varies with the type of susceptible flock [21]. Additionally, different grazing patterns at various locations add to the heterogeneities of the infection network [30], [28]. Thus, the type of livestock, the fraction of flocks of different species grazing away from, and grazing into a specific region, and the distribution of slaughter houses and meat markets in a particular geographical location characterize the parameters in our stochastic prediction model.

At discrete time steps, our model generates a probability of infection for every geographical location (node). If this probability of infection at any time step for a particular node is high, it is predicted to be infected. On the contrary, a low probability of infection refers to a node not being infected but remaining susceptible. Thus, at any time step t we define three sets of nodes N , $N_i(t)$ and N_s as the following:

$$\begin{aligned} N &= \{x : x \in \text{All nodes}\} \\ N_{inf}(t) &= \{y : y \in \text{Infected nodes}\} \\ N_s(t) &= \{v : v \in \text{Susceptible nodes}\} \end{aligned} \tag{3.1}$$

The node sets N , $N_{inf}(t)$ and $N_s(t)$ are related as: $N_{inf}(t), N_s(t) \subseteq N$.

Initially, at time $t = 0$ all nodes are equally susceptible and no node is infected.

Thus, $N_{inf}(t) = \{\phi\}$, $N_s(t) = N$.

After the infection is seeded at time $t = 1$, $N_{inf}(t)$ and $N_s(t)$ change at each following time step, thus affecting the probability of infection at each node. At any given time step t , the probability ($\zeta_{t,i}$) that a node i , does not receive infections from infectious neighbors j , is given as the following [31], [32]:

$$\zeta_{t,i} = \left\{ \prod_{j \in N_{inf}(t)} \left(1 - \frac{\omega_{i,j} \beta_{i,j} p_{t-1,j}}{\omega_{i,j} \beta_{i,j} + \delta_j(t)}\right) \right\} \forall i \in N \quad (3.2)$$

For any node i , the second term in parentheses is the probability that at time step t , the node i receives infection from an infected node j . On subtracting this probability from unity we obtain $\zeta_{t,i}$. It is noteworthy that at any infected node j , there is a proportion of infected population (proportional to $\omega_{i,j} \beta_{i,j}$) and a proportion of recovered/removed population (proportional to $\delta_j(t)$). Out of these two subsets of the population the infected node, only the infected population contributes towards the transmission of infection. Thus the effective proportion of population at any infected node contributing to disease transmission is $\frac{\omega_{i,j} \beta_{i,j}}{\omega_{i,j} \beta_{i,j} + \delta_j(t)}$.

- $p_{t,i}$ is the probability of a node i being infected at time t and is given by the following [32]:

$$p_{t,i} = \{1 - (1 - p_{t-1,i})\zeta_{t,i} - \delta_i(t)p_{t-1,i}\} \forall i \in N \quad (3.3)$$

- $\delta_i(t)$ = the rate at which the FMDV is removed from a node i , or the disease cure rate. This value is estimated as the number of livestock removed due to slaughtering at a particular node at a particular time instant t .

- $\omega_{i,j}$ = the weightage associated with a link between node i and j .

- $\beta_{i,j}$ = the rate at which infection is incident on node i due to an infected neighbor j .

As a novel contribution, we parameterize the weights $\omega_{i,j}$ on links of a fully connected contact network using the data on human interventions between locations and the impact of wind.

$$\omega_{i,j} = \left\{ c_m \left[\frac{m_i * m_j}{d_{i,j}} \right]^{-\gamma_m} + c_w \left\| \frac{\vec{D}_{i,j} (\vec{W}_{i,j} \cdot \vec{D}_{i,j})}{|\vec{D}_{i,j}|^2} \right\| \right\} [\forall i, j \in N, i \neq j] \quad (3.4)$$

- m_i = the scaled measure of population density in location i
- m_j = the scaled measure of population density in location j
- $d_{i,j}$ = the scalar distance between node i and j
- $\vec{D}_{i,j}$ = the vector corresponding to the separation between node i and j in terms of the latitudinal and longitudinal coordinates.
- $\vec{W}_{i,j}$ = the vector corresponding to the velocity of wind blowing between node i and j in terms of the latitudinal and longitudinal coordinates.
- c_m, c_w = the constants due to human interventions and wind respectively.
- γ_m = the exponent to represent scale-free nature of human interventions [33].

Here, the first term represents the nature of human intervention between two nodes using the gravity model. The extent of human interventions between two locations is estimated by the population densities and the distance between the two locations applied using the power law [34]. The second term represents the magnitude of the wind vector projected onto the separation vector between node i and node j . However, the wind data [29] is an incomplete dataset with missing data. Thus, we had to estimate the wind at each node to the closest location in the data base.

Next we parameterize the FMD transmission characteristics $\beta_{i,j}$ that maps the flow of infection through the contact network.

$$\begin{aligned} \beta_{i,j} &= \{(S_i \sigma_i n_i M_i)(T_j \tau_j n_j M_j)K(d_{i,j})\} \\ &[\forall i \in N, j \in N_{inf}(t), i \neq j] \end{aligned} \quad (3.5)$$

- S_i = susceptibility of livestock in location i
- σ_i = fraction of the infected livestock grazing into the location i .
- n_i = number of susceptible livestock in location i .
- M_i = weightage associated with node i due to the presence of grazing farms and slaughter houses.
- T_j = transmissibility of infected livestock in neighborhood j .
- τ_j = fraction of the infected livestock grazing away from the location j .
- n_j = number of infected livestock in location j .
- M_j = weightage associated with location j due presence of grazing farms and slaughter houses. Missing data was estimated using weightage associated with the number of meat markets.
- $K(d_{i,j})$ = Kernel function based on the distance between location i and j .

Thus, $\beta_{i,j}$ depicts the rate of transmission of infection from node j to node i due to the grazing movements of infected livestock in location j into locations with susceptible livestock. A non-linear kernel function characterizes the relative risk of infection with varying Euclidean distance between nodes (3.6). We assume that the relative spread of infection decreases exponentially with increasing distance which is parallel to the general assumption that FMD infection at a particular location affects regions within a 10 km radius as shown in Figure 3.1. The parameter c_j is a multiplicative constant that depicts the rate of infection incidence in node i due to infected neighbor j , such that $c_j = 1$ under normal conditions of infection spread without mitigation strategies. Additionally, the parameters a and γ are known as the kernel offset and kernel exponent respectively which are estimated from the data set during the model calibration phase.

$$K(d_{i,j}) = c_j \left\{ 1 + \frac{d_{i,j}}{a} \right\}^{-\gamma} \quad (3.6)$$

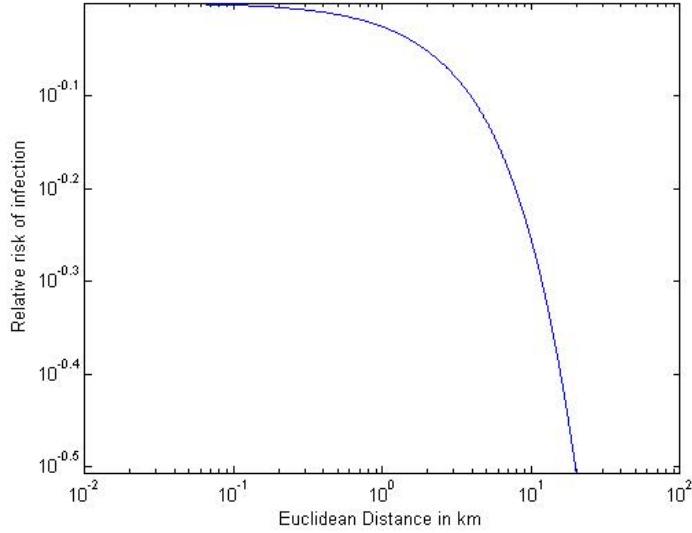


Figure 3.1: Relative risk of infection represented using the transmission kernel as a function of Euclidean Distance.

3.1 Model Assessment

Predictive performance of our model is assessed in comparison to three other network based models. The four network based models under analysis are defined as:

- W = The proposed Weighted, Parameterized model.
- NW = Weighted, Parameterized model without considering the impact of wind.
- $B1$ = Unweighted, Parameterized Baseline model with $\omega_{i,j}$ in eqn. 3.4 constant.
- $B2$ = Unweighted, Unparameterized Baseline model with $\omega_{i,j}$ in eqn. 3.4 and parameters $S_i, \sigma_i, T_j, \tau_j$ and M_j in eqn 3.5 constant.

We assess the contribution of wind in the spread of FMD by comparing the predictability of model W against model NW . Besides, the contribution of a weighted underlying contact network and disease transmission parameters that are novel to our model are assessed by comparing model W with baseline model $B1$ and $B2$ respectively.

3.1.1 Model Calibration

The parameters specific to the network based models need to be estimated so as to fit the data set of Turkey, Iran and Thailand accurately. For calibration purpose, we consider 60% of the data set, which is FMD infection incidence reports for a 15 month period from January 2005 March 2006, as training data. For optimal parameter set estimation of the four network based models, we perform the Nelder Mead optimization, wherein we minimize the root mean squared error (RMSE) iteratively between the true monthly probability distribution of infection and the simulated probability ($p_{i,t}$) across all nodes as shown in Figure 3.2.

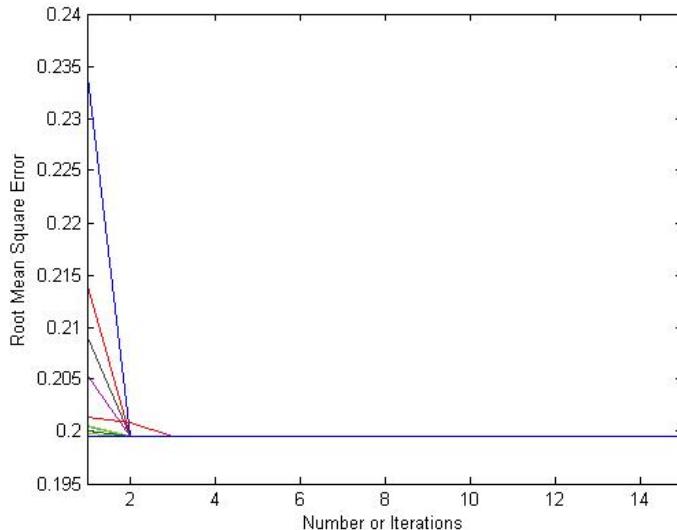


Figure 3.2: Iterative minimization of Root Mean Squared Error by Nelder Mead Optimization

The model W requires optimized values for its set of five parameters ($c_m, \gamma_m, c_w, a, \gamma$). While model NW requires optimization of its set of four parameters (c_m, γ_m, a, γ) and baseline models $B1, B2$ require an optimized set of 2 parameters (a, γ) respectively. All 50 parameter sets were simulated for 100 instances of randomly selected threshold probabilities (p_{th}) in the range [0,1], to compute the spatial prediction performance of each network based model. A threshold probability (p_{th}) is defined such that if the probability of infection at a particular node $p_{i,t} \geq p_{th}$, then node i is said to be infected, or else node i is susceptible. Spatial performance is measured in terms

of average sensitivity and specificity of the predictions over the training data set such that the nodes actually infected according to the data set, when predicted to be infected, are marked as true positives (TP). The nodes that are actually uninfected and predicted as uninfected are marked as true negatives (TN). However, the nodes that are actually infected and predicted to be susceptible are false negatives (FN), while the nodes actually susceptible but predicted to be infected are false positives (FP). The statistics of sensitivity and specificity are defined as the following:

$$sensitivity = \frac{TP}{TP + FN} \quad (3.7)$$

$$specificity = \frac{TN}{TN + FP} \quad (3.8)$$

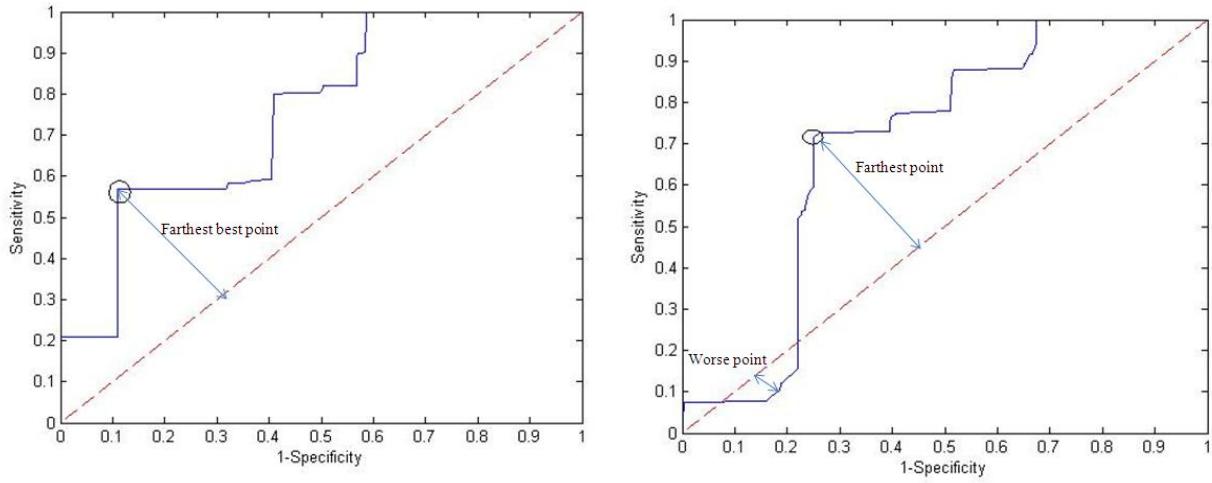
$$(3.9)$$

The optimal parameter set is denoted by the best point marked by the ring in the Receiver Operating Characteristic curve that is plotted to show the performance of sensitivity against specificity for each parameter set as shown in Figure 3.3.

The optimal parameter sets for the calibrated network models are summarized in Table 3.1.

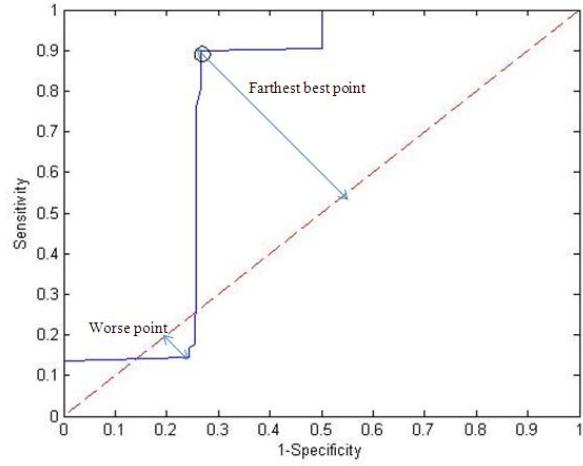
Table 3.1: Model Parameters

Country	Model	c_m	γ_m	c_w	Kernel offset (a)	Kernel exponent (γ)
Turkey	W	0.5709	4.053	0.5642	459.3133	27.338
	NW	0.4056	3.3248	-	215.74	49.486
	B1	-	-	-	98.98	42.272
	B2	-	-	-	170.0467	52.062
Iran	W	4.4276	0.0434	0.6916	489.495	42.662
	NW	4.9144	0.4417	-	398.84	29.64
	B1	-	-	-	219.79	16.124
	B2	-	-	-	16.8227	24.288
Thailand	W	4.0547	0.5588	0.1838	249	61.428
	NW	2.9243	0.2809	-	493.54	36.734
	B1	-	-	-	1742.5	47
	B2	-	-	-	248.215	35.798



(a) Turkey Data set

(b) Iran Data set



(c) Thailand Data set

Figure 3.3: Receiver Operating Characteristic curve for determining the optimal parameter set for model W for the three data sets. The optimal parameter set is denoted by the parameter set corresponding to the position marked by the dark ring on the curve.

3.1.2 Model validation

The calibrated network based infection models are validated for a validation data set that corresponds to the remaining 40% of the data set defined by a period of 9 months from April 2006 through December 2006. Also, an aggregated validation is performed on the entire data set defined by a period of 24 months from January 2005 through December 2006.

Sensitivity and specificity as defined in eqn 3.7 are the primary statistical parameters that characterize the spatial predictions. Additionally, some other spatial prediction statistics are defined as the following:

$$\text{Positive Prediction Value}(PPV) = \frac{TP}{TP + FP} \quad (3.10)$$

$$\text{Negative Prediction Value}(NPV) = \frac{TN}{TN + FN} \quad (3.11)$$

$$\text{Accuracy} = \frac{TP + TN}{TN + TP + FP + FN} \quad (3.12)$$

The ability of a network based model to predict an infected node to be infected is depicted by a high value of sensitivity and PPV. Specificity and NPV denote instances when an uninfected node is correctly predicted to be uninfected. Accuracy of any model refers to the closeness of the prediction from the models to the actual data.

Temporal prediction is assessed using the coefficient of determination (R^2) and symmetric mean absolute percentage error ($sMAPE$) which is an accuracy measure based on percentage (or relative) errors. If Y_i refers to actual data series and X_i refers to the fitted data series, n refers to the number of data points, $STOT$ is defined as the total sum of errors and SSE is the sum squared errors, then $sMAPE$ and R^2 are defined in Equations 3.13 and Equation 3.14 respectively. A higher of value R^2 and a lower value of $sMAPE$ signifies a well fitted model.

$$sMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|Y_i - X_i|}{(Y_i + X_i)} \quad (3.13)$$

$$SSE = \sum_{i=1}^n (Y_i - X_i)^2 \quad (3.14)$$

$$STOT = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (3.15)$$

$$R^2 = 1 - \frac{SSE}{STOT} \quad (3.16)$$

Predictive performance of the models is evaluated in a Bayesian framework using the Deviance Information Criterion (DIC), which is particularly useful in Bayesian model selection problems where the posterior distributions of the parameters sets of the models have been obtained by Markov chain Monte Carlo (MCMC) simulations. Deviance is defined as follows:

$$D(\theta) = -2\log(p(y|\theta)) + C \quad (3.17)$$

- y : the predicted probability of infection
- θ : parameter set
- $p(y|\theta)$: likelihood function. Whenever the outcome of a predictive model is observed by varying the underlying parameter set, the likelihood function is defined in terms of the variable parameter set (θ).
- C : constant that cancels out in all calculations that compare different models, and which therefore does not need to be known.

Next, the measure of goodness of fit is defined as the following:

$$\bar{D} = E^\theta[D(\theta)] \quad (3.18)$$

A smaller value of \bar{D} ensures a well fitted model. Next, the effective number of parameters of the model is computed as following:

$$p_D = \bar{D} - D(\bar{\theta}) \quad (3.19)$$

Here, $\bar{\theta}$ is the expectation of θ . The larger this is, the easier it is for the model to fit the data. Finally, DIC is defined as follows [35]:

$$DIC = p_D + \overline{D} \quad (3.20)$$

Thus, the complexity associated with each of the four predictive models is estimated using DIC . However, it is important to note that the likelihood function regarding the predicted probability of infection for a particular parameter set, if defined with respect to the true probability of infection as in [36], it defines the Divergence distance of the predictive model from the true infection probability distribution. Here, we define the likelihood function as the absolute predicted probability of infection for a particular parameter set such that a lower value of DIC ensures a better model. The statistical results after comparison of the four network models over the entire data set for 24 months and the validation data set of 9 months is summarized in Table 3.2.

Thus, spatial predictions of weighted network based models are clearly better than unweighted models $B1, B2$ by an acceptable margin. However, the accuracy of spatial and temporal predictions from model W and NW are found to be comparable. Additionally, the model W when compared to NW has a higher sensitivity, but a lower PPV. These observations result from the fact that we have considered an incomplete wind data set wherein we the estimated missing data for computation. However, it is noteworthy, that in spite of an incomplete wind data set, model W demonstrates a higher specificity, NPV, R^2 , $sMAPE$ and a lower DIC than all other network based models. Thus, we infer that although the positive prediction of infection improves in the model W when compared to NW , the negative test for infection in both these models is almost similar. Further analysis of the contribution of wind using wind-plume models may enhance the granularity of predictive performance in future.

Additionally, we observe a better spatial and temporal prediction performance for the entire data set as compared to the validation data set. This is true since the model parameters were estimated based on the calibration data set, and thus the validation data set suffers a from greater error as compared to the calibration data set. Also, from the Turkish data set it is evident that the first 12 months from January 2005 through December 2006, are characterized with lesser

Table 3.2: Statistical Analysis of Spatial and Temporal Prediction

Country	Statistics	Entire Data				Validation Data			
		W	NW	B1	B2	W	NW	B1	B2
Turkey	sensitivity	0.711	0.661	0.546	0.377	0.721	0.667	0.554	0.372
	specificity	0.945	0.959	0.706	0.613	0.943	0.969	0.539	0.376
	PPV	0.722	0.749	0.494	0.327	0.878	0.918	0.535	0.361
	NPV	0.876	0.859	0.749	0.633	0.802	0.778	0.583	0.389
	accuracy	0.878	0.872	0.719	0.626	0.847	0.837	0.595	0.449
	R^2	0.993	0.955	0.725	0.492	0.993	0.972	0.629	0.417
	sMAPE	9.391	14.914	21.190	33.181	12.628	16.691	25.513	41.705
	DIC	5896.4	6168.5	7245.2	7567.2	-	-	-	-
Iran	sensitivity	0.776	0.716	0.482	0.275	0.754	0.692	0.447	0.198
	specificity	0.974	0.974	0.924	0.864	0.975	0.977	0.941	0.846
	PPV	0.889	0.883	0.778	0.362	0.905	0.903	0.800	0.291
	NPV	0.937	0.921	0.801	0.802	0.926	0.909	0.769	0.769
	accuracy	0.929	0.915	0.816	0.73	0.922	0.908	0.792	0.690
	R^2	0.964	0.930	0.609	0.443	0.952	0.919	0.559	0.326
	sMAPE	8.393	11.524	22.937	34.167	9.214	15.702	35.538	41.658
	DIC	1953.63	2018.56	2130.6	2301.4	-	-	-	-
Thailand	sensitivity	0.782	0.737	0.567	0.514	0.773	0.691	0.506	0.427
	specificity	0.963	0.947	0.945	0.923	0.982	0.953	0.951	0.921
	PPV	0.852	0.594	0.583	0.406	0.931	0.613	0.604	0.358
	NPV	0.939	0.974	0.945	0.952	0.932	0.971	0.934	0.943
	accuracy	0.929	0.922	0.902	0.888	0.932	0.931	0.898	0.877
	R^2	0.882	0.818	0.472	0.225	0.700	0.487	0.428	0.205
	sMAPE	12.675	21.525	28.893	54.240	22.285	31.065	47.608	81.205
	DIC	1146.7	1467.9	1653.8	1947	-	-	-	-

outbreak reports than the next 12 months from January 2006 through December 2006. Hence, the probability of error in infection prediction is lesser in the initial 12 month period which is a part of the calibration data set. Similar observations for the Iranian and Thai data set show lower prediction errors in the training period owing to the lower number of outbreaks in the pre-outbreak period from January 2005 through June 2005.

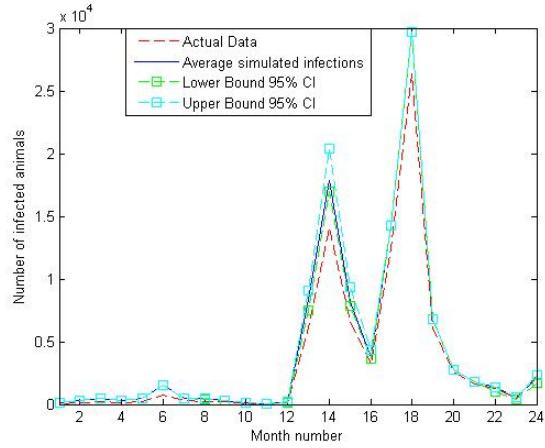
Temporal predictions regarding the number of infected livestock per month are depicted for each network model in Figure 3.4 and Figure 3.5.

The simulations bear evidence to the temporal predictability of all the four network based models. We observe that model W has a tighter confidence interval as compared to the other models. Additionally, weighted network based models provide better temporal predictions than unweighted baseline models. The baseline models $B1$ and $B2$ suffer from over predictions and greater prediction errors.

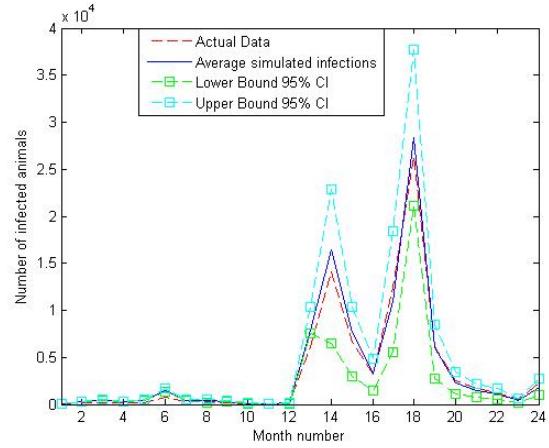
3.2 Simulation of Infection Networks

Having calibrated and validated our model W , we further analyze the FMD infection networks thus produced. We perform Monte-Carlo simulations with randomly selected threshold probabilities to analyze the various infection routes that can be taken during the spread of FMD. An important assumption for our simulations is that the state of complete removal of a node from the infection network is achievable only if either all the herds at that particular node have been culled, or they have all recovered from the infection such that they are immune to acquiring the infection in future time steps. Since our simulation is limited to a specific time period of observation, we do not encounter any node becoming completely recovered or removed from the infection network at any time step.

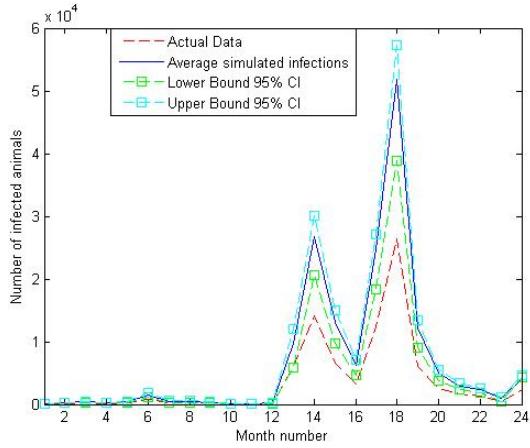
The simulation proceeds as the following: we initialize all the nodes as equally susceptible to receive FMD infection before January 2005. Next, we seed the infection according to the incidence reports [37] in January 2005. Following this, we use our mathematical model to obtain the probability of each node being infected in progressive time steps of one month. If this prob-



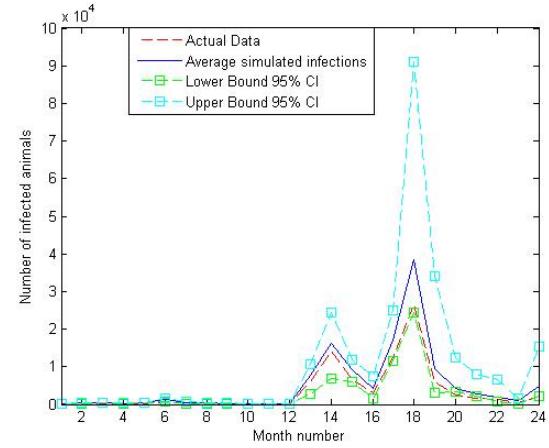
(a) Model W



(b) Model NW

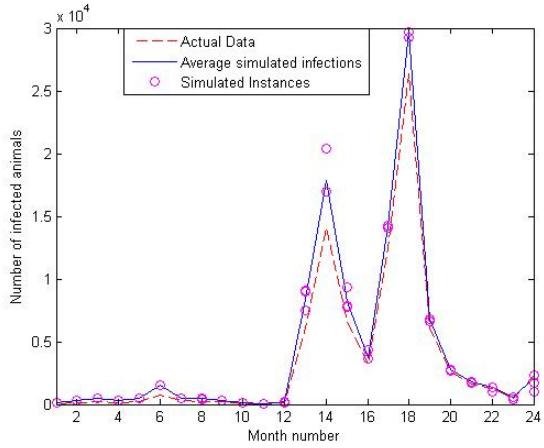


(c) Baseline Model B1

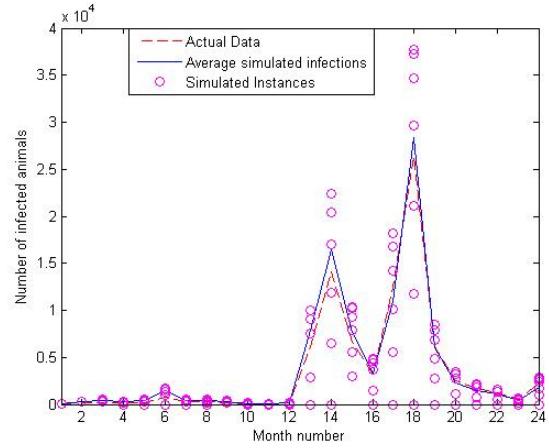


(d) Baseline Model B2

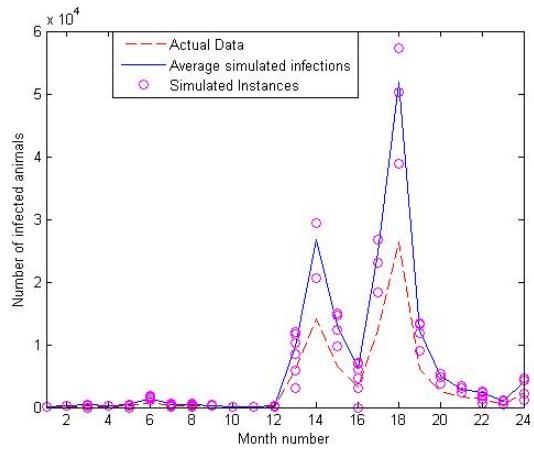
Figure 3.4: Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model W has the best predictions when compared to the others.



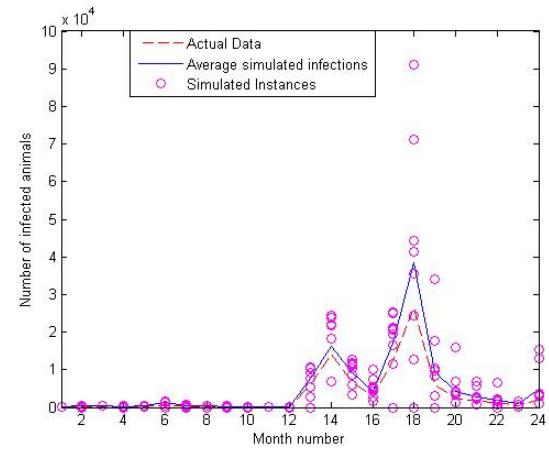
(a) Model W



(b) Model NW



(c) Baseline Model B1



(d) Baseline Model B2

Figure 3.5: Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.

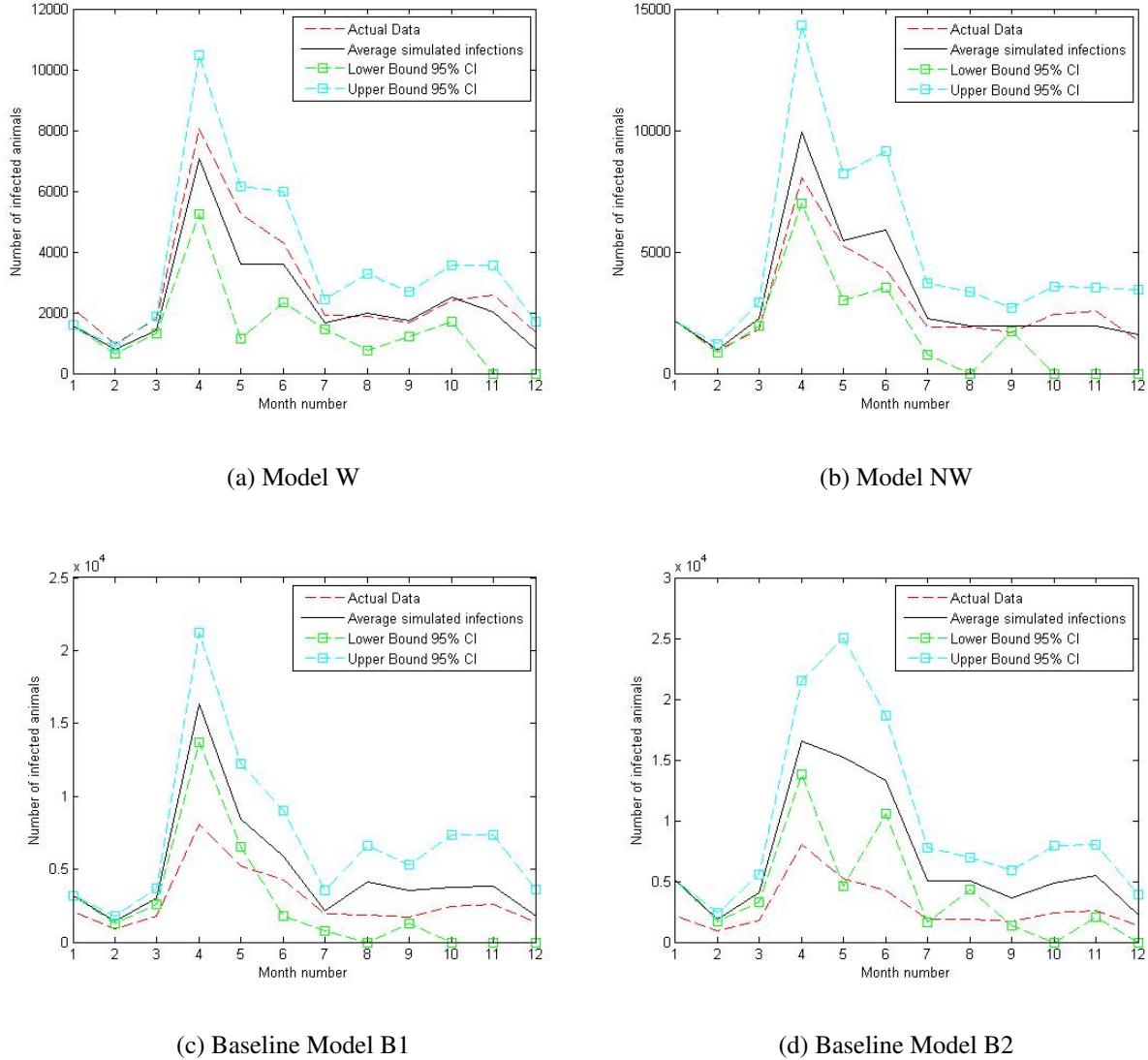


Figure 3.6: Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model W has the best predictions when compared to the others.

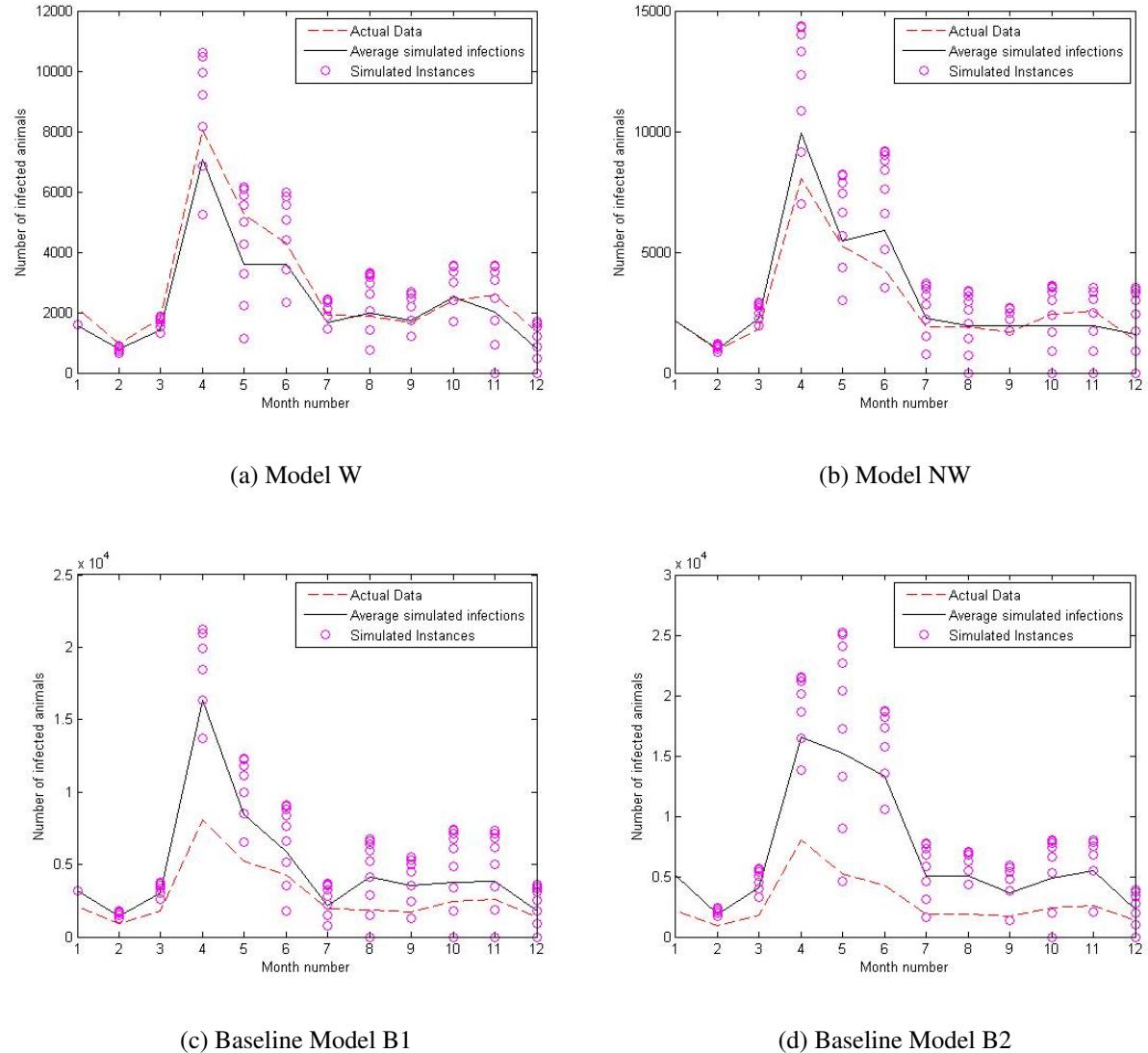


Figure 3.7: Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.

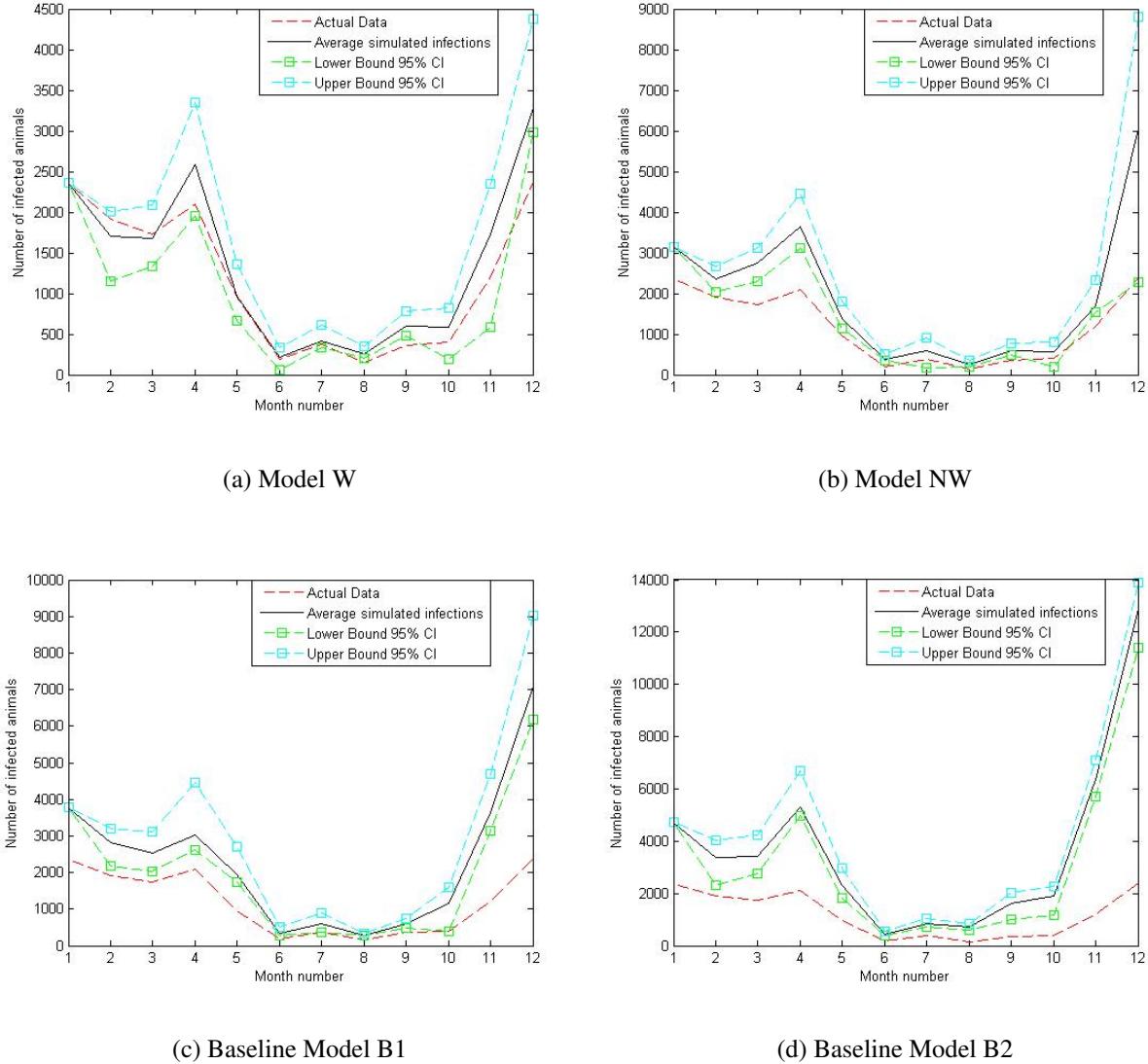
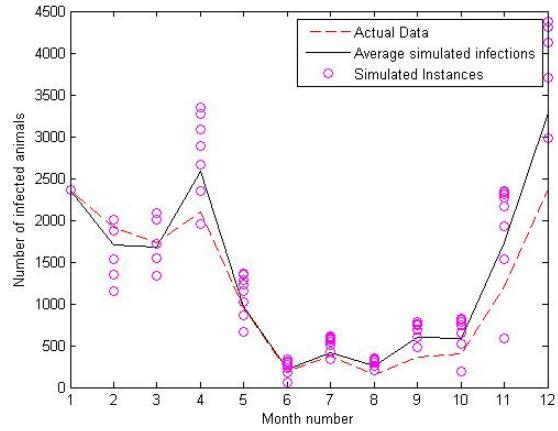
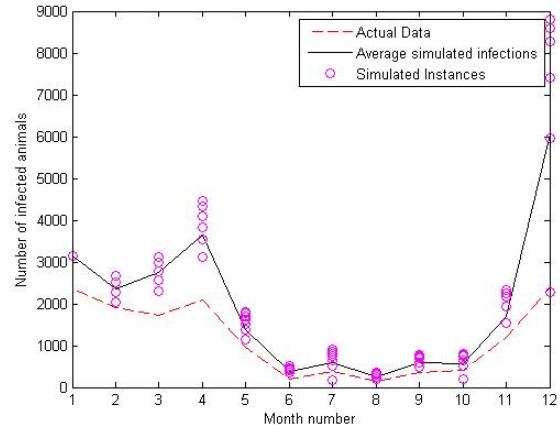


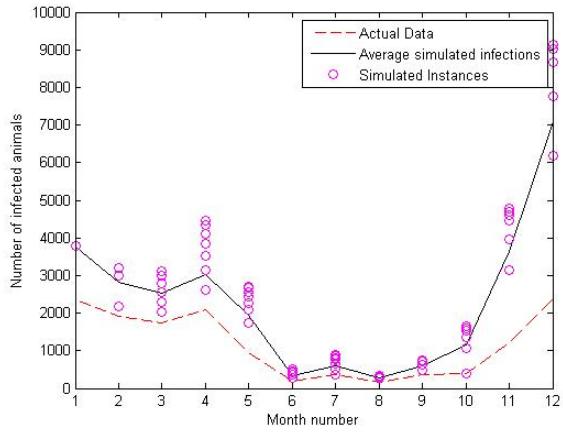
Figure 3.8: Monte Carlo Simulations indicating the 95% confidence intervals for temporal predictions produced by the network based models. The upper and lower confidence bounds along with the mean predictions are plotted against the actual data. Model W has the best predictions when compared to the others.



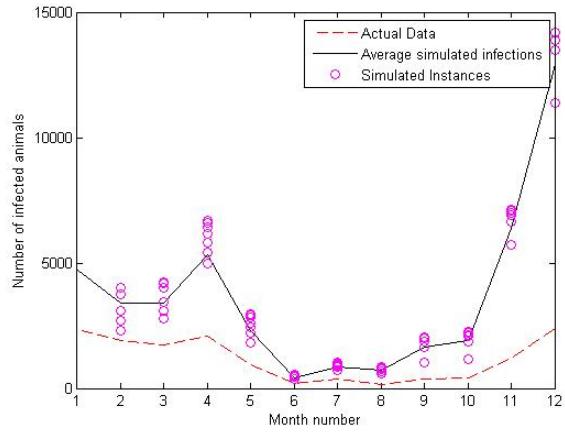
(a) Model W



(b) Model NW



(c) Baseline Model B1



(d) Baseline Model B2

Figure 3.9: Simulated runs corresponding to different randomly selected p_{th} are plotted against the actual data. The mean temporal predictions and 20 simulation runs are plotted out of the 100 simulation runs.

ability is higher than a threshold probability of infection, then that particular node is deemed to be infected. Next, the probability of any link contributing to the spread of FMD ($plink_{i,j}(t)$) is formulated as below:

$$plink_{i,j}(t) = \frac{\omega_{i,j}\beta_{i,j}(t)}{\omega_{i,j}\beta_{i,j}(t) + \delta_i(t)} \left(\frac{p_{t,i} + p_{t,j}}{2} \right) \quad (3.21)$$

Hence, simulated infection networks are realized at discrete time steps. It is however noteworthy that although model W realizes the transmission of FMD virus over a fully connected underlying weighted contact network, the infection networks traced out at various time instants are not fully connected.

Simulations enable visualization of different routes that can be traced out by the FMD infection network. A single realization of such a simulation is shown in Figure 3.10, 3.11 and 3.12. These simulations depict the dynamic growth of the infection network both spatially and temporally. Additionally the infection networks track the direction and velocity of the spread of the FMD virus based on which effective mitigation strategies may be coined.

3.3 Ergodicity of Infection Networks

To ensure realistic network structure of the dynamically growing infection networks, it is essential to assess if the underlying behavior of the network structure remains unchanged over the entire course of the epidemic [38][39]. For the model W , although the underlying weighted contact network is fully connected, the infection networks traced out at different time instants corresponding to different values of p_{th} are not necessarily fully connected. Thus, it would be feasible to construct infection networks over all possible time instants starting January 2005 through December 2006 using the a single threshold probability for determination of infected nodes and links, as well as by randomly selecting different threshold probabilities to realize infection networks in a particular time instant only. The concept of ergodicity in networks states that the mean underlying network feature shall remain unchanged for both such realizations. Hence, different snapshots of the infection networks will give the same estimate regarding the distribution of the size of the giant component growing with a certain reproduction ratio (R_0) defined as follows:

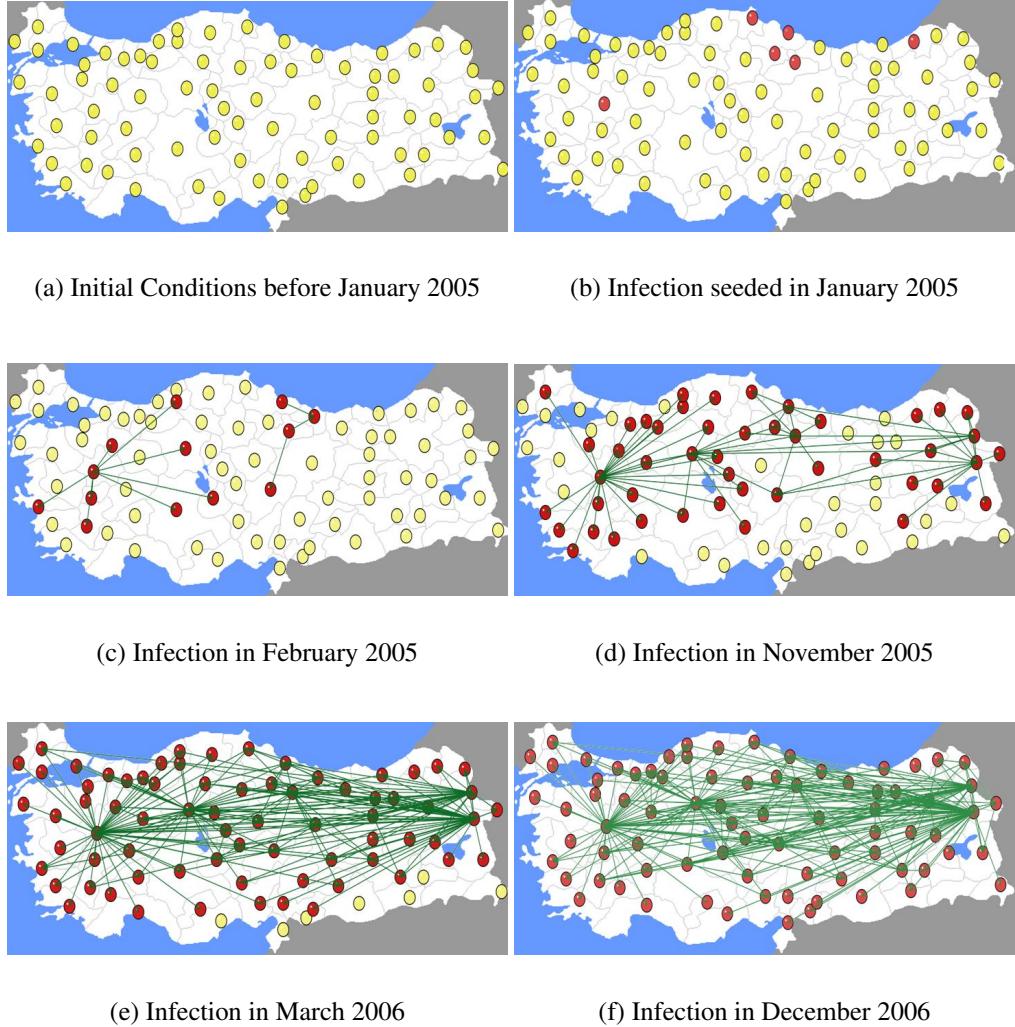
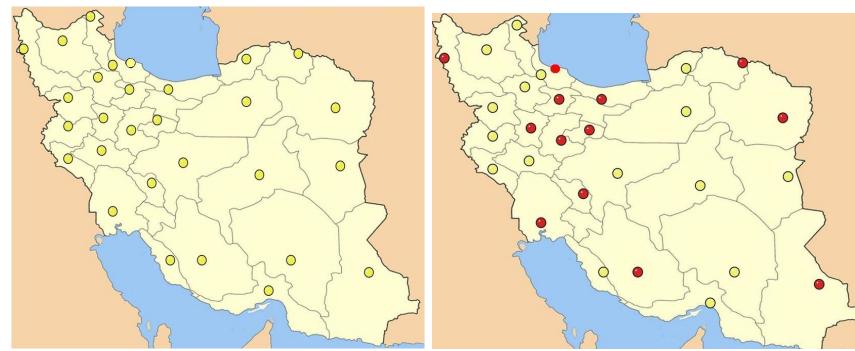
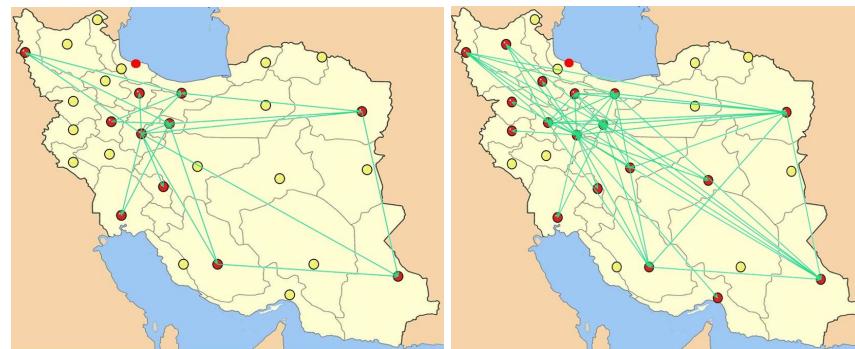


Figure 3.10: Foot and mouth disease infection network in Turkey after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [<http://kinemage.biochem.duke.edu/software/king.php>.]



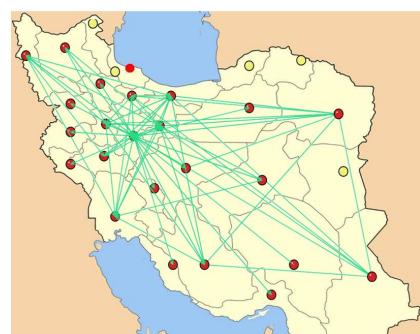
(a) Initial Conditions before January 2005

(b) Infection seeded in January 2005



(c) Infection in February 2005

(d) Infection in June 2005



(e) Infection in December 2006

Figure 3.11: Foot and mouth disease infection network in Iran after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [<http://kinemage.biochem.duke.edu/software/king.php>.]

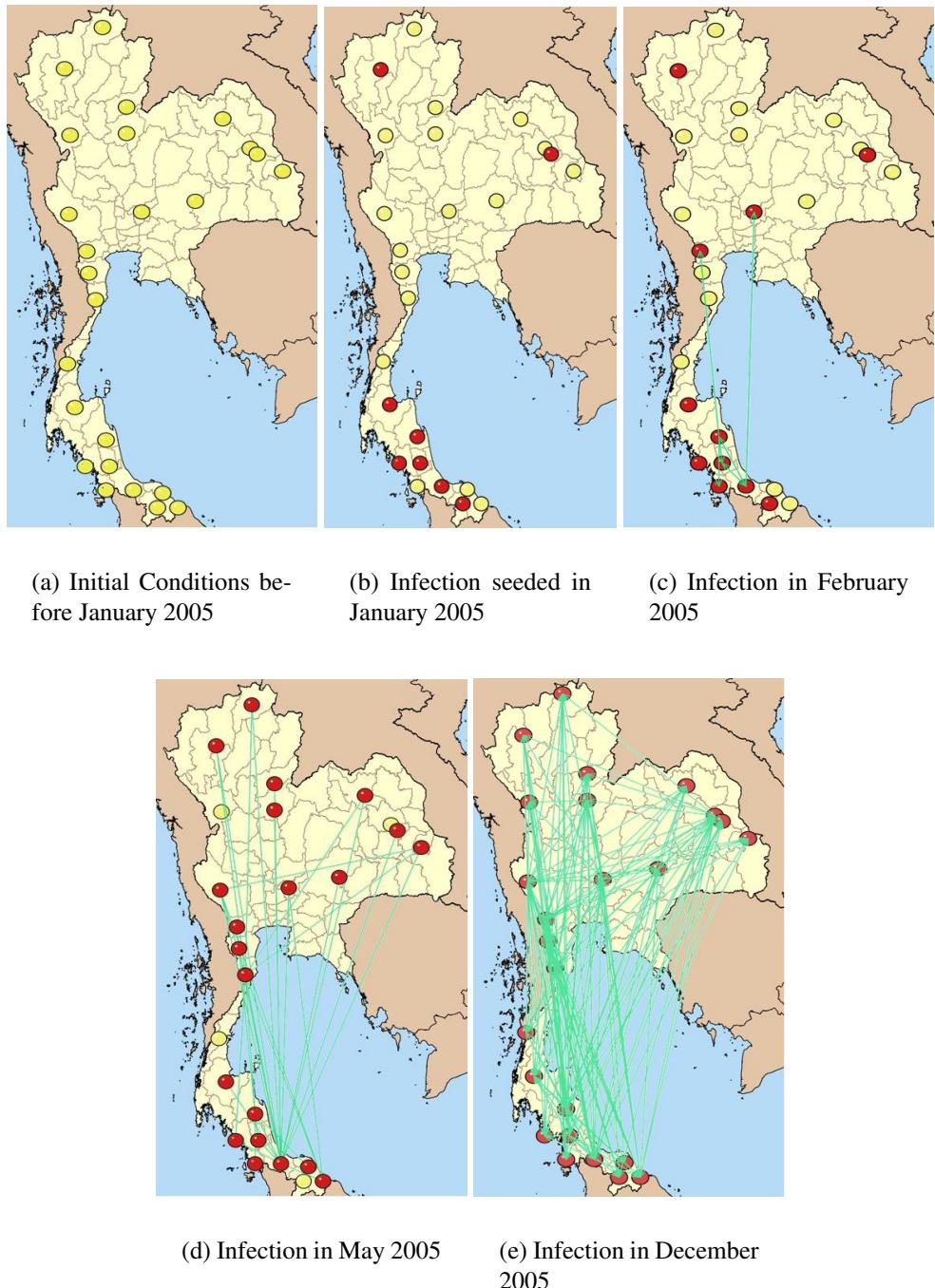
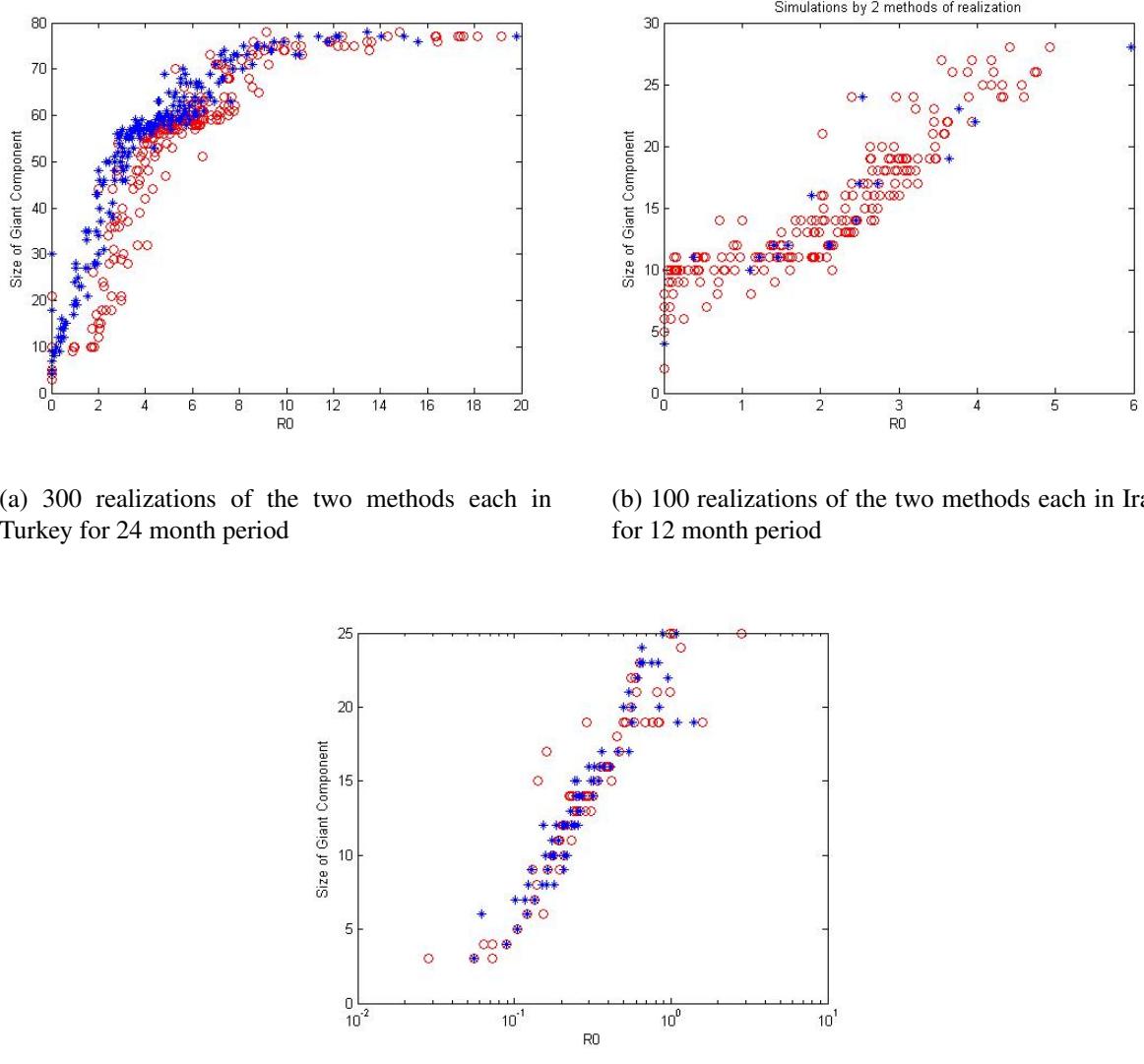


Figure 3.12: Foot and mouth disease infection network in Thailand after seeding the infection in January 2005. The links signify the direction in which FMD spreads from a previously infected node to a recently infected node. Visualization is generated using KING [<http://kinemage.biochem.duke.edu/software/king.php>.]

$$R_0 = \frac{\langle k_{in}k_{out} \rangle}{\langle k_{in} \rangle} \quad (3.22)$$

Here, k_{in} and k_{out} are the node in degree and node out degree for the infections networks. Figure 3.13 depicts the distribution of the size of the giant connected component against (R_0) such that it comprises of 10 realizations of infection networks with a single p_{th} in [0.01, 0.2], simulated over all the monthly time instants and 10 realizations of infection networks formed at a randomly chosen time instant for 10 different randomly chosen p_{th} in [0.01, 0.2]. We choose a small range of values for p_{th} to reduce the range of observations which eases our analysis. For a wider range of p_{th} in [0, 1], we would obtain similar inferences, but over a wider range of possible observations. The inference is that the distribution of the size of the giant connected component versus (R_0) is independent of the method of realization. Hence, any possible realization of infection networks will give similar insights on growth of the giant component. This is an important property to study the viability of mitigation strategies on the infection networks.

Thus, we have defined, calibrated and validated the performance of a weighted contact network predictive model. The comparative validation has been done against network based baseline models. However, there are existing learning-based predictive models that may be applied to predict the probability of infection spatio-temporally and temporally. In the next chapter, we shall study the performance of such learning-based predictive models to further validate the predictive performance of our model.



(a) 300 realizations of the two methods each in Turkey for 24 month period

(b) 100 realizations of the two methods each in Iran for 12 month period

(c) 50 realizations of the two methods each in Thailand for 12 month period

Figure 3.13: The growth of the size of the Giant connected Component for different values of R_0 using two methods of realization. In the first method infection networks are realized by using a single randomly selected value of p_{th} in $[0.01, 0.2]$ over all possible time instants in the particular data set. In the second method, infection networks are realized by using a randomly selected value of p_{th} in $[0.01, 0.2]$ over a particular randomly selected time instant. Both realizations provide similar estimates regarding the rate of growth of the giant connected component versus the secondary reproduction ratio. Since the slope of both realizations is almost similar, ergodicity of the dynamic infection networks may be assumed.

Chapter 4

LEARNING-BASED PREDICTION MODELS

Predictive epidemiology refers to the analytical study of disease dynamics to predict future outbreaks in space and time so that effective mitigation strategies can be implemented to curb the recurrence of epidemics. Since epizootic diseases like the Foot and Mouth Disease (FMD) raise several political, administrative, economic and welfare issues, it is imperative to analyze the disease dynamics to facilitate adequate preventive measures, especially in countries that report recurring epidemic outbreaks instances. Since the FMD outbreak in the United Kingdom in 2001, several analytical spatio-temporal models have been developed to spatially locate such epidemic outbreaks in time [18], [10], [11], [12], [13], [16]. However, it is important to address that spatio-temporal models have parameters of a possibly global structure. Such structures allow region-independence and adaptability of the models by taking information regarding the environment and neighborhood of geographical locations expressed in terms of model parameters. But, in the absence of the sensitive spatial parameters, we may attempt to train a local model on local data with latent parameters to mimic the predictive performance of spatial predictive models. These local models can be learning-based temporal predictive models which, when trained for predictive purposes in a local region, can provide predictions regarding the evolution of infection with time. The novel contribution of this chapter is that we study local information regarding the temporal evolution of infection that is hard-coded in geographical regions, by using learning-based

models that mimic the global parameters of a spatio-temporal model. Additionally, we simulate instances of mitigations strategies to study the cost-effectiveness of culling, vaccination and movement strategies to reduce the total number of infected livestock at the end of a period under study. Also, the utility function to assess the cost-effectiveness of mitigation strategies is defined in terms of the percentage reduction in the total number of infected livestock to the total cost incurred in million US dollars.

Numerous learning-based models have been developed so far to achieve temporal epidemic predictions. For example, neural network models have been argued to effectively model the dynamics of temporal data [40], while time series models have been applied for forecasting the incidences of influenza-like illnesses (ILI) in France [41]. Also, Bayesian networks are useful for reasoning under uncertainty in artificial intelligence which not only detects an outbreak, but also estimates how acute the epidemic is [42], [43], [44]. Regressive models have also been implemented to fit and predict outbreak related data [25] [45]. Additionally, learning-based prediction models have found their importance in predicting wheat leaf wetness [46], [47], soy-rust [48] in plants and critical diseases like influenza [49], malaria [50], [51] and SARS [52] in humans. However, such models have not found any application in prediction of global epizootic epidemics like FMD so far. Our study aims at analyzing the temporal prediction capability of various temporal prediction models and applying them for spatial predictions of FMD epidemic outbreaks in time.

Learning-based predictors are proactive methods for the classification of epidemic severity and for the development of preemptive disease mitigation strategies. They are good tools to analyze infection spread patterns without relying on background spatial information which is generally unknown or estimated. Due to the independence of learning-based models from the background information, they incur a higher prediction error when compared to spatio-temporal predictive models [19]. The reason being that learning-based models require a considerable amount of training data in order to predict well. While an under-trained model produces higher prediction errors, an over-trained model will depict a high variance in its predictions. Such confidential training data in the form of infection incidence reports are often not enough for training such models well. Ad-

ditionally, learning-based predictive models require that the data being used for training and the predicted probability of infection are well correlated. In the absence of correlation, such models fail to provide effective disease predictions. However, it is noteworthy that a trained predictive model can estimate logical bounds to the evolution of disease infectiousness and susceptibility over time [49], [53]. Such bounds can eventually be used for development of mitigation strategies to curb the epidemic impacts thus predicted. Here, we primarily categorize the period under study into two categories. The period wherein the number of reported outbreaks and infected livestock remains at a steady low value is defined as the pre-outbreak period. Contrastingly, the outbreak period is defined as the period when the number of infected livestock increases rapidly to a very large value. The transition period between the pre-outbreak and the outbreak period, is the time when the epidemic sets in. Thus, a well trained model is one which can predict the transition period, and also estimate the total number of infected livestock in the epidemic period under study.

In this chapter, we propose a few learning-based prediction models that can be used to study the temporal evolution of FMD infection and susceptibility at different administrative districts in Turkey. The predictive models when trained for each administrative district separately, can be used to predict the probability of infection and the probability of susceptibility to the FMD virus in future time instants. The different temporal prediction models are: neural networks, autoregressive models and Bayesian networks backed up with Monte-Carlo simulation models. Neural networks are non-linear models to detect sudden fluctuations in infection incidence data and respond accordingly. Contrarily, autoregressive models map the randomness in correlated data by trying to estimate a probability distribution function to fit the time series data in the best possible way. Additionally, Bayesian networks may be applied in instances where no good approximations regarding the probability distribution functions are feasible so that Bayesian estimators based on historical data can be used to predict future occurrences of infection. Also, Monte Carlo simulations help to achieve the gold standards regarding the 95% confidence intervals in the infection and susceptibility probability distribution. Additionally, we evaluate the performance of each prediction model in simulating mitigation strategies, and we eventually justify their predictive performances in terms

of the effectiveness of the various mitigation strategies proposed.

Additionally, we have simulated and analyzed the effectiveness of mitigation strategies for FMD based on the temporal infection predictions in space. From these simulations, we infer that vaccination and movement ban strategies are effective in impeding the spread of the FMD Virus (FMDV) before the onset of an epidemic outbreak, whereas premise culls are important to severely impede the spread of FMDV after an outbreak as set in. In countries which report recurrences of FMD such as the example of Turkey, it may be effective to develop adequate infrastructure to facilitate mass vaccinations for long term disease mitigation.

4.1 Experimental Data

Since learning-based models require considerable data in the training phase, we study FMD predictions using the infection incidence data in Turkey from January 2005 through December 2006. The 24 month period under study is divided into two segments, the training data set, which corresponds to the first 15 months (60% of entire data), and the validation data set, which corresponds to the last 9 months (40% of the entire data set). We study the probability of infection incidence, and the probability of infection susceptibility at 79 administrative districts in Turkey denoted by nodes in Figure 1, assuming that each node is independent and identical in nature. Thus, we compute the probability of infection on data set I and probability of susceptibility on data set S . For the infection data, the training data set is I_t , while the validation data set is I_v . Similarly, for the susceptibility data, the training data set is S_t , while the validation data set is S_v .

4.2 Statistics for model parameterization

We compute spatio-temporal variation in infectivity and susceptibility by studying the temporal evolution in the probability of infection and probability of susceptibility at different regions in space (nodes). Thus, the predictive performance of the learning-based models is analyzed with respect to the actual data such that all prediction errors are probabilistic. The models predict the probability of infection and the probability of susceptibility at each node in progressive time steps

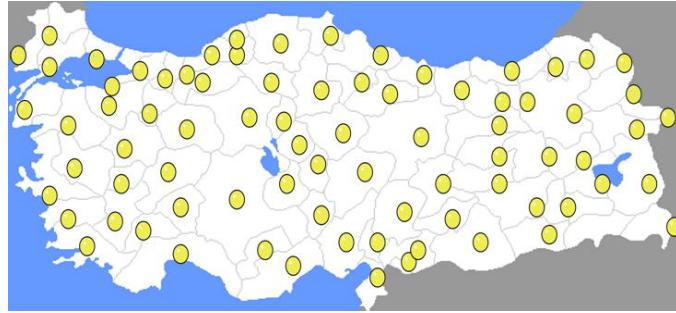


Figure 4.1: Map of Turkey depicting 79 administrative districts represented by circular nodes. The temporal evolution of FMDV in each node is studied separately.

of one month, which is compared to the actual probability of infection/susceptibility calculated by the ratio of infected/susceptible animals at a node at a certain time step over the total number of infected/susceptible animals all over Turkey in that time step.

Predictive performance of each model is analyzed in terms of the Mean Squared Error (MSE) defined as Equation 4.1

$$MSE = \frac{\sum_{i=1}^N (p_{act} - p_{pred})^2}{N} \quad (4.1)$$

- p_{act} = Actual probability from data set. It is calculated by the number of infected or susceptible animals at a particular location at a particular time step over the total number of infected or susceptible animals at all locations at that time step respectively.
- p_{pred} = Probability of infection or susceptibility predicted from the models.
- N = Total number of locations under consideration. In our case, we have 79 locations being studied.

Root mean squared error (RMSE) is the square root of the MSE.

Additionally, the goodness of model fit is defined in terms of $sMAPE$ and R^2 defined in Equation 3.13 and Equation 3.14 respectively. Besides, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) are the measure of goodness for a model such that a model with a lower AIC/BIC has higher likelihood to fit the actual data while incurring a lower computational

complexity due to parameterization. In Krueger et al [54], it is written ‘‘BIC provides a quantitative index of the extent to which each model maximizes correspondence between the observed and model predicted variances and covariances while minimizing the number of parameters. Better fitting models have more negative values, and the difference in BIC values relates to the posterior odds—the odds ratio formed by taking the probability that the second model is correct, given the data, over the probability that the first model is correct given the data. When comparing models, a difference in BIC of 10 corresponds to the odds being 150:1 that the model with the more negative value is the better fitting model and is considered very strong evidence in favor of the model with the more negative BIC value.’’

AIC is the measure of goodness of fit for a model such that a model with a lower AIC has higher likelihood to fit the actual data. If L is the maximized value of likelihood function for the estimated model, k is the number of parameters to be estimated by the model and n is the number of data points, then if we assume the observations to be identically normally distributed, AIC is given below.

$$AIC = 2k - 2\ln(L) \quad (4.2)$$

$$AIC = 2k + n\ln(MSE) \quad (4.3)$$

Similarly, BIC is defined below for normally distributed error with error variance(σ_e^2).

$$BIC = \ln(\sigma_e^2) + \frac{k}{n} \ln(n) \quad (4.4)$$

Kullback Leibler (KL) divergence distance is a method used to measure the goodness of fit by estimating a non-symmetric measure of distance between two distributions, the actual probability distribution (P), and the probability distribution predicted by a model (Q). This method relies on the concept of relative entropy given in the following equation, such that the model with the lowest KL divergence distance is the best fit model.

$$D_{KL}(P||Q) = \sum_i P(i) \frac{P(i)}{Q(i)} \quad (4.5)$$

4.3 Model Description

This section describes the application of neural networks, autoregressive models, Bayesian networks and Monte-Carlo simulations for temporal epidemic predictions. The method for parameter estimation is described, and the predictive performance of each model is also analyzed.

4.3.1 Neural Network Models

Starting with measured data from some known or unknown source, a neural network may be trained to perform classification, estimation, simulation, and prediction of the underlying process generating the data. Hence, neural networks, are software tools designed to estimate correlations in data. A Neural Network(NN) is a set of processing units called neurons and connections with adjustable weights that get modified during the learning process [55]. Usually a NN has a multi-layer structure with one input layer, one or more hidden layers and one output layer. The entire system gets trained by virtue of supervised learning mechanism called *backpropagation*, by utilizing a selective portion of the data called the training data set. During the training process, the weights and biases associated with each neuron are adjusted to minimize the root mean squared error. We apply the gradient-descent algorithm to adjust the weights and biases in each training step. The step following training is validation. In this phase, the *feed forward algorithm* runs on the validation data set and the output thus obtained is matched with the actual output for validating the prediction accuracy of the multi-layered neural network.

While constructing a mathematical model of a biological neuron, usually each weight is defined as w_{ji} , which reflects the connective impact that unit i has on unit j . w_{ji} can have negative or positive values. Negative values imply inhibitory connections between units i and j , while positive values are associated with excitatory connections. Within each unit a summer is present to add all the inputs, multiplied by their respective weights. This sum is denoted as s_j . Furthermore, there is an external bias b_j which is equivalent to a weight applied to a constant input with value of 1. This unit also includes a nonlinear activation function $F_j(\cdot)$. Given an input s_j , the output is

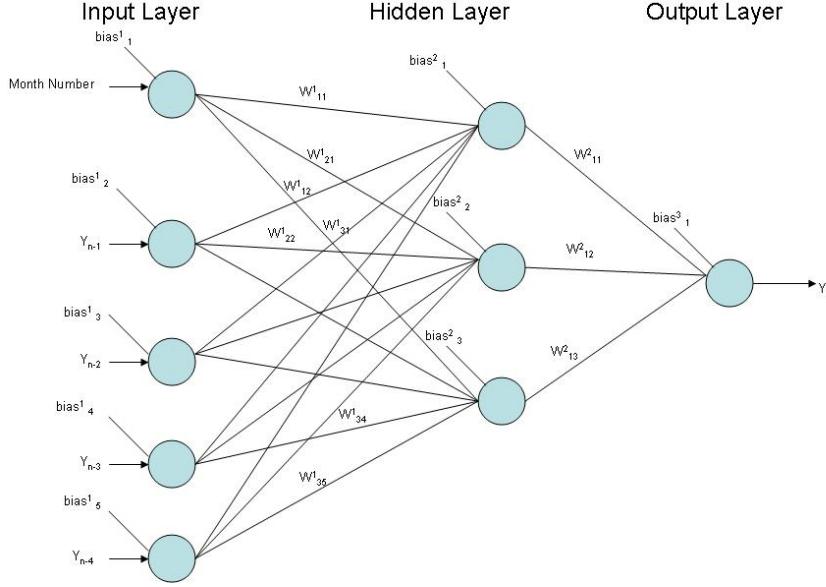


Figure 4.2: Showing a Multi-Layered Neural Network.

defined as:

$$s_j = \sum_{i=1}^{j-1} w_{ji} y_i + b_j \quad (4.6)$$

$$y_j = F_j(s_j) \quad (4.7)$$

$$F_j(s_j) = \frac{1}{1 + e^{-s_j}} \quad (4.8)$$

We implement a three layered neural network with an input layer, hidden layer and output layer using a sigmoidal activation function. The reason for choosing the sigmoidal function is that it is a continuous function that simplifies the backpropagation process. Besides, the derivative of a sigmoidal function can be easily calculated for weight updation using the Chain rule. Also, the inputs are the scaled monthly infection incidence levels at a particular location in previous time instants and the output is the predicted infection incidence at the same location in the next time instant. The input and target data are normalized between 0.1-0.9 although the desired ideal input and output valued range between [0,1], in reality, 0 and 1 are asymptotes to the sigmoidal activation function and hence they must be avoided since to reach asymptotic values, enormous weights and/or input values are required. For a particular set of inputs if t_o is the target output,

y_o is the output produced by the neural network, μ is the learning rate of the gradient descent algorithm, then the error function(E^p) and the corresponding weight updatation rule ($\Delta\omega_{oi}$) using gradient descent is given in Equation 4.9.

$$E^p = \frac{1}{2} \sum_o (t_o^p - y_o^p)^2 \quad (4.9)$$

$$\Delta\omega_{oi} = \mu(t_o - y_o)y_i \quad (4.10)$$

Besides using learning rate, another way to avoid oscillation is to add a portion of past gradient $\Delta\omega_{oi}(t - 1)$ in addition to the current gradient $\Delta\omega_{oi}(t)$ when updating weights according to Equation 4.11 where (t-1) refers to the previous training iteration and α is the momentum factor.

$$\Delta\omega_{oi}(t) = \mu(t_o - y_o)y_i + \alpha\Delta\omega_{oi}(t - 1) \quad (4.11)$$

$$\omega_{oi} = \omega_{oi} + \Delta\omega_{oi}(t) \quad (4.12)$$

Having understood the underlying features of the neural network, the first important task in building a neural network is the selection of the number of neurons in the input layer and the hidden layer. By experience from all the applications of NN, the preference goes to the structure in which there are fewer neurons in the hidden layer than neurons in the input layer. Also, according to Ockham's razor principle, it's more possible that a simpler computing model has better generalization abilities. Additionally, there exists a problem known as the bias/variance dilemma: when the model is too small it will generally be biased, but when the model is too large its parameter estimates can have large variance leading to large variance in the output.

Thus for estimating the best size of the neural network we implement the following set of rules:

1. The neural network is trained using the data sets I_t, S_t and validated using I_v, S_v respectively. Thus the errors for all these 4 data sets are studied separately to find the best neural network structure that minimizes RMSE variation for all the 4 data sets.
2. There will be only one output neuron since we desire single step predictions only.

3. We wish to estimate the size of of a single neural network that shall predict for all the existing nodes. Thus, we spot the node with maximum fluctuations in data S, and I respectively, following which, we estimate the optimal neural network structure to that fits the data of the node with most variations. It is assumed that a neural network which responds to high fluctuations will respond to lesser fluctuations too.
4. We vary the number of neurons in the input layer from 2 through 11, since the pre-outbreak period for the data set of Turkey shows the onset of an epidemic at the end of the twelfth month and hence, to predict an outbreak in time, the pre-outbreak data (comprising of 11 months) must be sufficient to predict the time of onset of an epidemic in the following time step. Also, the neurons in the input layer correspond to the month number and scaled probability of infection/susceptibility in previous time instants. Additionally, due to the lack of data, the size of input layer must be kept minimally small although the input layer cannot have only 1 neuron since a $1 \times 1 \times 1$ neural network would intuitively be biased.
5. It is established that the number of neurons in the hidden layer are at most the number of neurons in the input layer for a minimized network structure.
6. Thus we compare the predictive performance of all possible combinations of neural networks from $2 \times 1 \times 1$ through $11 \times 11 \times 1$. The network with the smallest variance in error and smallest average error is the best network.
7. For computational purpose, we implement the neural networks with momentum 0.18, learning rate 0.5, and 10 generations of 200 iterative cycles for each network.
8. At the end of each 200 iterative cycle the RMSE between the probability of infection/susceptibility and the actual data is computed. The average and standard deviation of the RMSE from the 10 generations is used to determine the best neural network structure.

The result of the above comparison is shown in Table B.12, B.13,B.14 which depict the combinations of input and output neurons and the respective average RMSE and standard deviation

of the RMSE. The respective surface plots are shown in Figure 4.3, 4.4. We look for the network structure with lowest standard deviation in the RMSE of error and considerably low average RMSE such that the network has least number of neurons in it. Thus we infer that the neural network with 4 input neurons and 4 hidden neurons has the smallest deviation in RMSE for a major portion of the iterative process. Thus we select 4x4x1, neural network as the optimal size.

Neural Network Parameterization

Having determined the best size of the predictive neural network, we examine the predictive performance of a 4x4x1 neural network to predict the probability of infection and probability of susceptibility at each 79 administrative district locations. We denote each district location by a unique node ID and we examine the predictive performance of the neural network on data sets I, S separately. For AIC, BIC calculations, it is important to note that the number of parameters to be estimated is equal to the size of S_t, I_t times the number of weights and biases that are updated by each backpropagation. In our case, for a 4x4x1 network, we have 29 such weights and biases. The predictive performances for each node on the data sets are elucidated in Table B.4, B.5,B.6,B.7 respectively.

Finally, the predictive performance on data set I, S are shown in Figure 4.5. Here the input refers to the actual probability in the training data set. The output is the predicted probability for the fitted training data set and the validation data set also. A sample fit for node ID 52 is shown in Figure 4.6.

4.3.2 Autoregressive Models

The second temporal prediction approach is based on time-series analysis of the FMD outbreak. An autoregressive (AR) model is a random process that models the randomness in a natural phenomena such as epidemic spread. These models predict the output at a time instant t based on the outputs of previous time instants by modeling the randomness in correlated data. To ensure correlation of data, we consider the evolution of probability of infection and probability of susceptibility at each node separately, thus ensuring the data to be correlated in space. The first step towards

time-series modeling is to subtract the mean value of data followed by eliminating the seasonality and trend in data. Next, the error in the probability is modeled and hence it becomes possible to fit a probability distribution function at each node corresponding to probability of infection and probability of susceptibility. In our analysis, since the data is sparse and having irregular trends, we analyze sliding windows of varying sizes for prediction. The method for implementing the sliding windows in training and validation phase is described as follows:

- Training phase: For data set I, if the window size is 6, we use the probability of infection at time instants 1 through 6 to predict the probability of infection at time step 7. In the next iteration, we use the actual probability of infection from time instant 2 through 7 to predict the probability at time instant 8 and so on.
- Validation phase: For data set I, if window size is 5, we use the probability of infection at time instants 1 through 5 to predict the probability of infection at time step 6. In the next iteration, we use the previously predicted probability of infection from time instant 2 through 6 to predict the probability at time instant 7 and so on.

The window size and order of autoregressive models with the lowest Akaike Information Criterion and lowest Mean Squared Error (MSE) is selected as heuristically optimal and used for future predictions. Since we wish to minimize parameter estimation, we consider only autoregressive models without the moving average parameters.

$$X(t) = \phi_0 + \sum_{i=1}^{i=p} \phi_i X(t-i) + \epsilon_t \quad (4.13)$$

Here, ϵ_t is additive white Gaussian noise which models the randomness part in the data while $X(t), X(t-i)$ are the probability of infection/susceptibility at time instants $t, t-i$ respectively. ϕ_0 is a constant and p is the order or the autoregressive model. Also, the order of the autoregressive model can be estimated using the Yule-Walker Equations [56]. To find the best data windows size and order of an AR model pertaining to each node separately, we first analyze the seasonality of infection incidence. From Figure 4.7 we see that the seasonality of infection incidence is six

months, i.e. the peak season with high probability of infection lasts for six months. Hence, the order and window size of the AR models must be at most equal to 6 since we wish to reduce the error that will creep into the predicted probability of infection during the transition from a peak season to an off-peak season of infection in Turkey.

Autoregressive Model Parameterization

Hence to estimate the best data window size and order of AR models (p) at each node separately, we perform the following operations:

1. For all combinations of window size from 2,3,4,5,6 and order $p= 1,2,3,4,5$, we compute the AIC, BIC and MSE of the fitted model on data sets S_t, I_t , which are given in Figures 4.8, 4.9.
2. The window size and order corresponding to lowest AIC/BIC is selected as the best parameter choice for that particular node. For the AIC/BIC calculation, it is important to note that the number of parameters to be estimated shall be equal to the total number of coefficients estimated for each node for data sets S,I.
3. The best parameter set at each node is used to fit the data in the training set and predict probability of infection/susceptibility for the validation set. The fitted data is evaluated in Table B.10,B.11,B.8,B.9 respectively.

The goodness of fit is shown in Figure 4.10. The predicted versus the input probability is depicted in Figure 4.11. The predictions have the randomness feature in them to a certain extent.

4.3.3 Bayesian Network Model

A Bayesian network is a directed acyclic graph representing a set of random variables and their conditional or probabilistic relationships. Since FMD outbreaks are random events, in the absence of a specific probability distribution function, they can be successfully modeled using probabilistic methods like Bayesian Networks for temporal predictions regarding the number of infected

livestock at any record time instant. Each random variable in the network is depicted by a node in the graph with connections between the parent and child nodes [55], considering input random variables to be parent nodes and the predicted output to be the child node respectively. A single connection leads a parent (causal) node to a child (influenced) node, thus depicts the conditional dependence between the child and parent nodes [55]. However, if there is no connection between two nodes, it indicates conditional independence. There is a conditional probability table for each child node, which can be computed by the prior probabilities of the parent nodes [55].

A single-layer discrete Bayesian network shown in Figure 4.12, is constructed with two input parent nodes representing the month type and the probability of infection/susceptibility in the previous time step, and one child output node representing the present probability of infection /susceptibility, such that inputs and outputs are classified into discrete levels. Based on the OIE Incidence reports regarding the FMD outbreaks, a conditional probability table and causal relationships are derived for each data set. Bayesian parameter estimation is carried out using Maximum Likelihood Estimation (MLE) and the mean expected output probability is compared with the actual data set. In our analysis, we consider 10 discrete input levels and 10 discrete output levels. Since we aim at magnifying the impact of input probability in the previous time instants on the output probability in the present time instant, thus we consider 5 levels of input probability and 2 types of month classifications ($5 \times 2 = 10$ input levels). The output probability, for the sake of even-ness has 10 discrete probability levels, each corresponding to a mean expected output probability level.

Bayesian Network Parameterization

It is necessary that the classification for the input parent nodes should be done in a way such that all the data points which have similar influences on the child output nodes are grouped into the same level. For the child node, the classification should be done so as to have as many levels as possible, with relevant number of data entries in each level [55]. Classification of the input month type is based on the number of infections reported each month. Now, survival of the FMDV in the environment depends on the initial concentration of virus in the material, the strain of virus,

the humidity, the pH and ambient temperature. Considering the seasonal average number of FMD outbreaks annually in Turkey, we find that reported foot and mouth disease outbreaks in summer (60%) is higher than in winter (40%)[30]. All regions in Turkey have lower incidence in December-February (winter) and September-October (autumn) with incidence rising over summer to mainly a peak in March to July [30]. Thus, we categorize the months for Turkey as the following:

- Month Type 1: Months of January, February, March, October, November, and December in each year. From the Figure 4.7 we observe that during these months, the number of infected livestock is comparatively lower than the rest of the months in that year cycle.
- Month Type 2: Months of April, May, June, July, August, and September in each year. From the Figure 4.7 it is evident that these months have a higher number of infected than the rest of the months in a year.

Next, Classification of previous time instant input probability level (I) is done in Table 4.1:

Table 4.1: Input Probability Level Classification

InputLevel(I)	Probability at time ($t - 1$)
1	< 0.0001
2	0.0001-0.007
3	0.0071-0.04
4	0.041-0.1
5	> 0.1

Following this, the classification of the output level(O) based on the output probability of infection/susceptibility is given by Table 4.2:

The conditional probability table (CPT) which gives the probability of occurrence of each output level given the probability of input in the previous time instant and the month classification type is given by Equation 4.14 and tabulated in Table 4.3:

$$P(\text{Outagellevel} = i | \text{Inputstate} = m) = \frac{N_i}{T_m} \quad (4.14)$$

Table 4.2: Output Probability Level Classification

OutputLevel(O)	Probability at time instant (t)
1	0-0.00002
2	0.000021-0.00005
3	0.000051-0.0014
4	0.00141-0.0035
5	0.00351-0.0080
6	0.0081-0.0133
7	0.01331-0.05
8	0.051-0.1
9	0.1-0.3
10	> 0.3

Table 4.3: The Conditional Probability Table

Output Level	1	2	3	4	5	6	7	8	9	10
Input Level 1	0.9720	0.0089	0.0115	0.0000	0.0000	0.0000	0.0000	0.0038	0.0025	0.0013
Input Level 2	0.0000	0.0000	0.0976	0.4634	0.4390	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 3	0.0000	0.0000	0.0000	0.0000	0.0667	0.3000	0.6333	0.0000	0.0000	0.0000
Input Level 4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3030	0.6970	0.0000	0.0000
Input Level 5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7241	0.2759
Input Level 6	0.9524	0.0253	0.0223	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 7	0.0000	0.0000	0.0978	0.5435	0.3587	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 8	0.0000	0.0000	0.0000	0.0000	0.1167	0.3250	0.5583	0.0000	0.0000	0.0000
Input Level 9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3143	0.6857	0.0000	0.0000
Input Level 10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7241	0.2759

Where,

N_i =Number of occurrences in outage level i .

T_m =Total number of occurrences in input state m .

It is important to note at this stage, that we construct a single CPT using the S_t, I_t of all 79 nodes. This is done since we want the CPT to be as non-sparse as possible and it would not be possible to populate such a CPT for each node individually, due to the lack of data. Thus, we construct the CPT using the training data sets and we validate the performance of the Bayesian network using the validation data sets S_v, I_v . Next, we evaluate the expected output probability

level as the prediction obtained from the CPT, which is based on the expected output for each input level given below.

$$E(N_l | Inputstate = j) = \sum_{k=1}^{10} P(O_k | I_j) \times Average(O_k) \quad (4.15)$$

Where,

N_l :Output Probability

O_k :Outage level=k.

I_j :Input state=j.

The average output probability corresponding to each output level is given in Table 4.4

Table 4.4: Average output probability corresponding to each output level

Output Level	Average Output
1	0.00010500000000000000
2	0.00035000000000000000
3	0.00095000000000000000
4	0.00245000000000000000
5	0.00575000000000000000
6	0.0106666666666667
7	0.0316666666666667
8	0.0750000000000000
9	0.2000000000000000
10	0.5000000000000000

The predictive performance of the Bayesian network for a node ID 52 is shown in Figure 4.13. Additionally, the input probability versus the predicted probability of infection/susceptibility against the ideal prediction curve is shown in Figure 4.14

4.3.4 Monte-Carlo Simulations

Monte-Carlo simulations (MCS) refer to a particular class of algorithms that rely on repeated random sampling to compute results. Since these simulations rely on repeated computation of random events, these methods are most suited to calculation by a computer and tend to be used

when it is unfeasible or impossible to compute an exact result with a deterministic algorithm or to simulatively verify the results obtained by other deterministic algorithms. In the Bayesian network model, we assume that the predicted value for each state is the expected value, which represents a point estimate for the output probability. But since a particular month type, and a particular previous time step probability input level are themselves composed of a number of entities, an input state represents a range of different values of factors, and thus it is only a rough classification of the effects of month and input probability level on the predicted output probability. Thus, the model is expected to have errors in prediction and we should simulatively find a range of values within which the observed numbers of outages are expected to lie.

MCS algorithm

Monte Carlo simulation is a common method to find out the confidence intervals and to simulatively calculate the expected output probability of infection/susceptibility [55]. Thus, the following steps illustrate the method of implementation of MCS.

1. We compute input state for a particular node at a given time step
2. We generate a uniform random number in $[0,1]$.
3. Using roulette wheel selection with this random number, we select an outage level based on conditional probability table.
4. We generate another uniform random number in $[0,1]$.
5. Using roulette wheel selection with this random number, we select a value of output probability which follows uniform distribution within one outage level.
6. The simulation is repeated 1000 times for each monthly time step for each node.
7. At the end of the simulation, we find the average output probability from the 1000 values corresponding to each node and each time step. This average value is compared to that obtained as output from the Bayesian network model.

- For computing the confidence interval, we sort the 1000 output values corresponding to each output probability and we select the upper limit for 95% confidence as the smallest integer X such that the percentage of all the numbers below X exceeds 97.5% of the 1000 data points. The lower limits are assumed to be the biggest integer which makes the percentage of all the number below it smaller than 2.5%.

Finally, we analyze the goodness of fit of the MCS in a sample node in Figure 4.15, and also the input probability versus the output probability comparison plots against the ideal output curve in Figure 4.16.

4.4 Model Analysis

The predictions regarding the probability of infection and the probability of susceptibility obtained using neural networks, autoregressive models, Bayesian network models and Monte-Carlo simulations is analyzed to evaluate the predictive performance of each model. The average of the RMSE for all 79 nodes between the data set S and data set I is shown in Figure 4.17

We can now assume that so far we have performed spatio-temporal predictions regarding the evolution of FMD, since we considered predictions at different administrative districts separately. At the end of these spatio-temporal predictions, the probability of infection at each node, if multiplied with the total number of infected animals in Turkey at any particular time instant shall yield the number of infected animals at that particular time step at the particular node. Thus, it is possible to estimate the number of infected animals in Turkey at a particular time step spatio-temporally. Having analyzed the spatio-temporal prediction of FMD using learning-based models, we further analyze the predictive performance of each method temporally, i.e. when predictions are made regarding the number of animals infected monthly in Turkey. It is important to note that intuitively the temporal predictions should have more error than spatio-temporal predictions owing to the lack of training data and the assumption that all locations in Turkey are correlated in space in the pre-outbreak and in the outbreak period of the epidemic.

For temporal predictions, neural network uses the same 4x4x1 network structure for predicting

the number of infected animals in future time instants. In auto-regressive models we observe that order $p = 1$ and $windowsize = 6$ yield the lowest AIC/BIC values. Bayesian network however has to be reevaluated for temporal predictions using the input level classification in Table B.1, output level classification in Table B.2 and CPT in Table B.3. We observe that the new CPT is more sparse than the spatio-temporal CPT.

The comparative performance of spatio-temporal and temporal predictions thus produced is shown in Figure 4.18 and Table 4.5.

Table 4.5: Comparative analysis of spatio-temporal versus temporal predictions

Model	Spatio-temporal		Temporal		R^2
	sMAPE	R^2	D_{KL}	sMAPE	
NN	11.2164	0.9816	0.0056	23.922	0.9373
BN	16.3127	0.9371	0.0095	12.5603	0.9086
MCS	18.164	0.9387	0.0094	12.6702	0.9106
AR	22.2083	0.9535	0.00833	28.1051	0.8956
MCS upper bound	44.925	0.5071	-	32.5821	0.68
MCS lower bound	30.5148	0.4398	-	38.6698	0.6921

Additionally, on comparing the KL divergence distance of spatio-temporal predictions for the data set S, we see neural network has a divergence distance of 0.0061, autoregressive models is 0.00794, Bayesian networks is 0.0122 and MCS is 0.0120. As an inference we see that neural networks have the lowest KL divergence distance, lowest error and best R^2 statistic. So it provides best estimates spatially and temporally regarding the evolution of FMD infection. Auto-regressive models incur higher error than neural networks, yet they fit the data aggregate and thus they incur a considerably good R^2 value. Bayesian networks on the other hand have a consistent error incurred temporally and spatio-temporally and hence they set good bounds to the extent of infection spread. Also, we see that the confidence intervals set by MCS by the upper bound and lower bound are higher than the predictions obtained by all the four learning based models. Hence, we can say that all the learning-based prediction models have acceptable prediction standards.

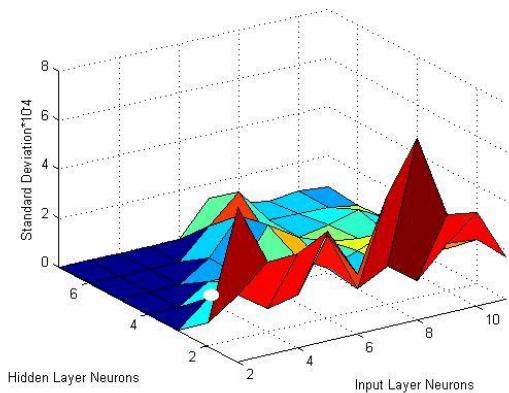
Finally, we compare the AIC and BIC of the learning-based predictive models and we infer that auto-regressive models incur a lower AIC/BIC since they require lower parameter estimation

than neural network models. Hence, if the number of nodes where spatio-temporal evolution is to be studied is large, the computational complexity of neural networks becomes very high and infeasible. Auto-regressive models in such cases provide lower computational complexity at the cost of prediction errors. Bayesian networks have a different scale for computing complexity. This statistic is called BIC_{score} and it is defined in Equation 4.16 which implies that a higher BIC_{score} indicates a better model.

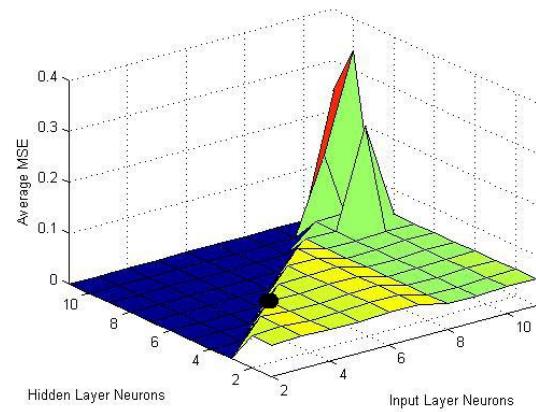
$$BIC_{score} = \sum_i \sum_m (\log_2 \theta_i - \frac{\log_2 M}{2} Dim[i]) \quad (4.16)$$

- i=Node in the Bayesian Network
- m=size of data set
- M= total size of data set
- θ_i = Bayesian estimator given by the CPT
- $Dim[i]$ =number of independent parameters with respect to a particular node. $Dim[G]$ is the total number of independent parameters in the graph G.

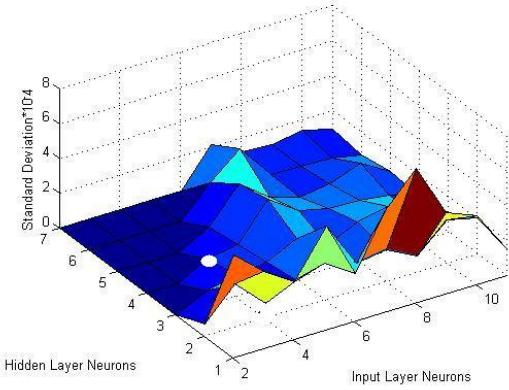
We observe that BIC_{score} for spatio-temporal Bayesian network predictions is -185.3445 while that of temporal predictions is -447.6649. This is indicative that Bayesian networks are better spatio-temporal predictors than temporal predictors. Finally, we observe that although the temporal prediction performance of learning based models is considerable, they are not as good as the network based spatio-temporal models since for the spatio-temporal model W , $sMAPE$ is 9.39 while R^2 is 0.99. Additionally, spatio-temporal models are good predictors in the absence of a large amount of correlated historical data which is otherwise required by learning-based models for predictive purposes. Thus, we may conclude that estimating background parameters related to the spread of epidemics using network based spatio-temporal models is a better method for epidemic prediction as compared to learning-based predictions.



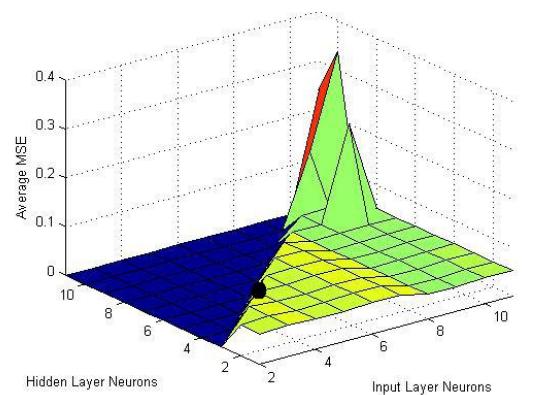
(a) Standard deviation of RMSE for I_t .



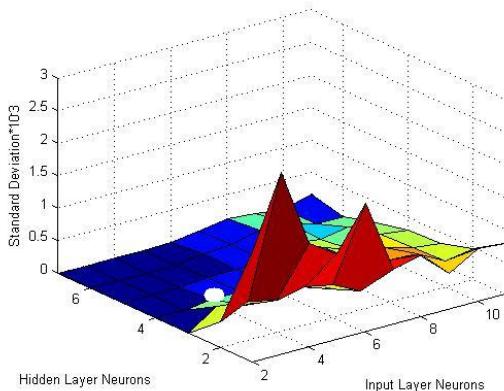
(b) Average RMSE for I_t .



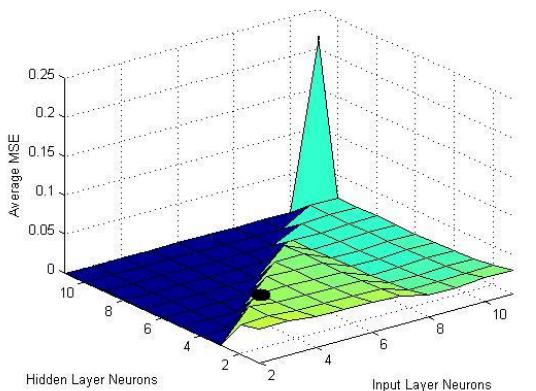
(c) Standard deviation of RMSE for I_v .



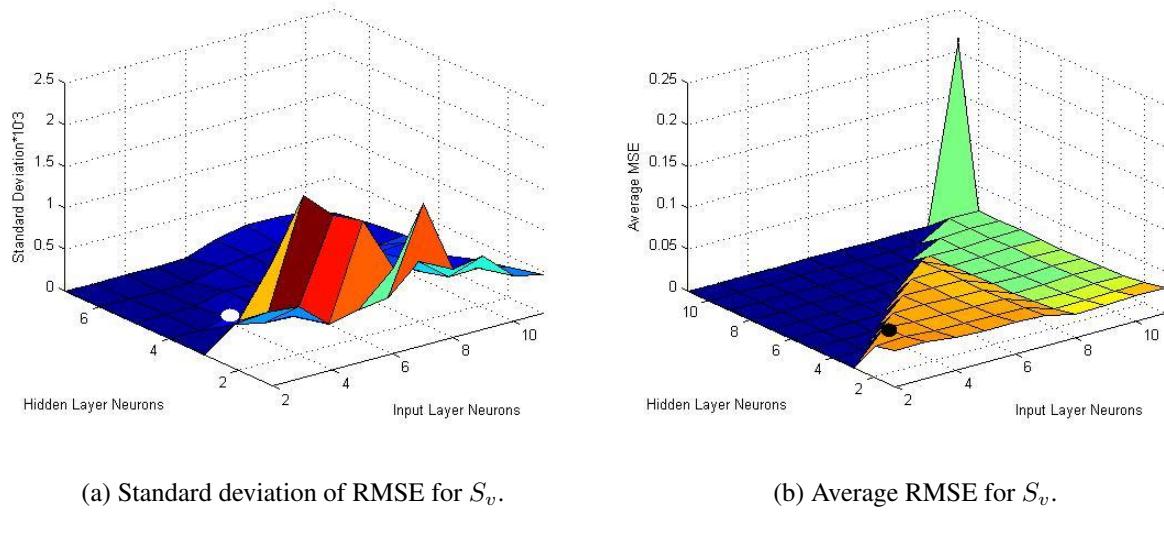
(d) Average RMSE for I_v .



(e) Standard deviation of RMSE for S_t .



(f) Average RMSE for S_t .



(a) Standard deviation of RMSE for S_v . (b) Average RMSE for S_v .

Figure 4.4: Standard deviation and average RMSE for data sets S_v

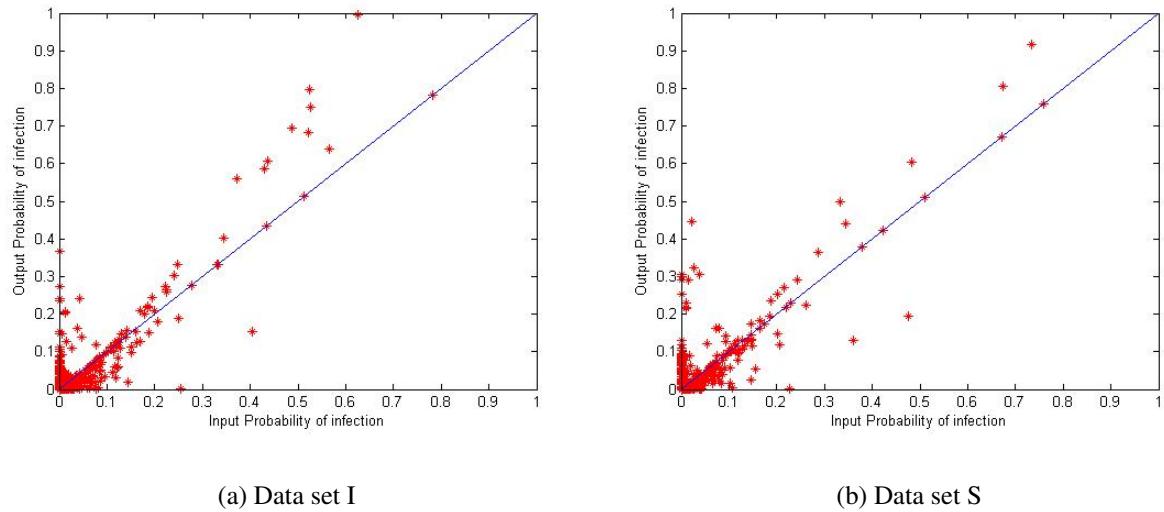


Figure 4.5: Output versus input characteristic feature for Neural Network towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the neural network. The straight line is the ideal output curve.

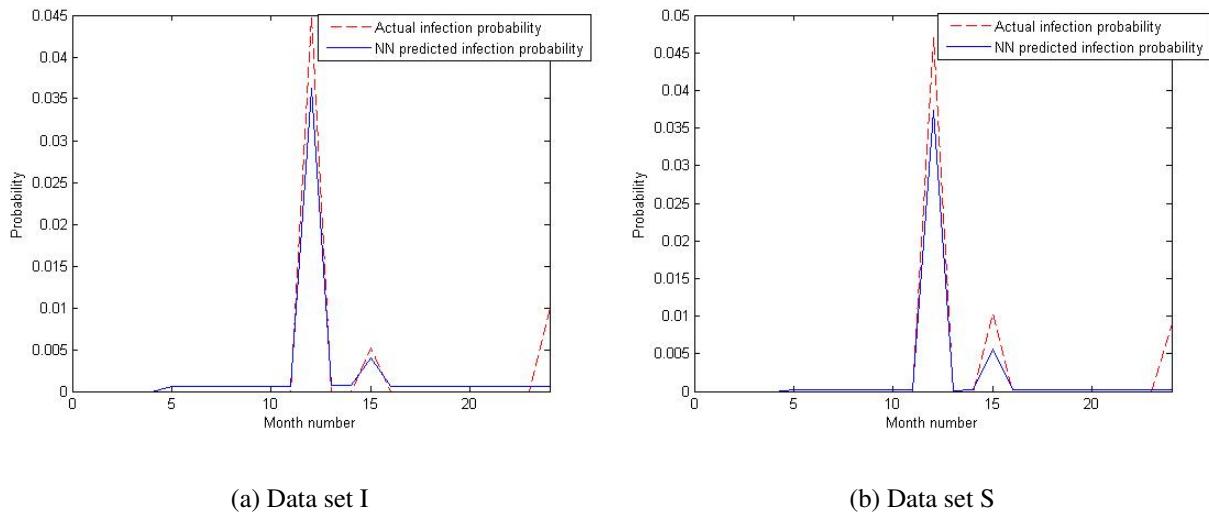


Figure 4.6: Neural network predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.

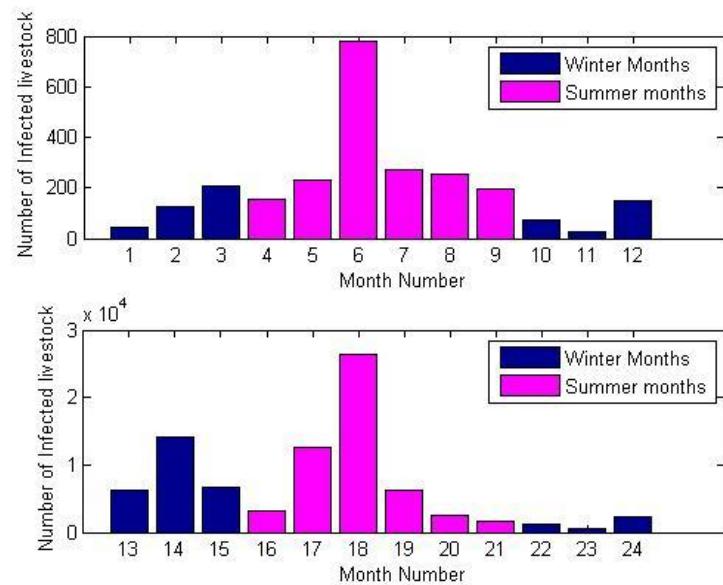
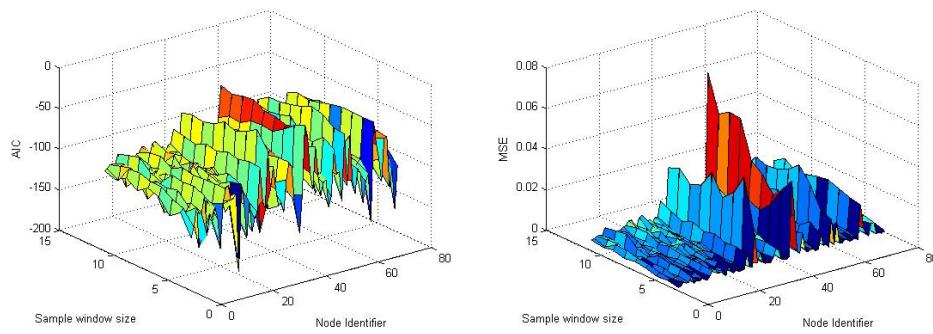
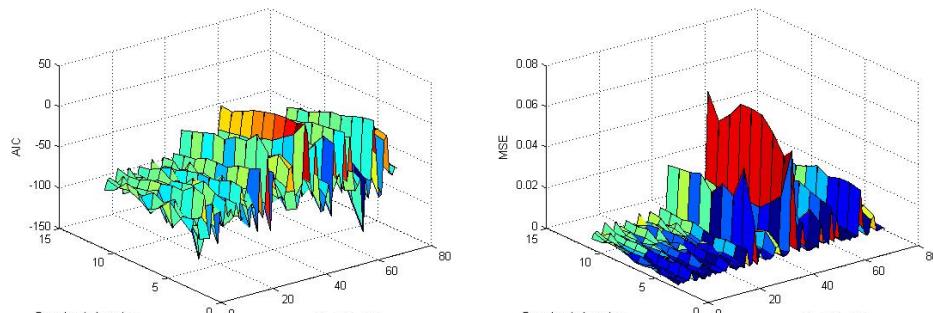


Figure 4.7: Seasonality of Infection Incidence in Turkey



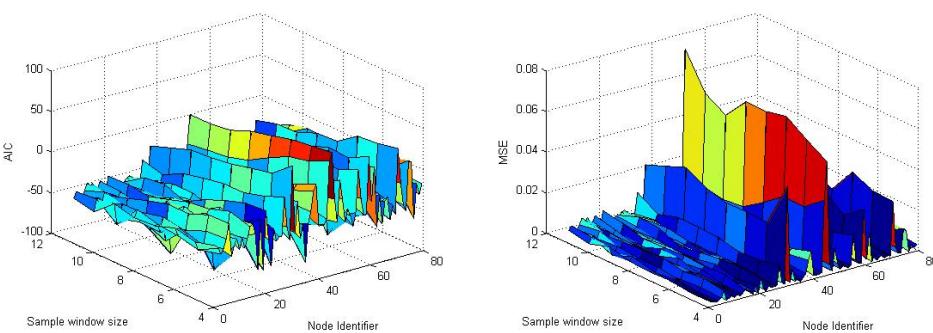
(a) AIC for AR(1)

(b) MSE for AR(1)



(c) AIC for AR(2)

(d) MSE for AR(2)



(e) AIC for AR(3)

(f) MSE for AR(3)

Figure 4.8: AIC and MSE for AR(1), AR(2) and AR(3) for all 79 nodes in Turkey.

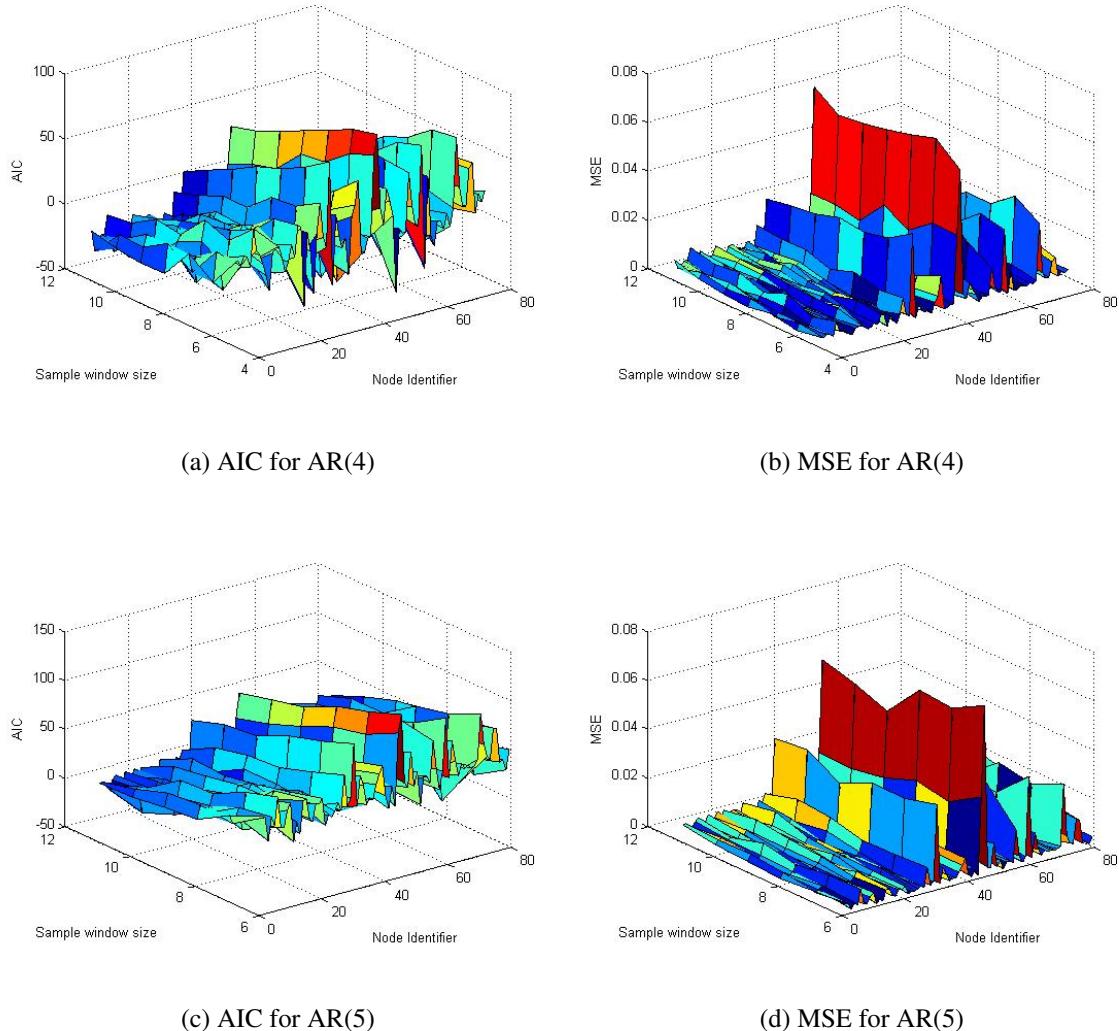


Figure 4.9: AIC and MSE for AR(4) and AR(5) for all 79 nodes in Turkey.

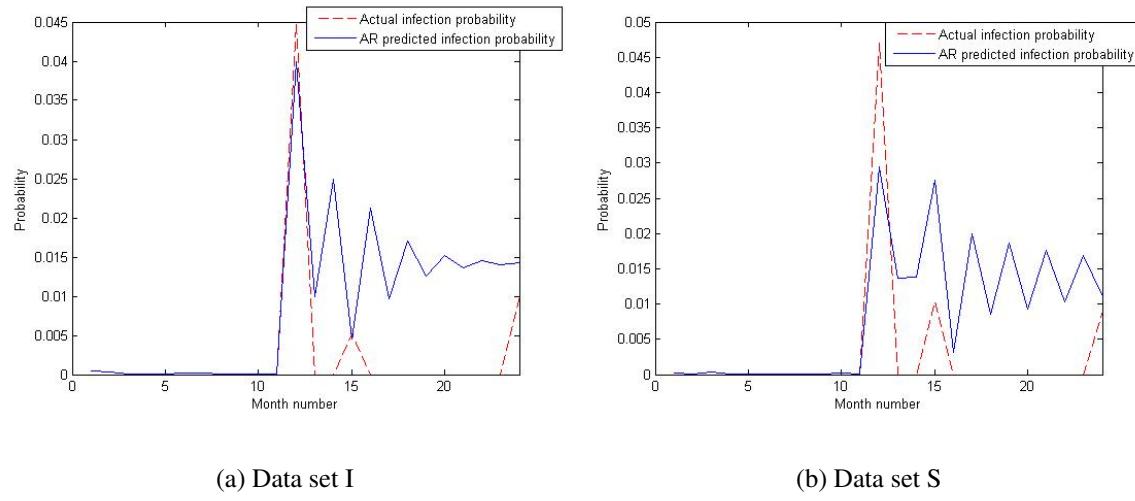


Figure 4.10: AR predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.

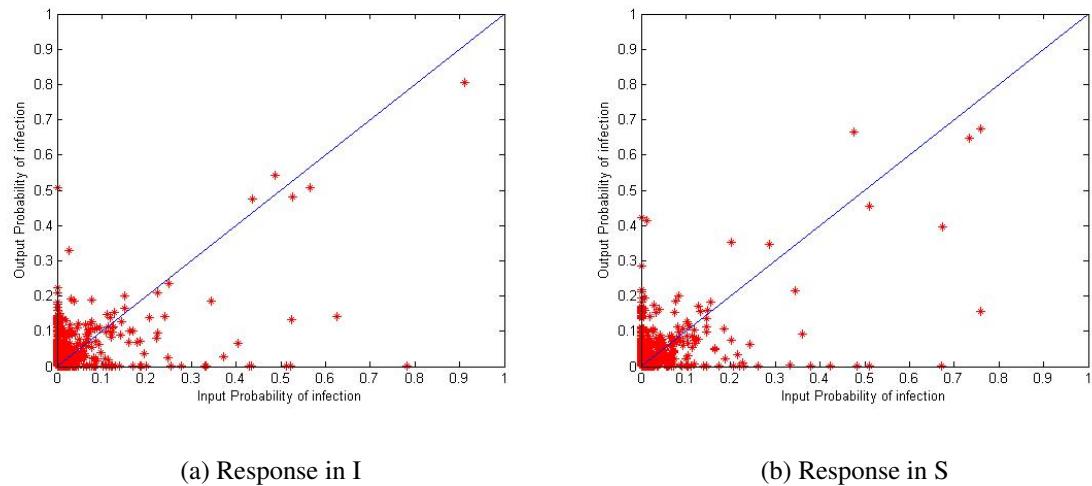


Figure 4.11: The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the autoregressive model.

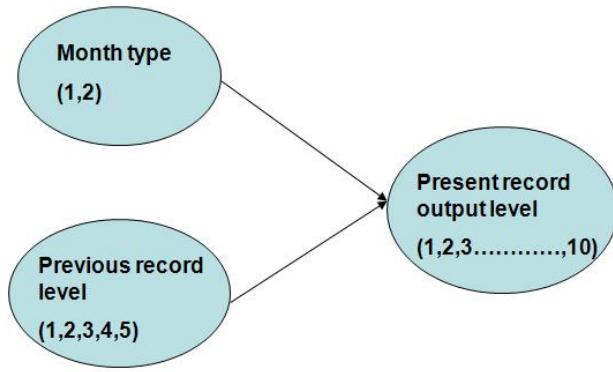


Figure 4.12: Showing a single layer Bayesian network model for predicting the probability of infection/susceptibility. The month type is classified into 2 types according to the peak and off-peak season for the virulence of FMD virus. The 'previous record level' input is the probability in the previous time step. The 'present record level' output is the probability in the current time instant that is predicted by the Bayesian network.

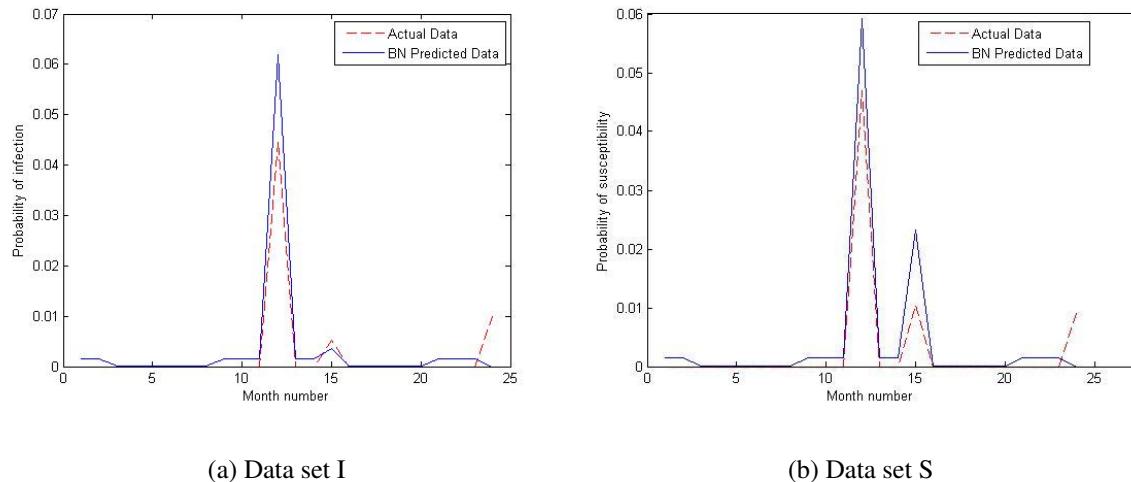


Figure 4.13: Bayesian network predicted probability of infection or susceptibility fitted against actual data on node ID 52 in Turkey.

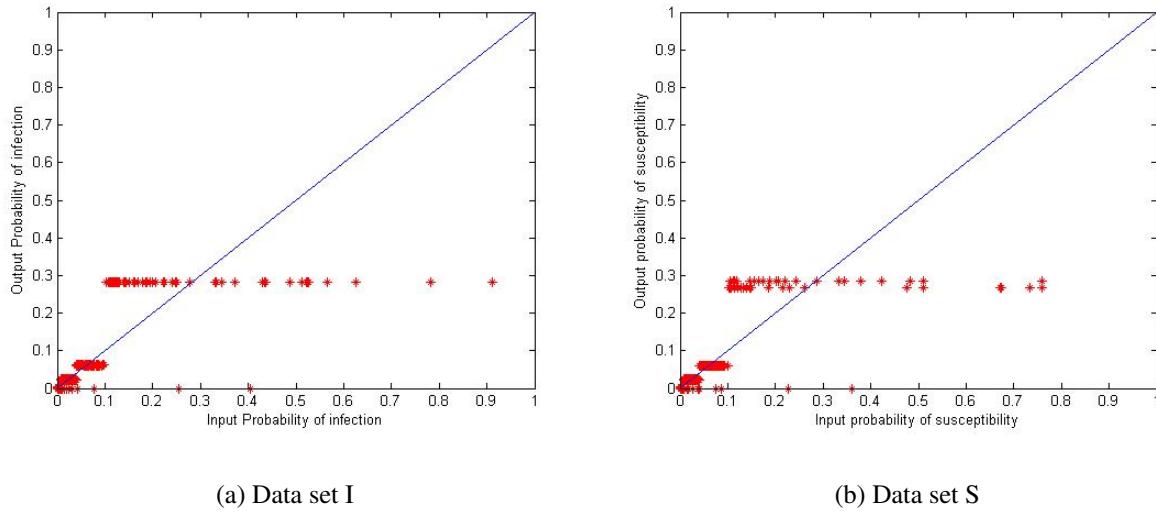


Figure 4.14: Output versus input characteristic feature for Bayesian network towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the Bayesian network.

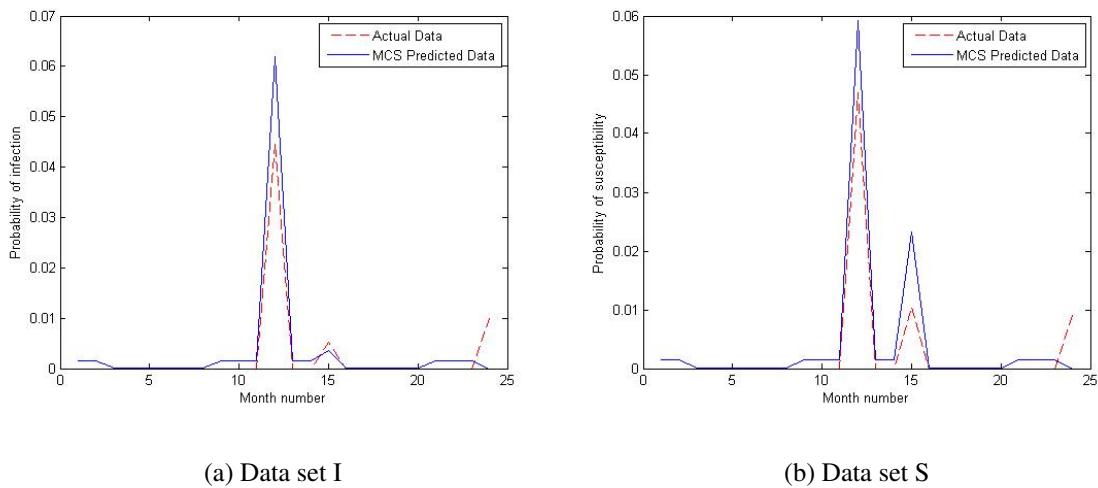
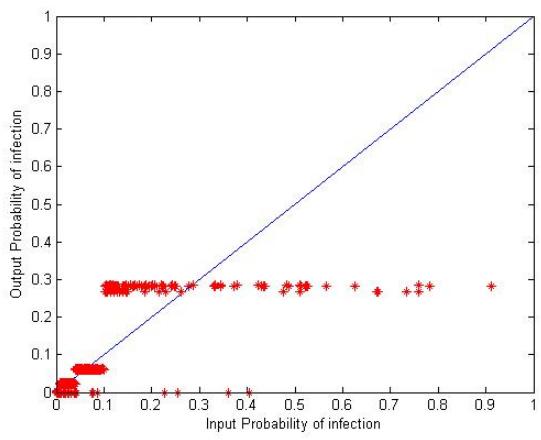
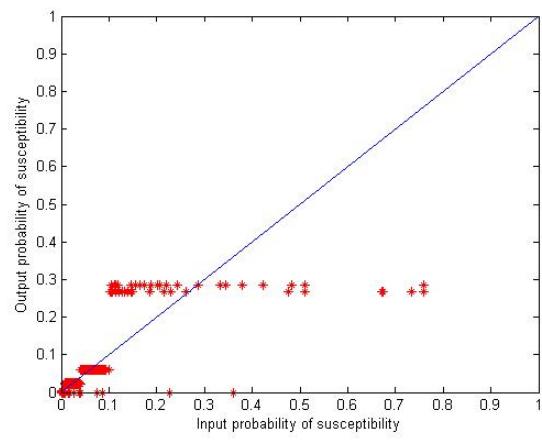


Figure 4.15: MCS predicted probability of infection or susceptibility fitted against actual data on node 52 in Turkey.



(a) Data set I



(b) Data set S

Figure 4.16: Output versus input characteristic feature for MCS towards data set I and S. The input is the actual probability of infection (I) or susceptibility (S) while the output is the probability predicted by the Monte-Carlo Simulations.

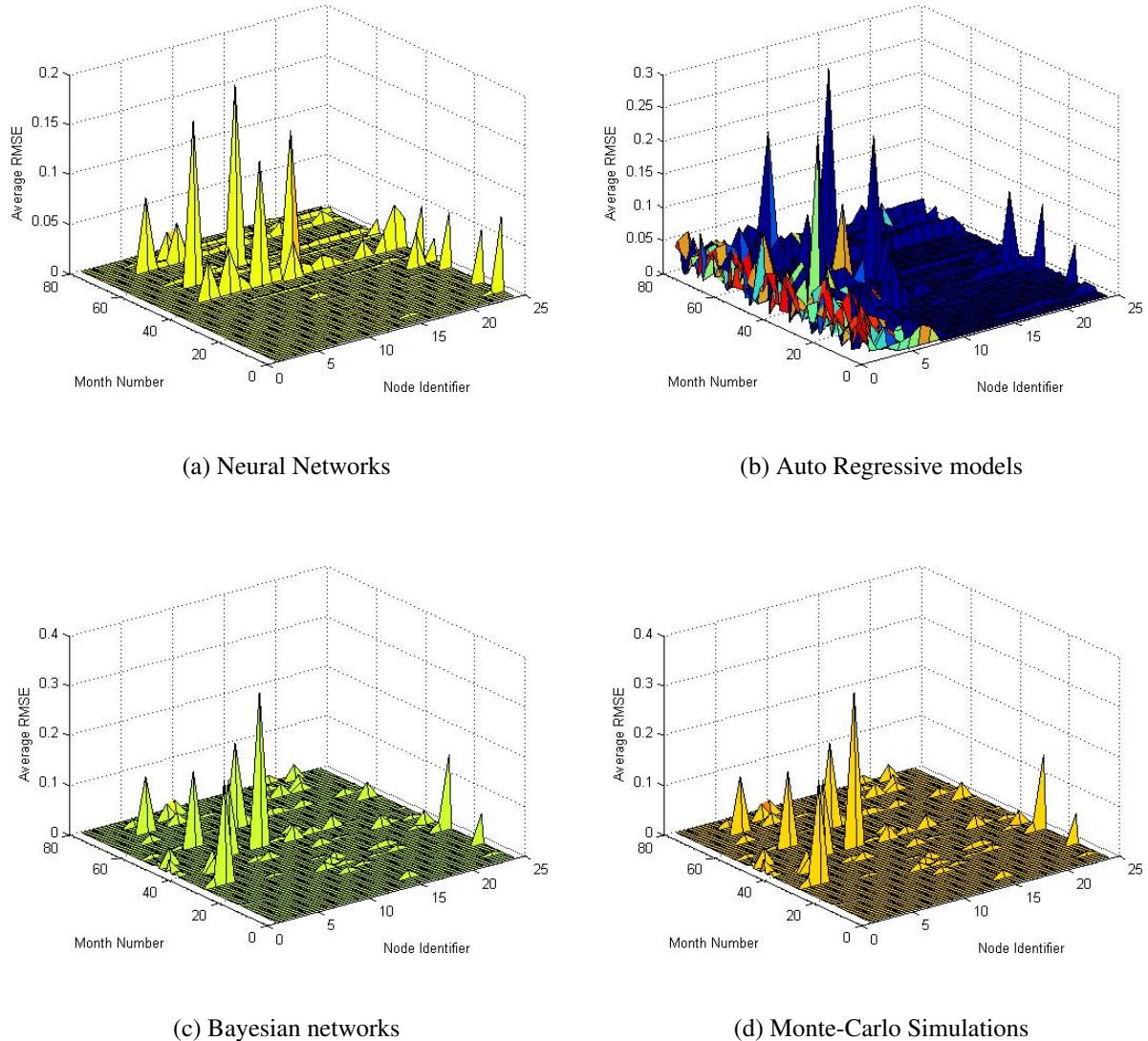
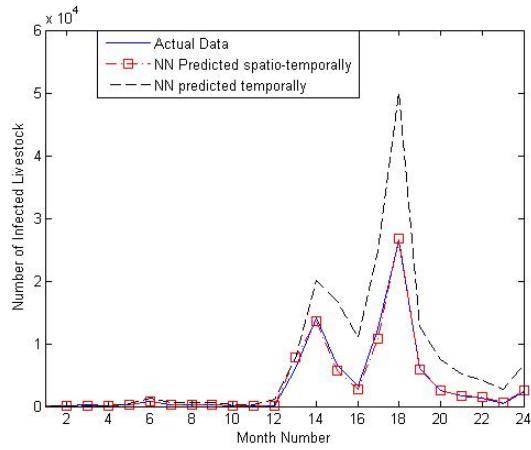
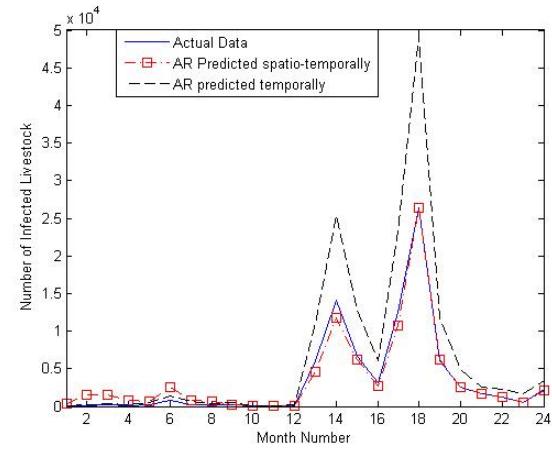


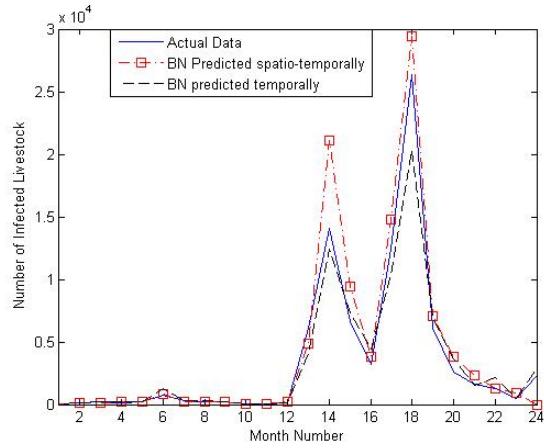
Figure 4.17: The average of RMSE between data set S and data set I for 79 nodes for 24 monthly time instants.



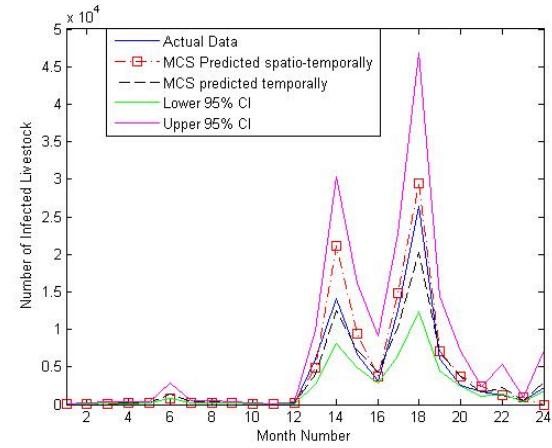
(a) Neural Networks



(b) Auto Regressive models



(c) Bayesian networks



(d) Monte-Carlo Simulations

Figure 4.18: Spatio-temporal versus temporal predictions for number of animals infected monthly.

Chapter 5

MITIGATION STRATEGIES

Mitigation of epidemics is necessary to alleviate the devastating impacts of FMD. Further, theoretical mitigation strategies based on isolation of nodes with high probability of infection [36] are practically infeasible since complete isolation of nodes is impossible in real-world situations. Also, there is a difference in the impedance to the spread of FMD due to the various practical control policies that are adopted such as: movement bans, vaccination, infected premise culls or dangerous contact culls. Though IP and DC culls are very successful in immediately retarding infection spread, they are highly unethical [57]. Conversely, vaccination and movement bans are ethical but they may be costly policies, and they may not retard FMDV spread as much as the culling strategies. Thus, it is absolutely imperative to formulate mitigation strategies that are both practically feasible, ethical and cost-effective too.

We have analyzed spatio-temporal models that model the evolution of FMD using network-based models and learning-based models as well. In this chapter we assess the impact of mitigation strategies using these predictive models. Re-occurrence of FMD epidemics in countries with un-developed livestock surveillance and infection incidence data-bases can be curbed by implementing suitable mitigation strategies based on the phase of the epidemic outbreak. For our analysis, we define an epidemic outbreak having two phases, the pre-outbreak phase where number of infection incidence reports are considerably low, and the outbreak phase wherein tens and hundreds of thousands of livestock are reported to be infected per month. We analyze mitigation strategies on both these phases separately and we attempt to draw some conclusions regarding the

mitigation strategies thus imposed.

Using the network-based spatio-temporal model W , we analyze the background structure of the infection networks thus formed to evaluate some network based properties which can be used to obtain suitable theoretical bounds for FMD mitigation based on the concept of network fragmentation. Hence, we rely on percolation theory, which provides the theoretical guideline to understand the effect on topology of a network due to node deletion [58] [59] since the main contribution of the percolation theory is that if an infinite connected cluster appears in a network, it implies that there is a high probability of an epidemic outbreak.

Network fragmentation helps us set theoretical bounds on the impact of mitigation strategies. These impacts may be in the form of overall reduction in the total number of infected livestock, the total direct costs incurred due to implementation of mitigation strategies, the total number of livestock culled and vaccinated. Since theoretical node isolation is practically infeasible, and since different mitigation tasks impede the spread of infection in a different manner, simulative practical mitigation strategies yield a better insight into the impact of various mitigation strategies. Mitigation tasks that are generally adopted to impede infection spread are premise culls, vaccinations and movement restrictions on livestock grazing and human movements as well. Thus, we study the economic impacts of practical mitigation strategies that have been simulated for analysis. For a complete study, we study the simulative mitigation strategies using the learning-based predictive models as well as the network-based predictive models.

5.1 Theoretical Mitigation

Control strategies to bring the FMD epidemic under control involve isolation of specific nodes which have a high probability of infecting susceptible nodes. Here, isolation refers to removing individual markets (through a movement ban) or farms (through culling) from the FMD infection network by total quarantine, vaccination, or culling of all infected and neighboring livestock. Additionally, we know that the removal of a node from the network interrupts the spread of the disease, and thus network may fragment into one or more clusters [60]. Thus, the size of the giant

component following removal of a node represents the largest possible size of the epidemic at that particular time instant. Also, the number of clusters resulting from the removal of a node is the measure of relative importance of that particular node. Hence, the fragility of the infection network is analyzed by deleting some particular nodes with certain characteristics, and then observing the subsequent characteristics of the network, for example, the size of the largest component, and the number of clusters. Deletion of nodes based on their betweenness centrality, closeness centrality and node degree provides an insight into the optimal control strategies that may be adopted to control the FMD epidemic.

Thus, greater the number of smaller clusters and lesser the size of the largest component, the more effective fragmentation of the network is achieved. In our analysis we have computed network fragmentation by removing certain nodes considering the following network properties:

1. greatest node degree: The degree of a node in a network is the number of links the node has to other nodes in the network. These links may be undirected or bidirectional. Thus, any node with highest node degree would refer to the most connected node, which has a lot of immediate neighbors that it can infect.
2. greatest betweenness centrality: This is the measure of the number of shortest paths that pass through a particular node in the network. A high betweenness centrality and low node degree would imply bridge nodes which can be highly important in the process of infection spread.
3. least closeness centrality: It is defined as the mean shortest path from a particular node to all other nodes reachable from it. A low closeness centrality would imply highly central location of a node and hence it can propagate infection to nodes in the periphery of the network very quickly.

Hence, all the three network properties, node degree, betweenness centrality and closeness centrality can be used to locate the highly significant nodes in the infection network. Significance of a node is defined in terms of the capability of a particular node to spread the infection in the next

time instant. It is thus imperative to decide which network property should be a good parameter to decide upon a highly significant infectious node, and hence we rely on the theory of ergodicity in dynamic infection networks for this purpose.

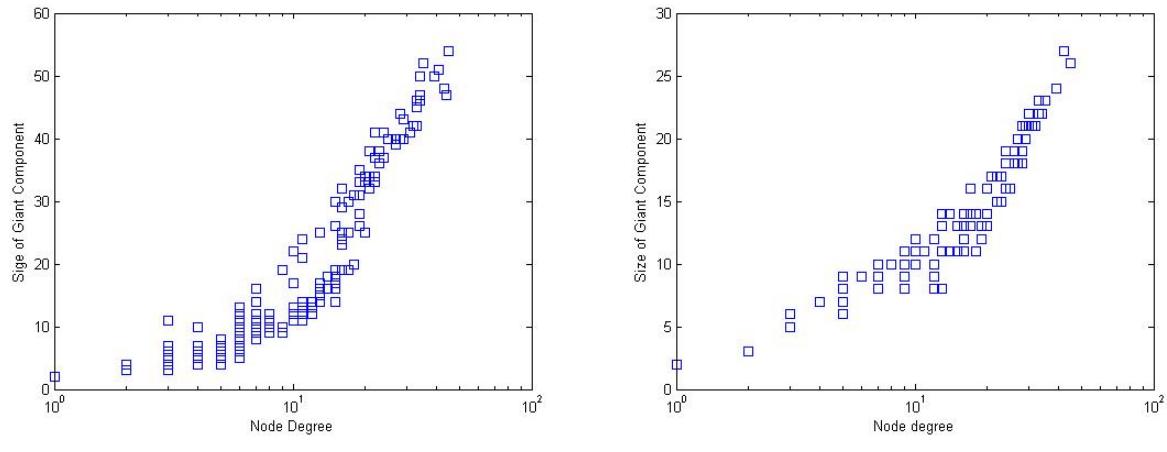
5.1.1 Ergodicity for theoretical mitigation

Earlier in Chapter 3, section 3.3, we have established that the infection networks display ergodicity in their growth rate. That is, irrespective of the realization, the size of the giant connected component versus the secondary reproduction number remains consistent. This concept is useful in understanding the parameters of the underlying network structure of the infection networks thus formed, which remain consistent irrespective of the realization. Since mitigating a single realization of infection networks is not the final goal, we need to identify the network property, which can fragment the infection networks as well as demonstrate ergodicity, or in other words, independence from the method of realization.

Thus, we extend the theory of ergodicity in infection networks to show that the distribution of the size of the giant connected component against the maximum node degree of the infection network is independent of the method of realization. for a range of $p_t h$ in [0.01, 0.2]. However, the size of the giant connected component versus greatest betweenness centrality and least closeness centrality do not demonstrate such consistency in realization as in Figure 5.2, 5.3. This is consistent with the fact that our infection network under consideration is not very large, and certain links in the network are responsible for excessive flow of infection through them. From the Figure 5.1, we infer that the size of the giant connected component versus the greatest node degree is consistent for different realizations, and hence targeted attacks on nodes with the greatest node degree shall consistently fragment any instance of an infection network.

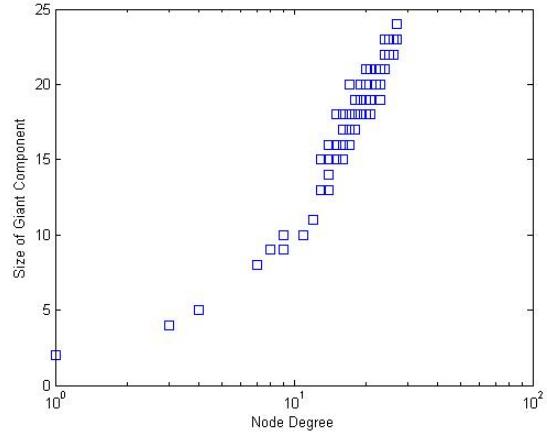
5.1.2 Theoretical network fragmentation

As implied by the ergodic behavior regarding the rate of growth of the giant connected component versus the greatest node degree, we implement infection network fragmentation by deleting/isolating nodes with the highest node degree from the infection network. The impact of theo-



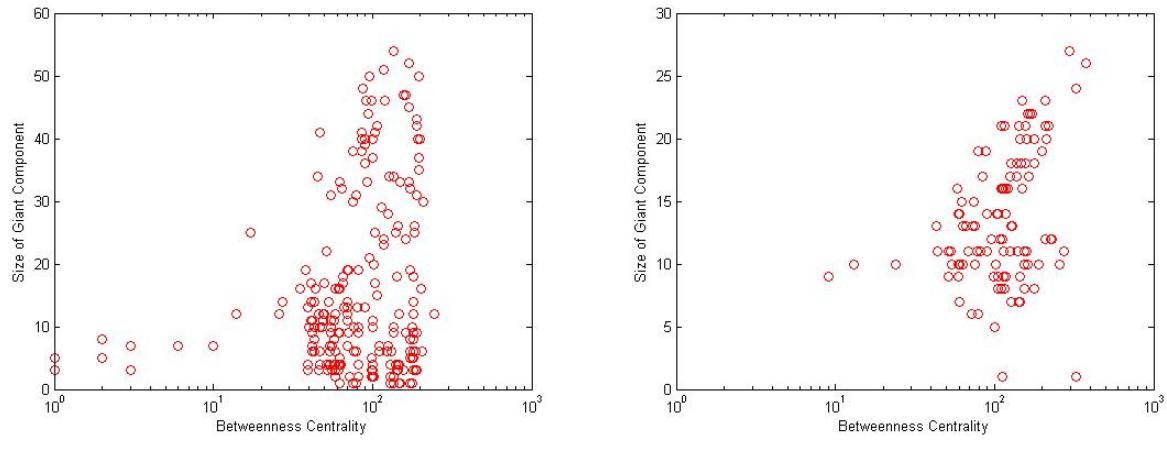
(a) Turkey for 24 month period

(b) Iran for 12 month period



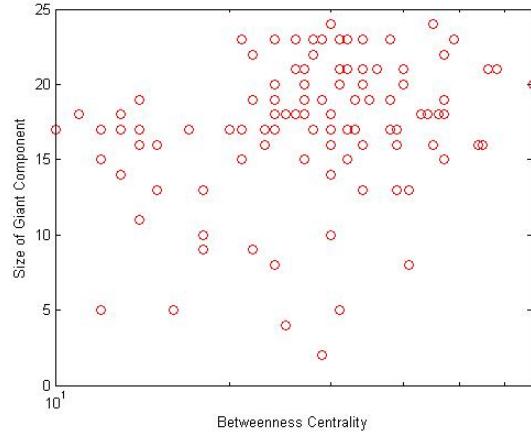
(c) Thailand for 12 month period

Figure 5.1: Ergodic behavior of the rate of growth of Giant connected component versus the highest node degree in the infection network.



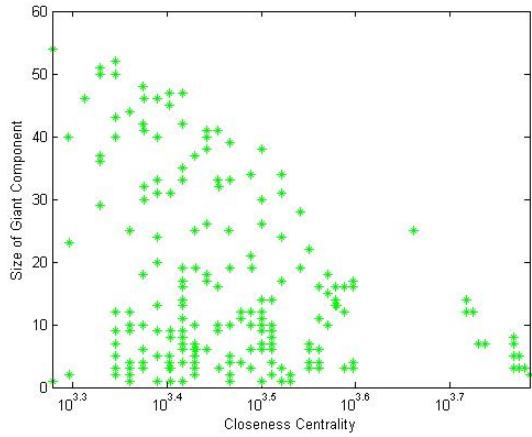
(a) Turkey for 24 month period

(b) Iran for 12 month period

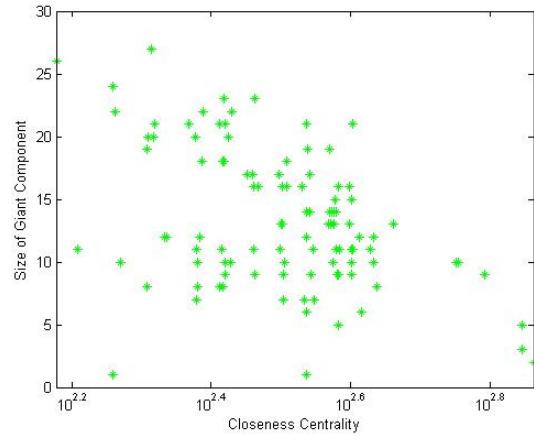


(c) Thailand for 12 month period

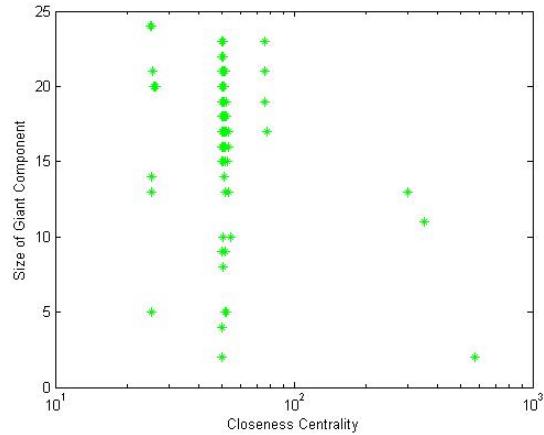
Figure 5.2: Non-Ergodic behavior of the rate of growth of Giant connected component versus the highest betweenness centrality in the infection network.



(a) Turkey for 24 month period



(b) Iran for 12 month period



(c) Thailand for 12 month period

Figure 5.3: Non-Ergodic behavior of the rate of growth of Giant connected component versus the lowest closeness centrality in the infection network.

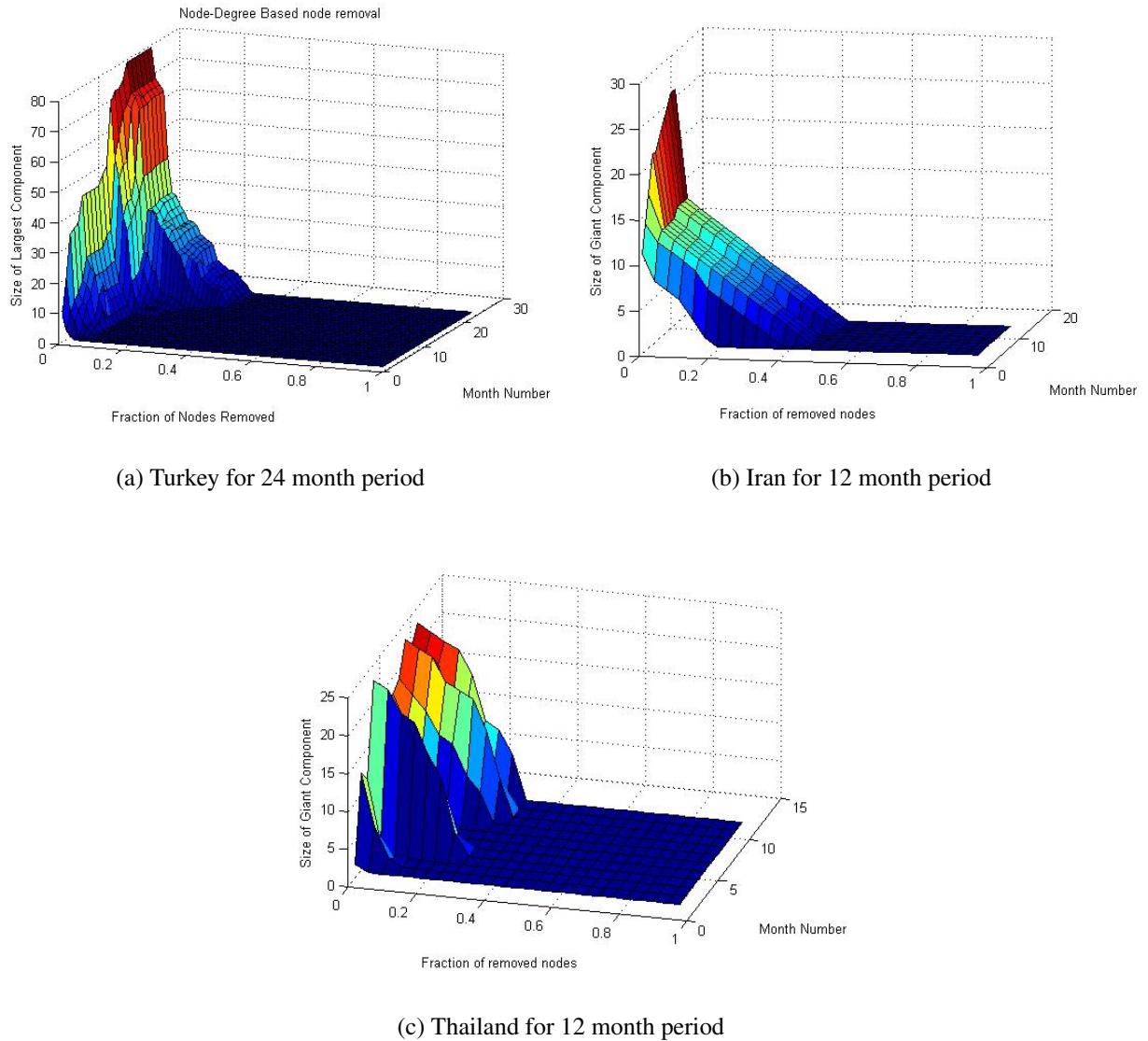


Figure 5.4: Theoretical mitigation by node removal based on highest node degree.

retical deletion of nodes from the infection network based on the greatest node degree is illustrated in Figure 5.4. A highly fragmented infection network implies a drastic reduction in the size of the largest component, thus making eradication strategies effective.

By applying theoretical node removal from the FMD infection networks, we observe that almost 8% node deletion based on greatest node degree for the infection networks in Turkey, results

in 40.8% reduction, while in Iran almost 9% node deletion brings about 38.2% reduction, and almost 10% node deletion in Thailand results in 36.8% reduction in the total number of infected livestock at the end of the period of study respectively. Thus, we set our theoretical targets to achieve a reduction around 40% in the total number of infected livestock by implementing practical mitigation strategies.

5.2 Practical Mitigation

Theoretical mitigation strategies provide an estimate regarding the target theoretical reduction in the number of livestock. Although network fragmentation theoretically checks the spread of infection by reducing the size of the largest component, it is practically not possible to obtain a similar level of node isolation in real-world infection networks. The primary reason being the difference in the impedance to the spread of FMD due to the various practical control policies that are adopted such as: movement bans, vaccination, infected premise (IP) culls or dangerous contact (DC) culls. Although IP and DC culls are very successful in immediate checking of infection spread, they are highly unethical [57]. Conversely, vaccination and movement bans are ethical but they are costly policies, and they fail to check immediate infection spread. Thus, it is absolutely imperative to formulate mitigation strategies that are both practically feasible, ethical and cost-effective too.

For network-based spatio-temporal models, mitigation strategies are simulated using the weighted network based model such that at the end of each month, each node is assigned a certain probability of infection $p_{t,i}$ based on which a certain mitigation task is performed at each node. The following mitigation tasks are adopted to control the spread of FMD by varying the constant c_i of the infected neighbor node in the transmission kernel function. Multiple realizations of these mitigation strategies for (p_{th}) in $[0.0, 0.3]$, provide a good estimate regarding the range of costs, reductions in the number of infected animals, and the respective cost-effectiveness of the mitigation strategies using algorithm 2. Along with the total number of infected animals at each progressive time step ($Inf(t)$), the total number of culled and vaccinated animals at each time

step ($cull(t)$, $vacc(t)$) is calculated according to the classification probabilistic level of $p_{t,i}$.

For learning-based spatio-temporal models, we simulate multiple instances of mitigation strategies which combine the implementation of several mitigation tasks to curb the impacts of an FMD epidemic outbreak. A particular instance of any mitigation strategy is realized by randomly selecting a threshold probability (p_{th}) in [0, 1] for any each node such that, if the probability of infection ($p_{t,i}$) at any node i at time step t , is greater than p_{th} , the node is deemed to be infected, or else the node remains susceptible. In the simulations we have restricted p_{th} in [0.01, 0.5] to reduce the range of variation in the mitigation strategies, since a smaller range of values eases analysis. Various mitigation tasks are implemented by introducing a multiplicative constant $c_{t,i}$ for each $p_{t,i}$. Since these mitigation tasks impede the spread of the FMDV, $c_{t,i}$ represents the extent of impedance by dampening the probability of infection. When no mitigation tasks are enforced, $c_{t,i} = 1$, and when mitigation tasks such as premise culls, vaccination and movement bans are enforced, $c_{t,i} = c$, where c is a constant dependent on the mitigation task.

In the simulation algorithm 1, $sus(i)$ is the total number of animals that can be infected in node i throughout the period of study, and $Inf(t)$ is the evaluated total number of infected animals across all nodes in time t . The mitigations strategies are applied from March 2005 and onwards, since the first two months are considered for sample run, and they also indicate the delay in detection of an FMD outbreak. It is noteworthy that the simulations of the mitigation strategies do not involve re-evaluation of the probability of infection after each time step as done in the case of network-based spatio-temporal models W , NW , $B1$, $B2$. However, the conclusions regarding the effectiveness of the mitigation strategies thus obtained, are parallel to that using the spatio-temporal analysis. The various mitigation tasks applied to impede FMDV spread are as follows.

1. Infected Premise cull (IP): This strategy is adopted at nodes with very high probability of incident infection. Accordingly, all livestock in the particular premise (node) are culled within next 24-48hours. Thus $c = 0.73$, since IP culls induce almost 27% impedance to the infection spread [61]. IP culls provide the most impedance to the spread of infection.
2. Dangerous Contact cull (DC): This strategy is adopted at nodes that are not yet infected but

they have an infected node in their neighborhood and hence, there is a high probability of infection in future time steps. Accordingly, all livestock in such a premise (node) are culled within the next 4-10 days, and thus $c = 0.87$, since DC culls can be estimated to yield 13% impedance to the virus spread [61].

3. Vaccination (Vacc): If the vaccine to the particular strain of FMDV is available, its administration will reduce the probability of future infections. The impedance to the spread of infection is represented by $c = 0.77$, due to almost 23% impedance by a potent oil-based vaccine.
4. No Movement Bans (NM): For nodes not having very high probability of incident infection, but with a potential of getting infected in the near future, the grazing movements of animals are banned. Human movements are also banned from such territories to prevent the spread of the virus. However, a high cost per day is incurred due to these movement restrictions and thus, movement bans are implemented by varying $c = 0.80$, to incur 20% impedance to infection spread. Although [13] states that movement restrictions could lead to as much as 50% reduction in the rate of disease transmission, we dampen this effect in our data set considering the granularity of the data set, and considering the delays in implementation of such movement restrictions.

Finally, we evaluate the impact of six different mitigation strategies on reducing the total number of infected animals at the end of the study periods for the three data sets of Turkey, Iran and Thailand. These mitigation strategies are identified by a set of mitigation tasks defined in Table 5.1.

To devise economic mitigation strategies, it is imperative to understand the cost-effectivenesses of the mitigation strategies. We compute the direct costs of implementation of mitigation tasks by assuming that the cost per head to breed cattle is 1133.5 US dollars and that of sheep is 121.0 US dollars [25]. Additionally, the cost of vaccinating cattle or sheep is 6.0 US dollars per head. However, vaccinations must be well planned by training vaccination teams, provid-

Algorithm 1 Simulative practical mitigation strategies

Randomly generate p_{th} in [0.01,0.5].
 Given $c_{1,i} = 1, c_{2,i} = 1$.

```

for  $i = 1$  to Number of nodes do
    for time  $t = 3$  to Number of months do
        Evaluate  $p_{th1}, p_{th2}, p_{th3}, p_{th4}$  such that
         $p_{th1} = 0.95 * p_{th}$ 
         $p_{th2} = 0.9 * p_{th}$ 
         $p_{th3} = 0.85 * p_{th}$ 
         $p_{th4} = 0.8 * p_{th}$ 
        if  $c_{t,i} * p_{t,i} > p_{th}$  then
             $Inf(t) = Inf(t) + c_{t,i} * p_{t,i} * sus(i)$ 
        end if
        if  $c_{t,i} * p_{t,i} > p_{th1}$  then
             $c_{t+1,i} = c$  for Task 1 due to high impeding infection
            If Task 1 is culling or vaccination,  $cull_i$  or  $vacc_i = c_{t,i} * p_{t,i} * sus(i)$ 
        else if  $p_{th1} \geq c_{t,i} * p_{t,i} \geq p_{th2}$  then
             $c_{t+1,i} = c$  for Task 2 due to considerably high impending infection
            If Task 2 is culling or vaccination,  $cull_i$  or  $vacc_i = c_{t,i} * p_{t,i} * sus(i)$ 
        else if  $p_{th2} \geq c_{t,i} * p_{t,i} \geq p_{th3}$  then
             $c_{t+1,i} = c$  for Task 3 due to considerable impending infection
            If Task 3 is culling or vaccination,  $cull_i$  or  $vacc_i = c_{t,i} * p_{t,i} * sus(i)$ 
        else if  $p_{th3} \geq c_{t,i} * p_{t,i} \geq p_{th4}$  then
             $c_{t+1,i} = c$  for Task 4 due to impending infection in near future
            If Task 4 is culling or vaccination,  $cull_i$  or  $vacc_i = c_{t,i} * p_{t,i} * sus(i)$ 
        else
             $c_{t+1,i} = 1$ , No mitigation task performed due to low impending infection.
        end if
    end for
end for

```

Table 5.1: Sequence of Mitigation Tasks in different Mitigation Strategies.

Strategy	Mitigation Task 1	Mitigation Task 2	Mitigation Task 3	Mitigation Task 4
1	IP	DC	Vacc	NM
2	Vacc	NM	-	-
3	IP	DC	NM	-
4	IP	DC	NM	Vacc
5	IP	Vacc	-	-
6	IP	NM	-	-

Algorithm 2 Simulative practical mitigation strategies

Randomly generate p_{th} in [0.0,0.3].
 $c_i = 1$ for $i \in N$ (*set of all nodes*)
Simulate infection networks for January and February 2005 (Month t=1, t=2 respectively).
for time $t = 3$ to *Number of months* **do**
 Evaluate $p_{th1}, p_{th2}, p_{th3}, p_{th4}$ such that
 $p_{th1} = 0.95 * p_{th}$
 $p_{th2} = 0.9 * p_{th}$
 $p_{th3} = 0.85 * p_{th}$
 $p_{th4} = 0.8 * p_{th}$
 for $i = 1$ to *Number of nodes* **do**
 if $p_{t,i} > p_{th}$ **then**
 $Inf(t) = Inf(t) + sus_{t,i}$
 end if
 if $p_{t,i} > p_{th1}$ **then**
 $c_i = c$ for Task 1 due to high impeding infection
 If Task 1 is culling or vaccination, $cull/vacc(t) = cull/vacc(t) + sus_{t,i}$
 else if $p_{th1} \geq p_{t,i} \geq p_{th2}$ **then**
 $c_i = c$ for Task 2 due to considerably high impending infection
 If Task 2 is culling or vaccination, $cull/vacc(t) = cull/vacc(t) + sus_{t,i}$
 else if $p_{th2} \geq p_{t,i} \geq p_{th3}$ **then**
 if Task 3 exists **then**
 $c_i = c$ for Task 3
 If Task 3 is culling or vaccination, $cull/vacc(t) = cull/vacc(t) + sus_{t,i}$
 end if
 else if $p_{th3} \geq p_{t,i} \geq p_{th4}$ **then**
 if Task 4 exists **then**
 $c_i = c$ for Task 4
 If Task 4 is culling or vaccination, $cull/vacc(t) = cull/vacc(t) + sus_{t,i}$
 end if
 else
 $c_i = 1$, No mitigation task performed due to low impending infection.
 end if
 end for
 Simulate infection network for month t .
end for

ing proper equipment, and preparing for quick transportation of vaccines to affected sites. We estimate such indirect costs at around 200,000 US dollars per geographic location. The cost to administer euthanasia or cull cattle is 16.5 US dollars per head and 2.31 US dollars per head for sheep [25]. Premise culling costs would thus include the breeding costs for livestock, the cost for the safe disposal of carcass, for administering euthanasia and decontamination costs. In case of movement restrictions, a cost of 157,968 US dollars is incurred per day for every geographic location [45]. Although the mitigation strategies suggested above are effective in reducing unnecessary and unethical livestock culls, it is imperative to understand their cost-effectiveness as well. These costs help in formulating the different mitigation strategies, wherein the priority of each of the four mitigation tasks is different. The cost-effectiveness of a mitigation strategy is defined as following.

$$effectiveness = \frac{\% \text{ Reduction in number of infected livestock}}{\text{Cost incurred (million US \$)}} \quad (5.1)$$

5.2.1 Learning-Based Models

We analyze the impact of the 6 mitigation strategies using the learning based models described in Chapter 4 on the data set in Turkey. Since the learning-based models incur a higher error in predictions, mitigation using these models do not reflect very accurate strategies quantitatively. However, the significances of the mitigation strategies in terms of their effectiveness and in impeding the spread of epidemics remains unchanged. Thus we wish to analyze the impact of each mitigation strategy towards reducing the total number of infected animals. The cost-effectiveness of these strategies is elucidated in Figures 5.5, 5.6.

Analysis of the impacts of the 6 mitigation strategies in terms of the percentage reduction in the total number of infected livestock at the end of the period of study, the cost of implementation of the mitigation strategies in million US dollars, the total number of culled and vaccinated livestock is presented in Table 5.2. Thus, we infer that strategy 5 incurs the most number of culled and vaccinated animals, because of which, this strategy results in the maximum reduction in the total

number of infected animals. Strategy 5 is thus very effective in impeding the rate of spread of FMDV when an epidemic has already set in. However, strategy 1 and 4 which rely on all 4 mitigation tasks have cost-effectiveness similar to each other, and they both incur a high cost of implementation. This shows that implementation of all 4 mitigation task merely increases the total cost of implementation without improving on the percentage reduction of the number of infected animals or the strategy *effectiveness*. Thus, it is important to plan the order of mitigation tasks instead of considering each task to be equally important. Also, we observe that the reduction in the total number of infected livestock is almost double for culling based strategies 1, 3, 4, 5, 6 when compared to vaccination based strategy 2; however, the *effectiveness* of strategy 2 is the highest. Thus, vaccinations must be preferred to culling strategies in the pre-outbreak period. Finally, we observe that strategy 2 and 5 which rely on vaccination strategies have the highest range of effectiveness. This shows that vaccination strategies need to be well planned, and a good vaccination program needs to be developed, to obtain high cost-effectiveness of mitigation strategy 2, otherwise vaccination will not be effective in impeding the spread of FMDV. Also, potency of vaccines play an important role in determining the cost-effectiveness of any vaccination strategy, since oil-based vaccines are considered to be more potent and effective than the water-based ones.

5.2.2 Network Based Models

From Table 5.3 and Figure 5.7, we infer that pre-emptive vaccination followed by movement restrictions imposed by mitigation strategy 2 is the most cost effective strategy. However, mitigation strategies that impose culls at the most probable nodes of infection are effective in reducing the total number of infected animals by as much as 50%. Also, IP culls when followed up with vaccinations as in mitigation strategy 5, are cost effective as well as they achieve a high reduction in the total number of infected animals.

From Figure 5.8 we infer that pre-emptive vaccination followed by movement restrictions imposed by mitigation strategy 2 is the most cost effective strategy. However, other mitigation strategies that impose culls at the most probable nodes of infection are effective in reducing the

Table 5.2: Cost-effectiveness of mitigation strategies based on learning-based predictive models on data set of Turkey.

Model	Strategy	%Reduction	Cost(million USD)	effectiveness	Culled	Vaccinated
NN	1	27.338-34.253	514.514-680.597	0.05-0.053	154-1108	0-207
	2	12.419-18.019	31.244-229.052	0.079-0.397	0	195-373
	3	26.488-35.023	443.186-614.489	0.057-0.06	154-905	0-199
	4	27.134-36.896	490.215-677.393	0.054-0.055	154-903	0-222
	5	23.663-46.794	87.892-360.019	0.13-0.269	2217-15295	0-1310
	6	26.54-33.891	507.738-767.864	0.044-0.052	154-1064	0
AR	1	27.055-37.178	371.233-557.004	0.067-0.073	110-1388	0-405
	2	13.23-22.299	23.511-80.703	0.276-0.563	0	140-1761
	3	26.961-37.552	367.508-523.402	0.072-0.073	110-1388	0-372
	4	27.879-36.669	371.624-548.711	0.067-0.075	110-1388	0-225
	5	28.183-34.972	360.779-536.617	0.065-0.078	110-1388	0-207
	6	27.606-36.569	372.846-561.948	0.065-0.074	110-1388	0
BN	1	20.177-31.029	501.88-768.027	0.04-0.042	483-965	0-21
	2	9.627-17.408	33.88-42.338	0.284-0.411	0	770-1224
	3	20.006-31.644	501.88-584.517	0.04-0.054	483-965	0
	4	20.002-31.872	501.88-906.258	0.035-0.04	606-965	0
	5	20.608-34.468	143.742-384.179	0.09-0.143	3318-8888	0-437
	6	20.278-31.675	501.88-753.125	0.04-0.042	585-965	0
MCS	1	20-32.19	501.88-660.437	0.04-0.049	483.004-965	0-21
	2	9.627-15.07	33.88-0.377	0.284-0.373	0	743-1224
	3	20.014-32.388	501.88-579.315	0.04-0.056	800-965	0-21
	4	20.107-33.666	501.88-753.479	0.04-0.045	610-965	0
	5	22.084-38.085	144.094-402.147	0.095-0.153	3976-8888	0-437
	6	20.087-31.84	501.88-821.285	0.039-0.04	585-965	0

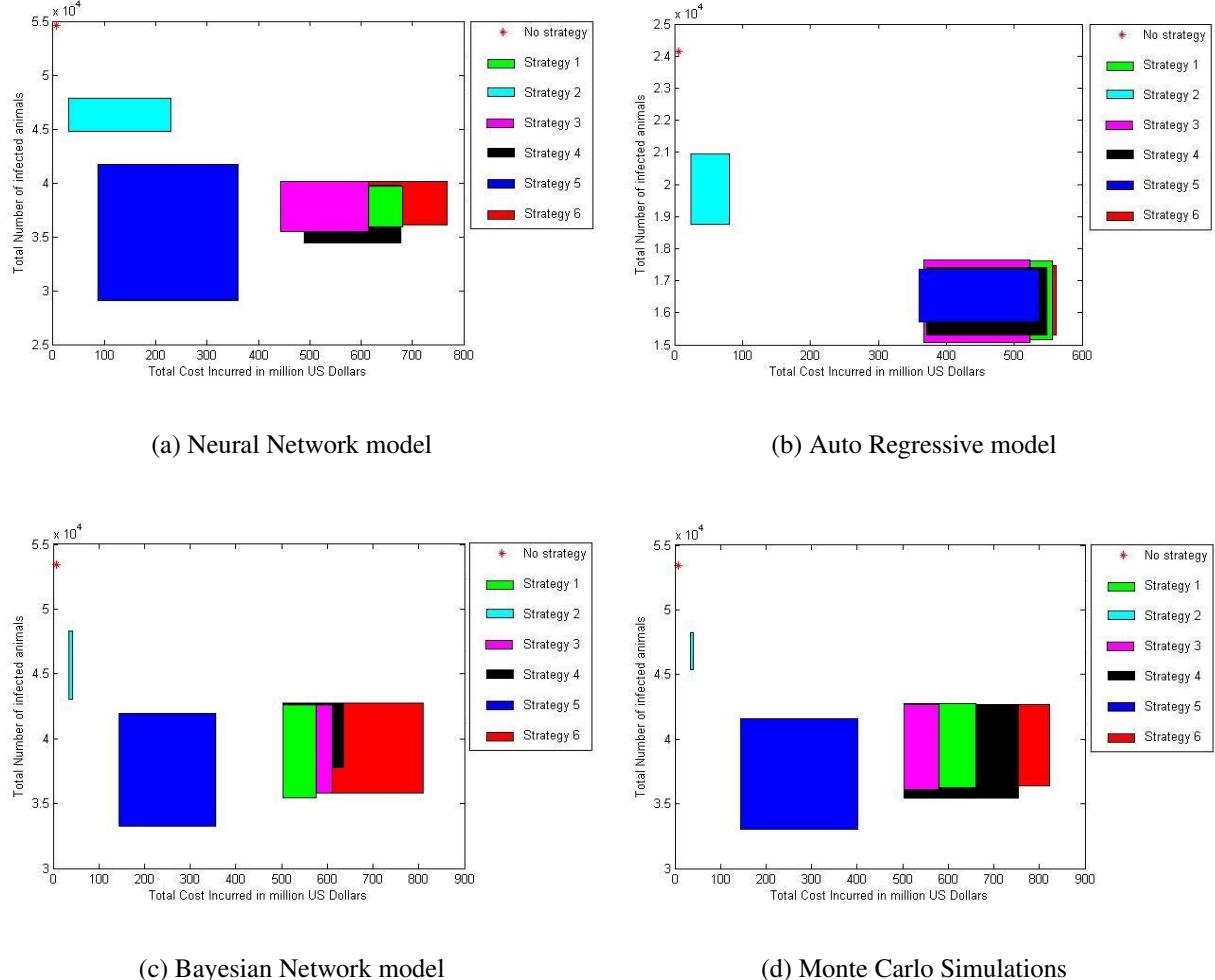
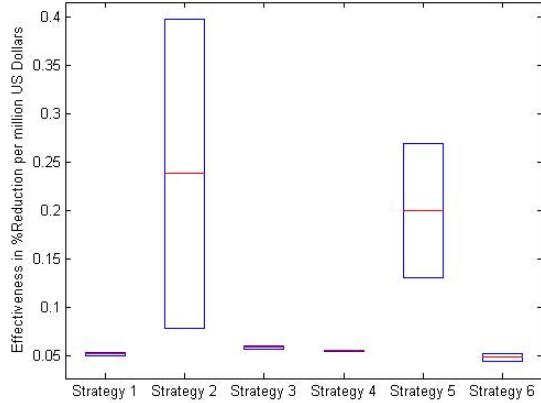
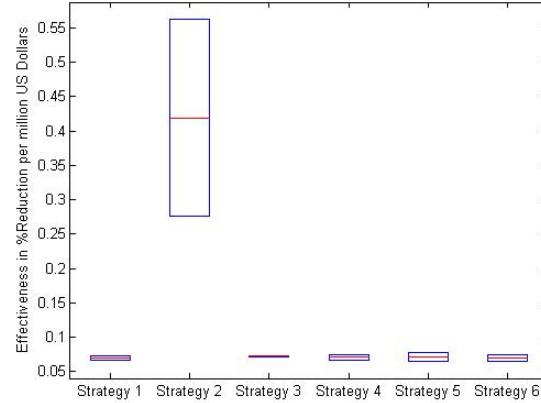


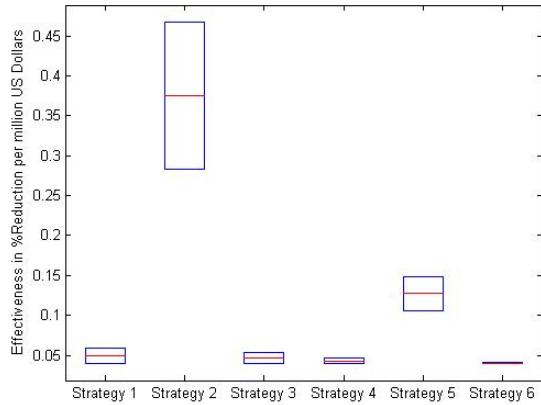
Figure 5.5: Total number of infected animals presented against the total cost incurred for the 6 mitigation strategies using learning-based predictive models.



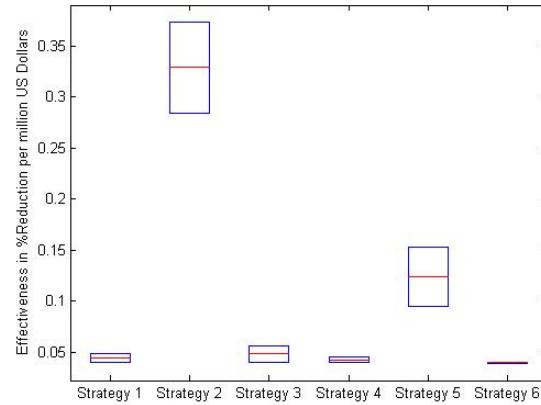
(a) Neural Network model



(b) Auto Regressive model



(c) Bayesian Network model



(d) Monte Carlo Simulations

Figure 5.6: Effectiveness of 6 mitigation strategies using learning-based predictive models.

Table 5.3: Cost-effectiveness of network-based mitigation strategies on data set of Turkey, Iran and Thailand.

Country	Strategy	%Reduction	Cost in US Dollars	effectiveness	Culled	Vaccinated
Turkey	1	50.109 - 60.997	472.97 - 524.74	0.105 - 0.116	33968 - 34015	0 - 10119
	2	29.927 - 31.535	199.50 - 310.36	0.150 - 0.101	0	448 - 38693
	3	45.439 - 57.56	473.13 - 571.013	0.096 - 0.101	378 - 34023	0
	4	50.122 - 57.801	473.00 - 574.55	0.101 - 0.106	3723 - 34015	0 - 49
	5	50.399 - 56.174	435.68 - 481.51	0.115 - 0.116	3723 - 34015	0 - 10119
	6	50.226-57.633	432.93 - 571.25	0.100 - 0.116	3723 - 34015	0
Iran	1	50.253 - 63.179	169.87 - 247.13	0.255 - 0.295	948 - 16574	0 - 13982
	2	26.974 - 33.544	23.00 - 37.26	0.900 - 1.172	0	0 - 14982
	3	52.378 - 66.108	159.19 - 241.51	0.273 - 0.329	0 - 14385	0
	4	50.912 - 65.213	168.39 - 252.82	0.257 - 0.302	894 - 16574	0 - 64227
	5	46.539 - 65.111	167.60- 276.14	0.235 - 0.277	948 - 15409	0 - 155404
	6	47.807 - 64.207	165.84 - 250.50	0.256- 0.288	948 - 14203	0
Thai	1	39.303 - 51.560	64.56 - 105.83	0.487 - 0.608	18169 - 19034	0-2709
	2	20.946 - 28.473	14.40 - 32.89	0.865 - 1.454	0	4172 - 4544
	3	39.303 - 52.357	60.24 - 108.79	0.481 - 0.652	18169 - 20490	0
	4	34.491 - 56.284	96.73 - 141.56	0.356 - 0.397	0	0 - 854
	5	39.132 - 51.966	51.35 - 56.21	0.761 - 0.924	18076 - 26212	0-2969
	6	39.110 - 47.092	60.15 - 87.56	0.537 - 0.650	14033 - 20533	0

total number of infected animals to almost 50%. Also, IP culls when followed up with vaccinations as in mitigation strategy 5, are cost effective as well as they achieve a high reduction in the total number of infected animals.

5.2.3 Impact of prediction errors on mitigation strategies

We quantify the impact of mitigation strategies using *effectiveness* following which we observe that predictions errors cause non-linear variations in the effectiveness of these mitigation strategies. Hence, it is imperative to analyze the variations in the effectiveness of fixed mitigation strategies caused due to prediction errors, which in turn analyzes the robustness of mitigation strategies against prediction errors. For instance, if any future spatio-temporal models claim as high as 80-90% improvement in prediction error, but if the effectiveness of the mitigation strategies change only by about 20%, we may say that our model W still holds good for understanding the impact of mitigation strategies. Hence, for all baseline models with higher prediction errors as compared to spatio-temporal model W , we analyze the impact on the effectiveness of the 6 mitigation strategies.

We analyze the change in effectiveness of each of the 6 mitigation strategies owing to degraded prediction from baseline network models. We observe the change in prediction error ($\Delta sMAPE$) in baseline models ($NW, B1, B2$) with respect to model W . For a constant set of mitigation strategies, we assess the impact of degraded prediction performance on the outcome of the mitigation strategies. The utility factor in the downstream assessment of the mitigation strategies is defined in terms of change in effectiveness ($\Delta effectiveness$) caused by the baseline models with respect to that caused by model W . While $sMAPE$ is prediction error in percentage, mitigation strategy *effectiveness* has a unit of percentage per million US dollars. The downstream impact for Turkey is elucidated in Figure 5.10, 5.11. The magnified impacts of prediction error on mitigation strategies in Iran and Thailand are shown in Figure B.1, B.2 and B.3, B.4 respectively.

We observe that although the gradient of $\Delta sMAPE$ increases sharply from model NW through model $B2$, the impact on mitigation strategies $\Delta effectiveness$ changes considerably

for mitigations strategy 2 that imposes vaccinations. Conversely, strategies that impose culling as the primary mitigation task such as strategy 1,3,4,5 and 6 do not depict a high variation in effectiveness. This fact is evident from Figures 5.10, 5.11. We observe that for the baseline model *NW*, 43% variation in prediction error results in almost 35% variation on mitigation effectiveness. Model *B1* incurs about 42% variation in effectiveness for a 52% change in prediction error, while model *B2* incurs about 47% variation in effectiveness for 57% change in prediction error.

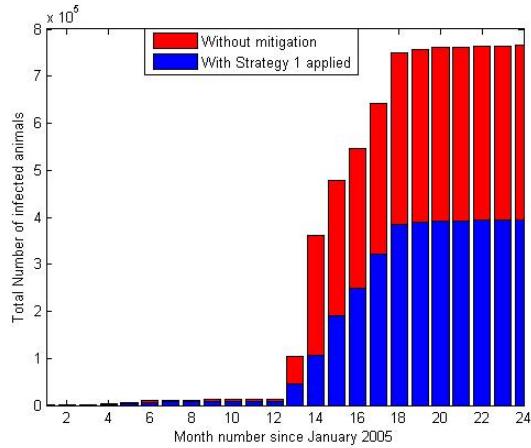
We have similar observations for prediction errors using learning-based models. Table 5.4 shows the range of variation in *effectiveness* of the 6 mitigation strategies ($\Delta eff1$ through $\Delta eff6$) with variation in prediction errors ($\Delta sMAPE$). Here, we analyze the variation in prediction error, and variation in cost-effectiveness of the mitigation strategies between the learning-based models and the spatio-temporal predictive model in [36]. Since this spatio-temporal model incurs lower prediction errors due to its global parameterization, we assess the impact of increased prediction errors, by the learning-based models, on the downstream utility function defined in terms of *effectiveness* of the mitigation strategies. We observe that for NN models almost 14% change in prediction error causes about 30% variation in *effectiveness*, while AR models show 24% variation in *effectiveness* for 56% prediction error. BN and MCS show around 18% variation in *effectiveness* due to a prediction error of 48%. Thus, we are able to assess the non-linear variation and robustness in *effectiveness* of mitigation strategies for prediction errors from the learning-based prediction models since, as much as 50% prediction errors incur as much as 30% variation in cost-effectiveness of mitigation strategies. Further, we find that the variation in prediction errors for NN and AR with respect to the spatio-temporal model *NW*, *B1*, *B2*, which requires parameters such as wind, human movement and grazing movement of animals, is lesser than that of other spatio-temporal models with lower parameterization. Hence, we infer that the predictive performance of learning-based models can be better than under-parameterized spatio-temporal models.

In this chapter, we have simulated instances of mitigation strategies such that the probability bands for determining the nodes with high infection incidence have been fixed at 0.95, 0.9,

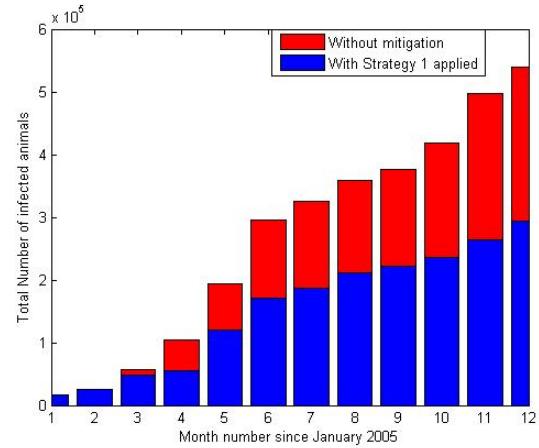
Table 5.4: Impact of prediction errors using learning-based models

Statistic	NN	AR	BN	MCS
min $\Delta sMAPE$	1.5469	12.5388	6.6432	8.4945
max $\Delta sMAPE$	2.6581	13.6500	7.7544	9.6057
min $\Delta eff1$	0.0352	0.0188	0.0452	0.0368
max $\Delta eff1$	0.0440	0.0242	0.0569	0.0572
min $\Delta eff2$	0.0701	0.1276	0.1505	0.1505
max $\Delta eff2$	0.2638	0.4291	0.2624	0.2245
min $\Delta eff3$	0.0661	0.0513	0.0689	0.0671
max $\Delta eff3$	0.0716	0.0580	0.0915	0.0915
min $\Delta eff4$	0.0420	0.0268	0.0613	0.0518
max $\Delta eff4$	0.0465	0.0296	0.0620	0.0618
min $\Delta eff5$	0.0006	0.0524	0.0129	0.0228
max $\Delta eff5$	0.1387	0.0654	0.0409	0.0359
min $\Delta eff6$	0.0415	0.0197	0.0533	0.0537
max $\Delta eff6$	0.0583	0.0373	0.0604	0.0636

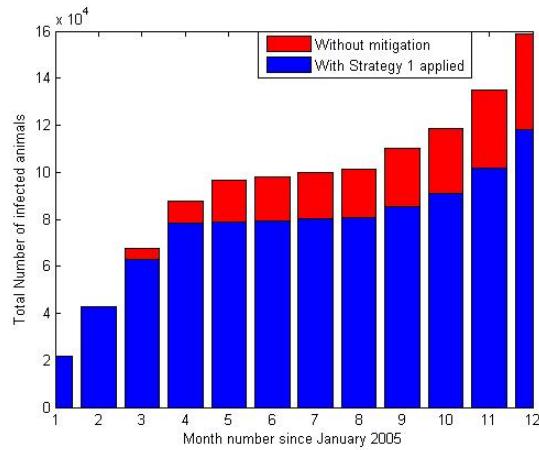
0.85, 0.8 times the p_{th} . In the next chapter, we eliminate these fixed probability bands by applying fuzzy dominance genetic algorithms to dynamically generate the threshold probabilities $p_{th1}, p_{th2}, p_{th3}, p_{th4}$ so as to minimize the total number of infected animals, the total direct costs, the total number of culled and the total number of vaccinated animals respectively.



(a) Turkey for 24 month period



(b) Iran for 12 month period



(c) Thailand for 12 month period

Figure 5.7: Reduction in the total population of infected livestock due to implementation of mitigation strategy 1 adopted in Turkey, Iran and Thailand. The bounds on this reduction elucidate the impact of immediate culls, potent vaccines and strict movement restrictions as well as delayed culls and less strict movement bans. The implementation of the mitigation strategy is assumed from March 2005 and onwards each data set. Delayed implementation shall result in a lesser reduction of the total number of infected animals.

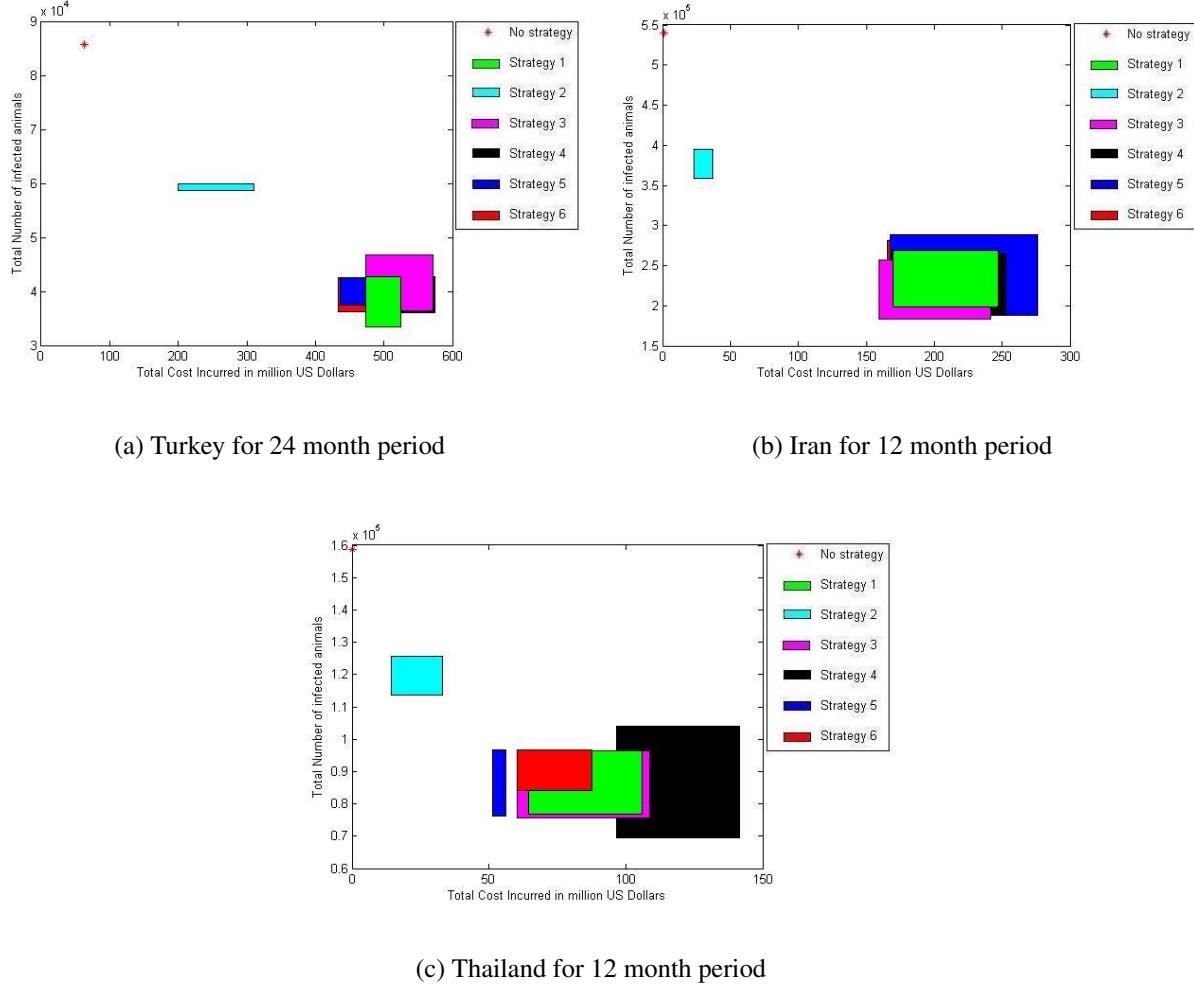


Figure 5.8: Total number of infected animals presented against the total cost incurred for the respective mitigation strategies.

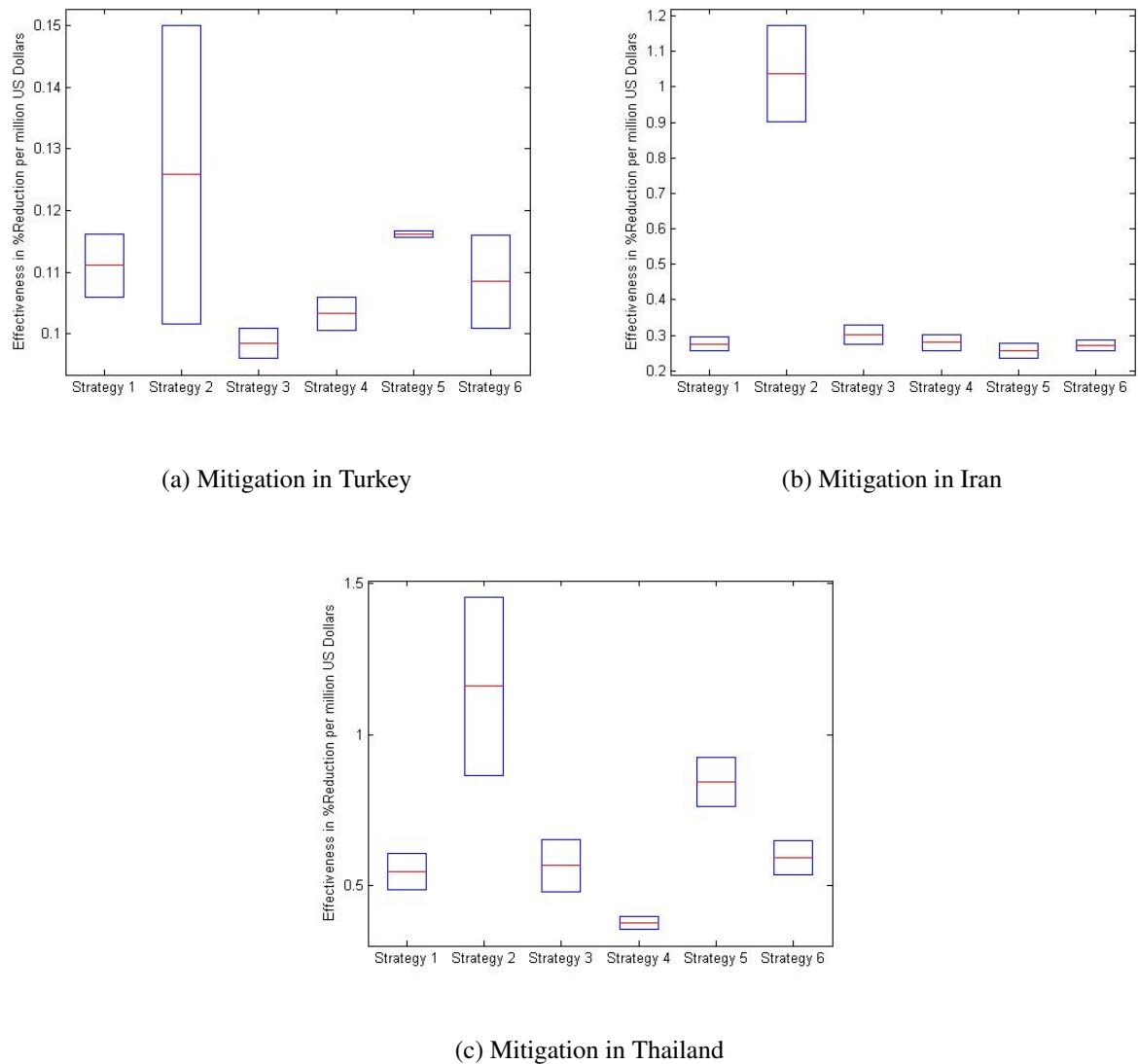
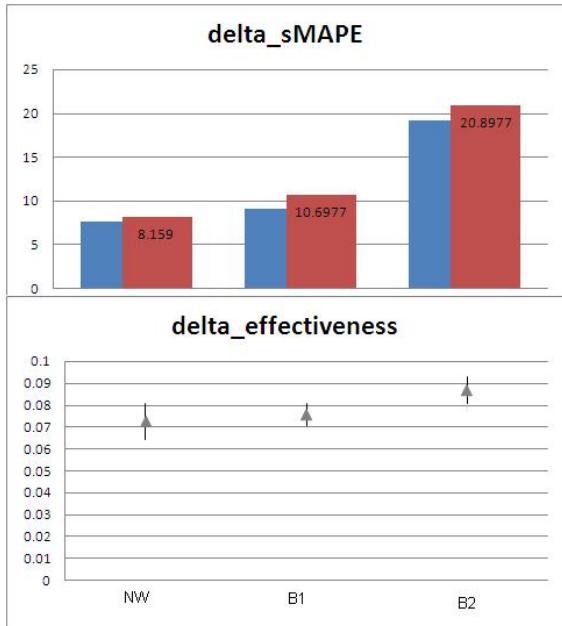
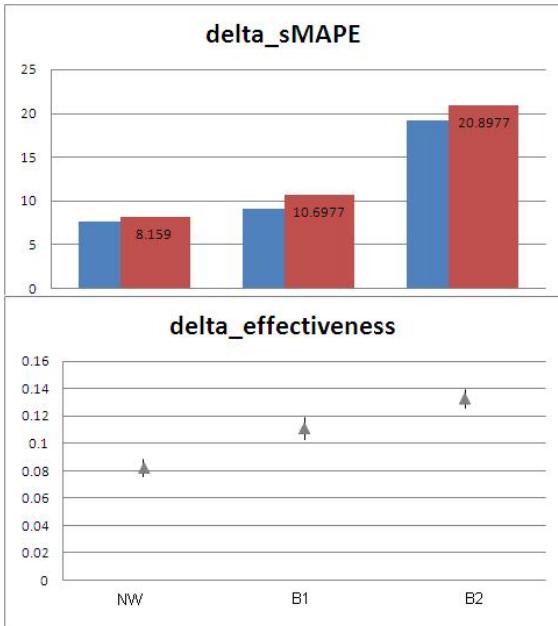


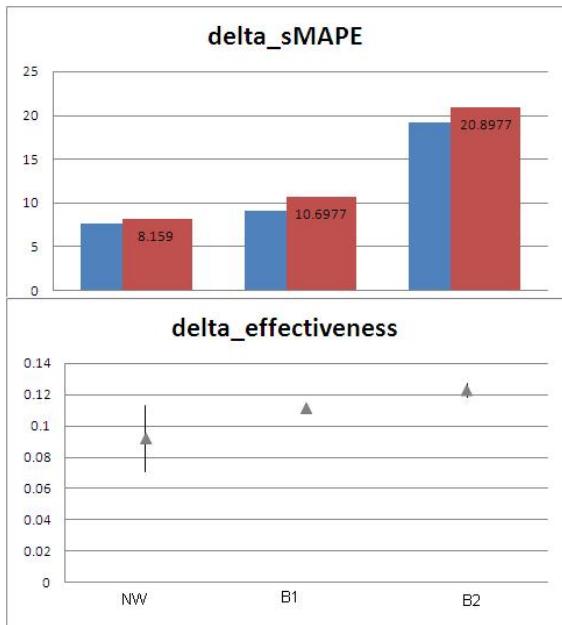
Figure 5.9: Cost-effectiveness of 6 mitigation strategies in Turkey, Iran and Thailand



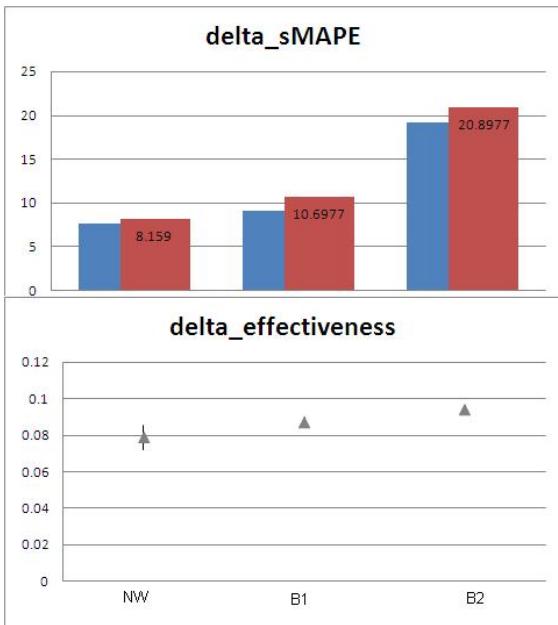
(a) Mitigation Strategy 1



(b) Mitigation Strategy 2



(c) Mitigation Strategy 3



(d) Mitigation Strategy 4

Figure 5.10: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Turkey. The prediction percentage error between the baseline models and model W($\Delta sMPAE$) is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.

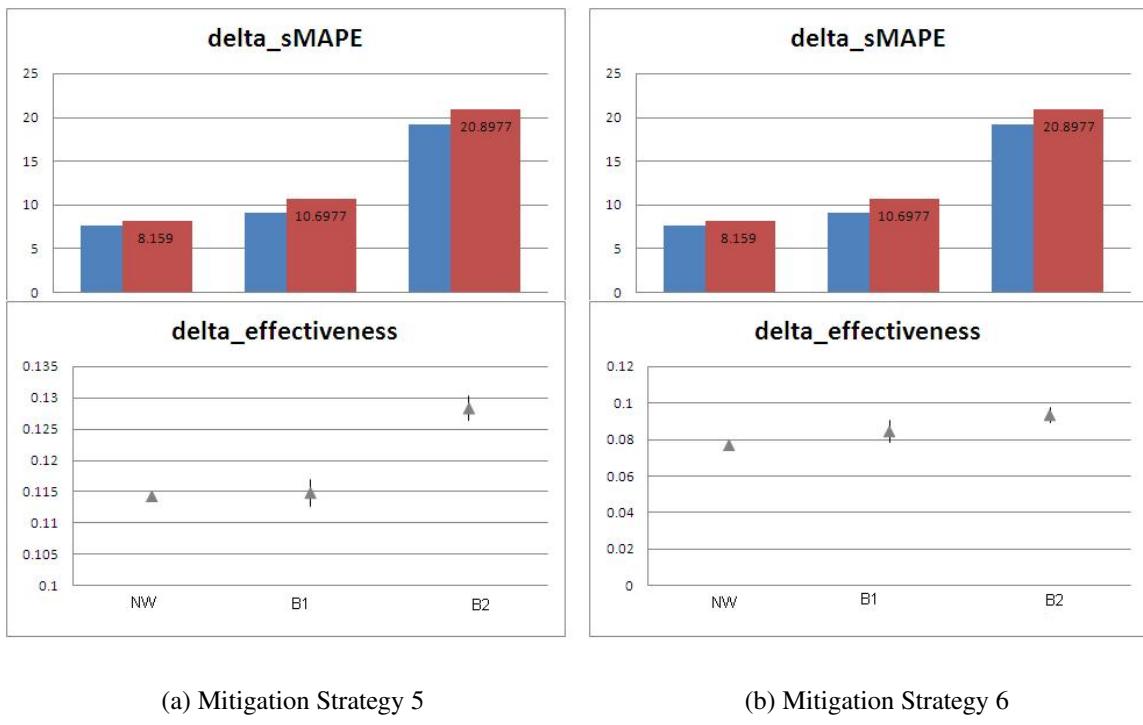


Figure 5.11: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Turkey.

Chapter 6

OPTIMAL MITIGATION USING GENETIC ALGORITHMS

A genetic algorithm (GA) is a way of mimicking and solving real-world optimizations problems by providing exact or approximate solutions. As a particular class of evolutionary algorithms (EA), GA use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. From an implementation standpoint, GA is a computer simulation in which a population of abstract representations (called chromosomes or the genotype of the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves towards better optimal solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and progresses in generations. In every generation, the fitness of every individual in the population is evaluated, multiple individuals are selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm.

6.1 Methodology for Genetic Algorithms

The steps for optimization using GA is described below:

1. Initialization: The initial population comprises of randomly generated individuals. The population size depends on the nature of the problem, but typically contains several hundreds or

thousands of possible solutions. Traditionally, the population is generated randomly, covering the entire range of possible solutions. Occasionally, the solutions may be “seeded” in areas where optimal solutions are likely to be found.

2. Selection: During each successive generation, a proportion of the existing population is selected to give rise to a new generation. Individual solutions are selected through a fitness-based process to minimize certain objective functions. Most fitness-functions are stochastic and designed so that a small proportion of less fit solutions are selected. This helps keep the diversity of the population large, preventing premature convergence on poor solutions. Popular and well-studied selection methods include roulette wheel selection and tournament selection.
3. Reproduction (Crossover and Mutation): The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation. For each new solution to be produced, a pair of “parent” solutions is selected for breeding from the pool selected previously. By producing a “child” solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its “parents”. New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated. Although reproduction methods that are based on the use of two parents are more “biology inspired”, some research suggests more than two “parents” are better to be used to reproduce a good quality chromosome. These processes ultimately result in the next generation population of chromosomes that is different from the initial generation.
4. Termination: This generational process is repeated until a termination condition has been reached such that a solution is found that satisfies minimum criteria.

The genetic algorithm has parameters determining the nature of mutation and crossover as well as the population size. These parameters can be tuned to improve performance for the problem class at hand.

Algorithm 3 Pseudo code for typical genetic algorithm

Choose the initial population of individuals (pop)
Evaluate the fitness of each individual in that population($fitness(pop)$)
repeat
 Select the best-fit individuals for reproduction
 Breed new individuals through crossover and mutation operations to give birth to offspring
 Evaluate the individual fitness of new individuals
 Replace least-fit population with new individuals
 Parameter Tuning
until Termination condition satisfied

6.2 Optimal Mitigation using GA

We begin our analysis using the weighted network based prediction model for FMD, W defined in Chapter 3 and the simulative practical mitigation strategy 1 as described in Chapter 5, section 5.2, wherein the nodes most susceptible to infection are administrated with IP culls, followed by DC culls, vaccination and no-movement bans. We study this strategy since it involves all the mitigation tasks in the order of criticality, and since it is the most critical mitigation strategy that requires optimization for an improved mitigation effectiveness. As a sample data set we analyze process of optimal mitigation in Turkey and accordingly we observe the total reduction in the number of infected animals from January 2005 through December 2006 due to implementation of the mitigation strategy that is calculated for all the administrative locations (nodes) collectively. The cost incurred due to the implementation, along with the total number of animals culled and vaccinated across all nodes over the 24 month observation period is also calculated. The aim of the optimal mitigation strategy is to choose an optimal set of threshold probabilities $pth_1, pth_2, pth_3, pth_4$, for each month, such that at the end of December 2006, there is a reduction in the total number of infected animals, low cost of implementation, less number of animals culled preventively and optimally less number of animals that are vaccinated. Thus, the set of solutions to be optimized is a set of 88 probabilities (i.e. 4 threshold probabilities for 22 months). We consider 22 months, since the infection is seeded in the first month and the predictive output for month 2 is a sample run before the mitigation strategy is actually introduced. Thus the mitigation strategy is implemented

from month 3 through month 24 ending in December 2006.

Thus, we apply genetic algorithm based on the concept of fuzzy dominance to minimize four objectives defined as below:

- Objective 1: Total Number of Infected animals (f_{NI}).
- Objective 2: Cost Incurred due to implementation of mitigation strategy. This cost is estimated in million US Dollars (f_{DC}).
- Objective 3: Total Number of animals culled as a preventive measure (f_{NC}).
- Objective 4: Total Number of animals vaccinated (f_{VC}).

6.2.1 Theoretical Framework

Without a loss of generality, we will assume that the multi-objective optimization problem involves the simultaneous minimization of M objective functions, $f_m : \Omega \rightarrow R^+$, $m = 1, 2, M$, where Ω is the set of all feasible solutions, or the search space of the multi-objective problem, and its image $[f_1(\Omega), f_M(\Omega)]$ is the objective function space [62], [63].

Consider two solutions, u, v in Ω . We say that u dominates v iff, for each objective m , $f_m(u) \leq f_m(v)$, and for at least one objective m , the inequality is strict, i.e. $f_m(u) < f_m(v)$. This relationship is written as $u \prec v$. Given any subset of solutions, S in Ω , the non-dominated set of S , N_S , is the subset of all solutions in S that are not dominated. The non-dominated set of Ω , N_Ω , is called the *Pareto set*. Its image in the objective function space is called the *Pareto front*.

6.2.2 Fuzzy Dominance

Unlike the previous approach which compares solutions that are Pareto-optimal, the method proposed in [64] considers a fuzzy approach to compare arbitrary solutions, i.e. ones that need not necessarily be Pareto-optimal. In order to apply this approach, it is necessary to first define a set of M monotonically non-decreasing membership functions, one for each objective function. Let these functions be $\mu_m : R \rightarrow [0, 1]$, $m = 1, 2, \dots, M$. Any solution u is said to m-dominate another

solution v , if and only if the condition $f_m(u) < f_m(v)$ is met. This relationship will be denoted as $u \prec_m v$. The degree of fuzzy m-dominance is equal to $\mu_m(f_m(v) < f_m(u)) \equiv \mu_m(u \prec^m v)$. Clearly, $u \prec v$ if a solution if for each m , $u \prec_m v$. The degree of fuzzy dominance can be by applying fuzzy intersection, using any t-norm:

$$\mu(u \prec v) = \bigcap_{m=1}^M \mu_m(u \prec^m v) \quad (6.1)$$

Equation 6.1 is equivalent to a t-co norm and is carried out over all M objectives. Next, we establish the concept of hierarchies between objectives to discriminate more important objectives from the less important ones. For hierarchy of objectives with L number of levels, $f(\cdot)$ and $g(\cdot)$ distinguish between primary and secondary objectives, we denote as $f^l(\cdot)$, $l = 1, 2, L$, to be an objective function in the l th level. Hence, in terms of this notation, the primary objectives are $f^1(\cdot)$, the secondary ones are $f^2(\cdot)$, and so on. When a solution u dominates another one, v along all the objectives $f^l(\cdot)$ of the l th level, we denote the relationship as $u \prec_l v$. In other words, for each objective m , $f_m^l(u) \leq f_m^l(v)$, $m = 1, 2, \dots, M_l$, where M_l is the number of level l objectives.

Next we introduce the idea of domination up to a level. This will be defined in a recursive manner. We say that u dominates v up to level l , if it u dominates v for the level l objectives, and if there exist other levels below in the hierarchy, u dominates v up to level $(l + 1)$ as well. In other words Equation 6.2 holds.

$$\begin{aligned} u \prec_{l+} &\equiv (u \prec_l v) \vee ((u \prec_l v) \wedge (u \prec_{(l+1)+} v)) \text{ for } l < L \\ &\equiv u \prec_L v \text{ for } l = L \end{aligned} \quad (6.2)$$

In the above equation, $u \prec_{l+} v$ denotes the relationship u dominates v up to level $l + 1$. Letting $\mu(u \prec_{L+} v)$ simply be equal to $\mu(u \prec_L v)$, the fuzzy membership can be computed readily from the above definition accordingly as Equation 6.3

$$\mu(u \prec_{l+} v) = \mu(u \prec_l v) \cup (\mu(u \prec_l v) \cap \mu(u \prec_{(l+1)+} v)) \quad (6.3)$$

We consider $u \prec v$ only if $u \prec_{1+} v$. Therefore, $\mu(u \prec v)$ is taken to be $\mu(u \prec_{1+} v)$. We next turn our attention to selecting a representative from any given solution sample S.

6.2.3 Mediality

Given a non-dominated set of solutions, S , in many applications, it is important to isolate one or more samples in it, that may serve as representative of the entire set. A method to apply fuzzy logic to identify such a solution is proposed. This method extends the concept of a median to multiple dimensions. Given any set of solutions S , let $S_m = u_i$ be the ordered set of the solutions in S , sorted along their values for the m th objective. We define the m -mediality of any solution u_i using Equation 6.4.

$$d_m(u_i) = \min(i, |S| - i) \quad (6.4)$$

where $|S|$ denotes set cardinality. It can be seen that the median in S_m is the one that has the highest value of m -mediality, at $fl(|S|)$, where $fl()$ is the floor function. The mediality of any solution is simply given Equation 6.5,

$$d(u_i) = \min_m(d_m(u_i)) \quad (6.5)$$

The minimum in Equation 6.5 is carried out over all M objectives. Under these circumstances, the median solution is given by Equation 6.6

$$med(S) = argmax(d(u_i)) \quad (6.6)$$

The above concepts can be extended using a fuzzy logic framework. Given monotonically non-decreasing fuzzy membership functions, $\sigma_m : R \rightarrow [0, 1], m = 1, 2, , M$, we define the fuzzy m -mediality of solution u_i to be $\sigma_m(d(u_i)) \equiv \sigma_m(u_i)$. In a manner similar to fuzzy dominance, the fuzzy mediality of the solution and the median solution can be defined using Equation 6.7.

$$\sigma(u_i) = \cap_{m=1}^M \sigma_m(u_i) \quad (6.7)$$

$$med(S) = argmax(\sigma(u_i)) \quad (6.8)$$

6.2.4 Evolutionary Algorithm

We used the fuzzy dominance based genetic algorithm [64] that makes use of the concept of fuzzy dominance. Unlike other multi-objective approaches that assume a crisp dominance measure for Pareto ranking, fuzzy dominance is adopted here. The algorithm considers fuzzy dominances between all solutions in the population to compute a measure reflecting its overall fitness, with nonzero solutions being assigned zero values. This allows ranking the individuals and selecting the non-dominated solutions based on their fuzzy dominance measure.

At each generation, binary tournament selection is used to selects candidate solutions for recombination. We have used a simple convex crossover and polynomial mutation for this problem. Elitism is used to archive the best solution candidates (Pareto front solutions): any other solution candidate in the population does not dominate them. The elite is not subjected to crossover and mutation but re-introduced into subsequent population for evaluation, to ensure that it is still the best solution candidate. The process is iterated for a desired number of generations.

We explore the application of the fuzzy dominance based evolutionary optimization algorithm in two cases. In Case-I, all objectives are treated equally as in a traditional multi-objective setting, and fuzzy dominance (as in [64]) is used. These four objectives are: f_{NI} , f_{DC} , f_{NC} , and f_{VC} .

In the second case (Case-II), the fuzzy dominance is calculated using the hierarchical dominance described earlier. Keeping in view the earlier considerations, we apply three hierarchies.

The genetic algorithm is applied using fuzzy dominance without hierarchical dominance and with hierarchical dominance. Without hierarchical dominance, all the four objectives have equal importance of minimization. While implementing hierarchical dominance, we assume the following hierarchy:

1. Primary Objective (Highest degree of importance): Objective 1 (f_{NI}).
2. Secondary Objective (Second highest degree of importance): Objective 2 (f_{DC}).
3. Tertiary Objective (Third highest degree of importance): Objective 3 (f_{NC}) and Objective 4 (f_{VC}).

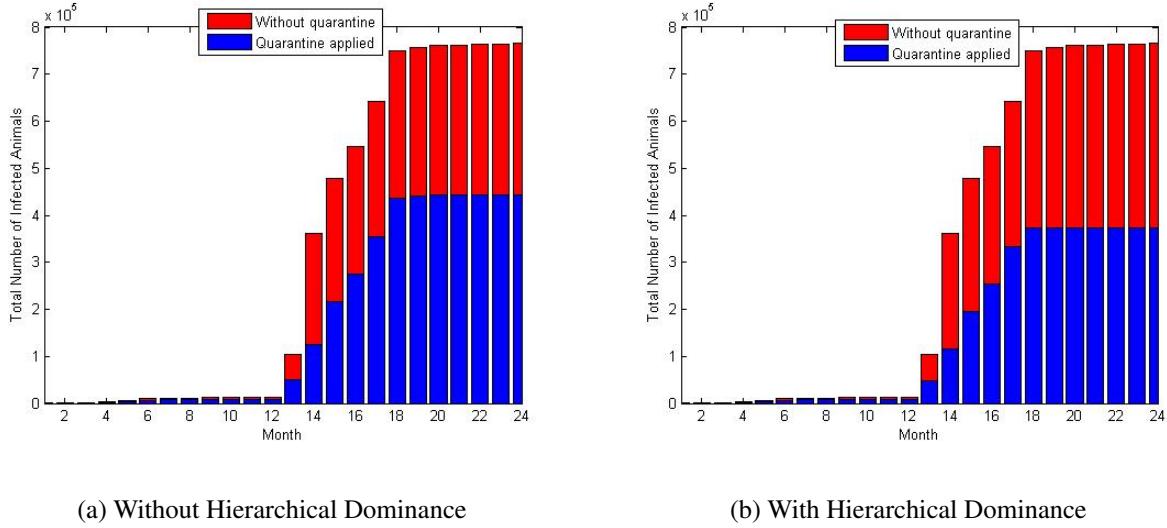


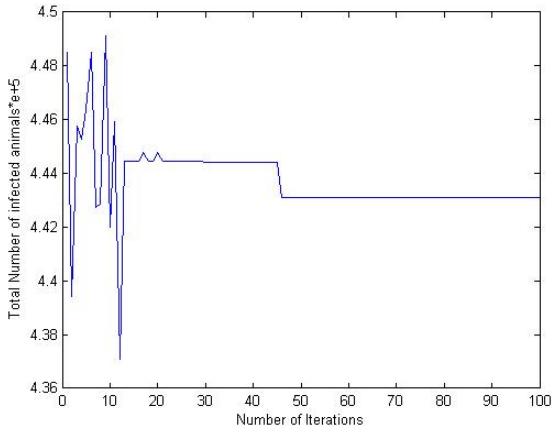
Figure 6.1: Reduction in the Total Number of Infected animals due to mitigation strategies without and with hierarchical dominance.

Due to the extremely high computational overheads (of the order of a several hours) each approach is run 5 times and the average is reported. The population size is 100, and the total number of generations is fixed at 50. Fuzzy mediality is used to extract representative solutions, hereafter referred to as the medial solution. No distinction is made here between the primary, secondary and tertiary objectives.

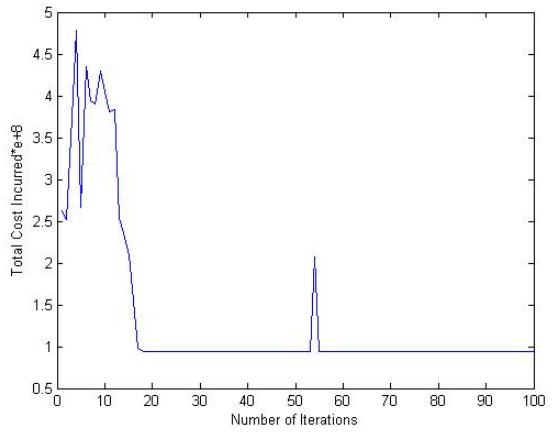
6.3 Results

Using the concept of fuzzy mediality, we trace the rate of convergence of the medial solution in the set of 100 solutions along every iteration. The expected observation would be faster convergence of primary objective (f_{NI}) and secondary objective (f_{DC}) using hierarchical fuzzy dominance than without hierarchy. However, the tertiary objectives (f_{NC} and f_{VC}) have a slower rate of convergence using hierarchical fuzzy dominance than without hierarchy.

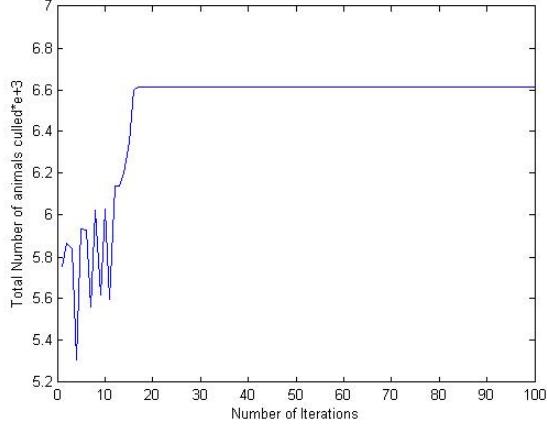
The medial best solution from the final solution archive after 100 runs of iteration is considered to be the optimally best solution. The impact of these optimally best solutions in impeding the spread of FMD virus is depicted in Figure 6.1.



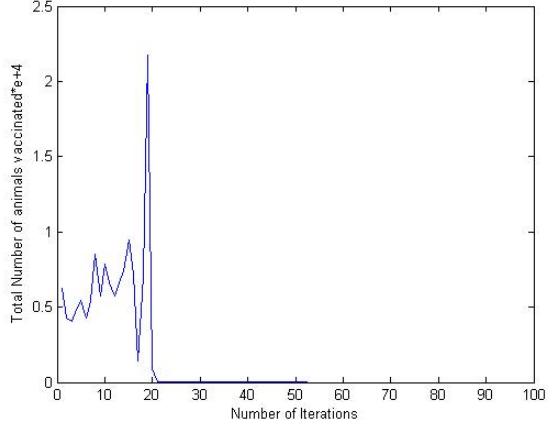
(a) f_{NI} : Total number of infected animals.



(b) f_{DC} : Total cost incurred due to mitigation strategies.



(c) f_{NC} : Total number of animals culled.

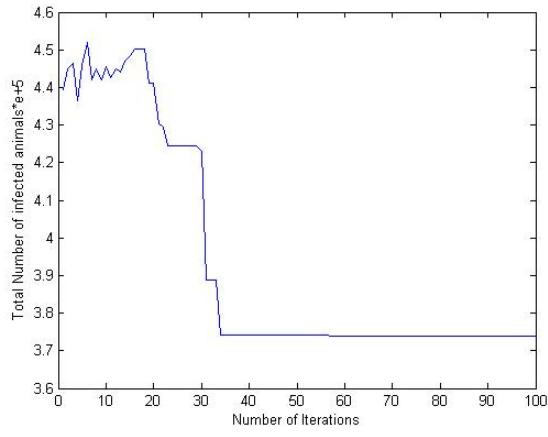


(d) f_{VC} : Total numbers of animals vaccinated.

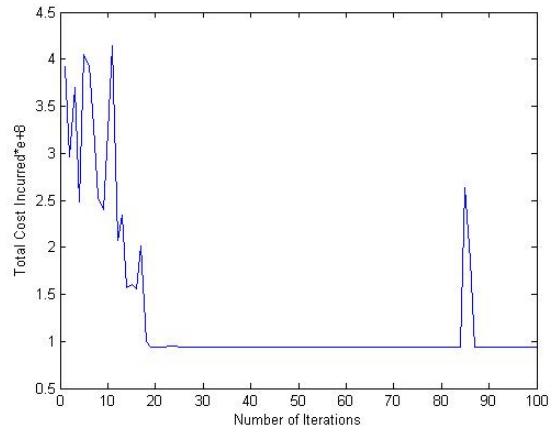
Figure 6.2: Optimization of f_{NI} , f_{DC} , f_{NC} and f_{VC} without hierarchical dominance

Initially, we observe the rate of reduction of the medial solution along the four objectives without using the concept of hierarchical dominance in Figure 6.2.

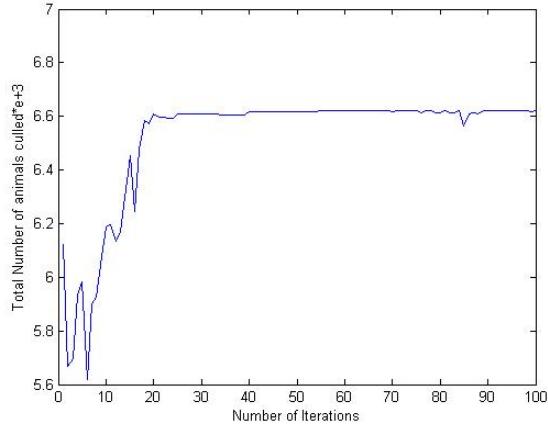
Next, we observe the rate of reduction of the medial solution along the four objectives using the concept of hierarchical dominance as Figure 6.3. Thus we infer from Figure 6.2 and Figure 6.3, that the rate of convergence for f_{NI} and f_{DC} is faster with hierarchy and the rate of convergence for f_{NC} and f_{VC} is faster without hierarchy. The final observations are tabulated below.



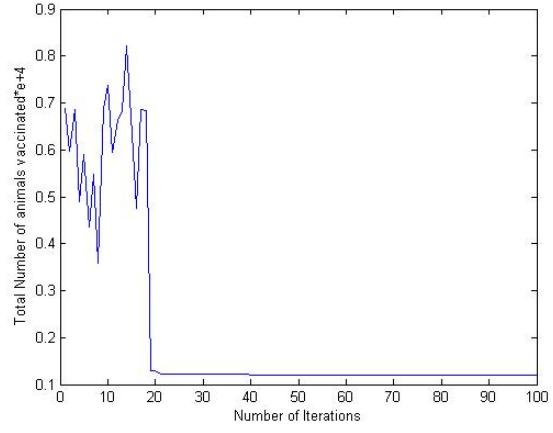
(a) f_{NI} : Total number of infected animals.



(b) f_{DC} : Total cost incurred due to mitigation strategies.



(c) f_{NC} : Total number of animals culled.



(d) f_{VC} : Total numbers of animals vaccinated.

Figure 6.3: Optimization of f_{NI} , f_{DC} , f_{NC} and f_{VC} with hierarchical dominance

Table 6.1: Optimization of 4 objectives

FDGA	%Reduction ¹	f_{NI}	f_{DC}	f_{NC}	f_{VC}
No Hierarchy	42.203%	443095	93.483	6611	0
Hierarchy	51.241%	373816	93.298	6622	1208

From Table 6.1, we infer that using hierarchical fuzzy dominance, the set of optimal threshold probabilities thus obtained bring about a greater reduction in the total number of infected animals than without hierarchy at the end of the observation period. Also, the cost incurred due to the implementation of mitigation strategies using hierarchical fuzzy dominance is lower than that without hierarchy. However, the hierarchical fuzzy dominance method focuses on optimal vaccination, and hence it results in a lot of susceptible livestock getting vaccinated. On the other hand, non-hierarchical fuzzy dominance method relies greatly on movement bans than on vaccinations.

Thus, we have considered a real-world problem regarding the spread of the FMD virus, thus resulting in an epidemic. We have introduced the concept of fuzzy mediality and verified its performance theoretically. Next we studied the optimization of medial solution to achieve an optimal mitigation strategy for FMD epidemics with and without the concept of hierarchical fuzzy dominance.

In conclusion, fuzzy dominance method provide an optimal solution to impede the spread of FMD while reducing the economic losses also. While non-hierarchical dominance based fuzzy optimization relies greatly on movement restrictions to impede the spread if infection, hierarchical fuzzy dominance suggests vaccination as a better policy. Both methods provide a cost optimal reduction in the impact of the foot and mouth disease. The approach suggested here applies multi-objective evolutionary optimization to obtain optimal values for the monthly thresholds to minimize the extent of foot and mouth disease. Instead of applying a traditional method of optimization, the new method proposed here, fuzzy hierarchical dominance, considers various levels of objectives, where objectives at a higher level are more important to optimize. The approach is tested to obtain optimal mitigation strategies, and the results show that the application of a hierarchy produces more optimal threshold values.

As future implementation, we need to explore the impact of hierarchical dominance on optimization of other mitigation strategies described in Chapter 5. Besides other data sets need to analyzed too for more conclusive results.

Chapter 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

The primary contribution of this work is the modeling of FMD dynamics in countries such as Turkey, Iran and Thailand, where FMD outbreaks have been inadequately recorded, reported and insufficiently modeled as compared to countries like the United Kingdom where extensive study has been done to predict and control FMD. To achieve modeling of disease dynamics in such countries with undeveloped databases, we characterize a meta-population based stochastic model that predicts the probability of infection at different locations in discrete time steps. The novelty of this model is that it considers a fully connected heterogeneous contact network between locations denoted by nodes in space, such that the heterogeneous interaction between locations can be depicted in terms of weights on links of the contact network. These weights are characterized by wind and human interventions between the nodes. Additionally, the parameterization of the spread of FMD dynamics is achieved by considering the grazing patterns of infected livestock, the species of livestock, and the number of meat markets and slaughter houses in the different regions.

Our mathematical predictive model is optimally calibrated and validated using the FMD infection data from Turkey (January 2005 through December 2006), Iran and Thailand (January 2005 through December 2005) from the FMD Bioportal. These data sets were selected for study due to their non-sparse nature among other confidential epidemic incidence data sets. The spatial and temporal predictability of the model is compared to that of unweighted and unparameterized

baseline models. Additionally, the contribution of wind in the spread of FMD is also studied. Our study lays foundation to the importance of wind in long distance spread of FMD virus. Although the weights and parameterizations increase precision in spatio-temporal predictability, the contribution of wind needs to be studied more extensively on other data sets in future to infer the effectiveness of its contribution. However, for the current data set on Turkey, Iran and Thailand, the predictive performance of our weighted network based model is noteworthy.

Next, we apply learning-based models to mimic the spatial characteristics of network based spread models. Our study shows that it is possible to generalize local learning-based models such as neural network, autoregressive, Bayesian network and Monte-Carlo simulation models to recover the latent spatial parameters in it. Learning-based models may offer an improvement over spatial models in epidemic predictions in the absence of data regarding spatial characteristics.

Additionally, we use the network-based and learning-based prediction models to simulate cost-effective mitigation strategies. Different mitigation tasks such as Infected Premise (IP) culls, Dangerous Contact (DC) culls, and Vaccination (V) and No Movement (NM) restrictions are simulated and their importance in impeding the spread of the FMDV is analyzed. Conclusively, we observe that IP and DC culls followed up by vaccinations result in a cost-effective reduction of the total number of infected livestock, and this strategy retards the rapid spread of the FMDV. However, the cost incurred in this process is quite high. Hence, this mitigation strategy may be preferred after the onset of an epidemic outbreak when immediate reduction in the number of infected livestock is mandatory. However, though a potent vaccination strategy, followed by movement restrictions, incurs a lower reduction in the total number of infected livestock, this strategy results in the most cost-effective control of the epidemic. Thus, vaccinations and movement ban strategies may be adopted before the onset of an epidemic outbreak to control the rapid spread of the FMDV.

Further, the variation in cost-effectiveness of various mitigation strategies with prediction errors from baseline network-based models and learning-based models provide insights into the robustness of our simulative mitigation strategies. Such downstream utility impacts of prediction errors in mitigation strategies analyze the non-linear relation between prediction error and

mitigation effectiveness variations.

As a variation to the simulative study of various mitigation strategies, we apply fuzzy-dominance genetic algorithms implemented using hierarchies between the objective functions to minimize three levels of objective functions. The hierarchical dominance is achieved by minimizing the total number of infected livestock at the end of the period under study as a primary objective, minimizing the direct cost of implementing these mitigation strategies as a secondary objective and minimizing the total number of unethical culls and vaccinations as the tertiary objectives. Conclusively, we can say that hierarchical dominance provides better reduction in the impact of an FMD outbreak than that compared to non-hierarchical optimizations wherein all the objectives are treated to be of equal priority.

It is noteworthy that the proposed simulative model W is iterative, and the time steps can be varied to suit the predictive requirements. Besides being efficient in epidemic predictions, the model can be used to simulate dynamic mitigation and control policies to curb unethical culls and reduce the economic adversities of FMD. Additionally, the proposed model is geographically independent, and it can be used to model the disease dynamics at other countries with undeveloped databases. However, there is a lot of scope for future work to complement our results and findings, some of which are given below.

7.2 Future Work

The spatial network based prediction model can be further improvised by parameterizing the spread of FMD virus in water since it is known that although the FMD virus spreads up to 40 miles in air, it spreads up to 180 miles in water. Since direction of water flow would be consistent, it would be interesting to notice the improvement in epidemic prediction due to the impact of flowing water bodies. Though, this addition in parameterization would require more granular data regarding the location and routes of water bodies, but such a parameterization might enhance the predictive performance considerably. Also, further analyses of the wind patterns by using wind-plume models in Turkey, Iran and Thailand will give better insights into the impact of wind in

FMDV spread.

Next, the prediction performance of the learning-based temporal prediction models may be enhanced for improved portability and enhanced global outlook. Besides, the predictive performance of other learning-based local models such as recurrent networks, moving average and wavelet-based prediction models to name a few may be analyzed for improved prediction performance. Additionally, a dynamic system that requires local training but which is capable of performing as good as the global prediction models is required, and this can be an interesting topic for future research. This would mean a system that contains network-based spread models as well as learning-based models inbuilt into it, such that, according to the amount of available data regarding global spread parameters, a suitable predictive model may be dynamically selected to provide infection incidence predictions.

The simulative mitigation strategies studied here may be altered and additional combinations of mitigation tasks may be studied. Additionally, the impact of delays in implementation of mitigation tasks, such as a delay in development and deployment of vaccines, behavioral delays due to in compliance with movement ban policies, or in compliance to DC culling strategies, may provide better insights into the critical nature of the mitigation strategies. Also, since we observe a notably high cost-effectiveness of mitigation strategies involving vaccination and movement bans in Turkey, Iran and Thailand, we may conclude that training of personnel, providing a good transportation infrastructure, and enhancing research facilities for quicker development of vaccines may be beneficial for long term eradication of FMD instead of suppressing the epidemic for a short term by imposing massive culling strategies.

Also, the concept of ergodicity in dynamic infection networks that has been simulative established in this work may be mathematically implied in future. The predictive output of our proposed network-based model along with the suggested mitigation strategies may be evaluated in comparison with the FMD simulator developed by USDA. Most importantly, a good visualization platform may be developed to enhance the implications from our model, and to present our predictive model in a better way in the future.

Finally, the optimal mitigation strategies developed in this work may be studied in detail for the data sets of Iran and Thailand in future to yield more conclusive results regarding the optimal mitigation strategies. So far, we have analyzed only the mitigation strategy 1 that combines IP culls, DC culls, vaccination and movement ban strategies using hierarchical and non-hierarchical Fuzzy Dominance Genetic Algorithm (FDGA). These results may be useful in studying the other mitigation strategies 2 through 6, to assess the optimality induced by FDGA in each of them. Also, the comparative study of FDGA versus NSGA2, may provide more conclusive results regarding the optimal mitigation strategies.

Bibliography

- [1] R. P. Kitching, A. M. Hutber, and M. V. Thrusfield, “A review of foot-and-mouth disease with special consideration for the clinical and epidemiological factors relevant to predictive modelling of the disease,” in *The Veterinary Journal*, vol. 169, 2005.
- [2] Origin of the UK Foot and Mouth Disease Epidemic 2001. [Online]: <http://www.defra.gov.uk/footandmouth/pdf/fmdorigins1.pdf>.
- [3] Foot-and-mouth virus: A global dilemma. [Online]: <http://archives.cnn.com/2001/WORLD/europe/02/27/farming.world/index.htm>.
- [4] Foot and Mouth Disease-FMD. [Online]: <http://www.eden.lsu.edu>.
- [5] US Department of Agriculture. Foreign Animal Report 1998: p41. [Online]: <http://www.aphis.usda.gov/lpa/pubs/fadrep.pdf>.
- [6] M. Woolhouse, “Foot-and-mouth disease in the uk: what should we do next time?,” in *J. Appl. Microbiol.*, pp. 126–130, 2003.
- [7] M. Woolhouse, D. Haydon, A. Pearson, and R. Kitching, “Failure of vaccination to prevent outbreaks of foot-and-mouth disease,” in *Epidemiol. Infect.*, pp. 363–371, 1996.
- [8] F. Moutou, “Epidemiological basis useful for the control of foot-and-mouth disease,” in *Comp. Immunol. Microbiol.*, pp. 321–330, 2002.
- [9] T. W. Bates, T. E. Carpenter, and M. C. Thurmond, “Benefit-cost analysis of vaccination and preemptive slaughter as a means of eradicating foot-and-mouth disease,” in *American journal of veterinary research*, vol. 64, pp. 805–812, 2003.

- [10] T. W. Bates, T. E. Carpenter, and M. C. Thurmond, “Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease,” in *American Journal of Veterinary Research*, vol. 64, pp. 205–210, 2003.
- [11] T. W. Bates, T. E. Carpenter, and M. C. Thurmond, “Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease,” in *American Journal of Veterinary Research*, vol. 64, pp. 195–204, 2003.
- [12] T. Carpenter, M. Thurmond, and T. Bates, “A simulation model of intraherd transmission of foot and mouth disease with reference to disease spread before and after clinical diagnosis,” in *J Vet Diagn Invest*, vol. 16, pp. 11–16, 2004.
- [13] N. Ferguson, C. Donnelly, and R. Anderson, “The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions,” in *Science*, vol. 292, pp. 1155–1160, 2001.
- [14] N. Ferguson, C. Donnelly, and R. Anderson, “Transmission intensity and impact of control policies on the foot and mouth epidemic in great britain,” in *Nature*, vol. 413, pp. 542–548, 2001.
- [15] N. Ferguson, M. Keeling, W. Edmunds, R. Gani, B. Grenfell, R. Anderson, and S. Leach, “Planning for smallpox outbreaks,” in *Nature*, vol. 425, pp. 681–685, 2003.
- [16] M. J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. . Wilesmith, and B. T. Grenfell, “Dynamics of the 2001 uk foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape,” in *Science*, vol. 294, 2001.
- [17] M. Keeling, M. Woolhouse, R. May, G. Davies, and B. Grenfell, “Modelling vaccination strategies against foot and mouth disease,” in *Nature*, vol. 421, pp. 136–142, 2003.

- [18] R. Morris, J. Wilesmith, M. Stern, R. Sanson, and S. M.A., “Predictive spatial modeling of alternative control strategies for the foot-and-mouth disease epidemic in great britain,” in *Vet. Rec.*, 2001.
- [19] S. Roy Chowdhury, C. Scoglio, and W. Hsu, “Evolution and control strategies of the foot and mouth disease epidemic on a weighted contact network,” in *EPIDEMICS2*, Elsevier, p. P2.07: Abstract, 2009.
- [20] M. Keeling, “Models of foot-and-mouth disease,” in *Proceedings of the Royal Society of London B*, vol. 272, pp. 1195–1202, 2005.
- [21] S. Alexandersen, Z. Zhang, A. I. Donaldson, and A. J. M. Garland, “The pathogenesis and diagnosis of foot-and-mouth disease,” in *Journal of Comparative Pathology*, vol. 129, pp. 1 – 36, 2003.
- [22] Foot and mouth Disease. [Online]: <http://en.wikipedia.org/wiki/Foot-and-mouth-disease-Taiwan>.
- [23] S. M. Crispin, “Foot-and-mouth disease: The vital need for collaboration as an aid to disease elimination,” *The Veterinary Journal*, vol. 169, no. 2, pp. 162 – 164, 2005.
- [24] J. Guitian and D. Pfeiffer, “Should we use models to inform policy development?,” in *The Veterinary Journal*, vol. 172, pp. 393 – 395, 2006.
- [25] M. Kobayashi, T. Carpenter, B. Dickey, and H. R.E, “A dynamic, optimal disease control model for foot-and-mouth disease:. i. model description,” in *Preventive Veterinary Medicine*, vol. 79, pp. 257–273, 2007.
- [26] J. Globster, H. J. Champion, J. H. Sorensen, T. Mikkelsen, D. B. Ryall, P. Astrup, S. Alexandersen, and A. I. Donaldson, “Airborne transmission of foot-and-mouth disease virus from burnside farm, heddonon-the-wall, northumberland, during the 2001 epidemic in the united kingdom,” in *The Veterinary Record*, vol. 152, pp. 525–533, 2003.

- [27] M. Garner and R. Cannon, “A report prepared for the australian meat research corporation,” in *Bureau of Resource Sciences*, p. 98, 1995.
- [28] Turkish Statistical Institute. [Online]: <http://www.turkstat.gov.tr/Start.do>.
- [29] Windfinder—Real time wind and weather charts Turkey. [Online]: http://www.windfinder.com/windreports/windreports_online_tr.htm.
- [30] M. Tufan, “Report of the foot-and-mouth situation in turkey from 1990 to 2002,” in *Consultant’s report to the Food and Agriculture Organisation of the United Nations*, p. 44, 2003.
- [31] K. Eames, J. Read, and W. Edmunds, “Epidemic prediction and control in weighted networks,” in *Epidemics*, vol. 1, pp. 70–76, 2009.
- [32] P. Schumm, C. Scoglio, D. Gruenbacher, and T. Easton, “Epidemic spreading on weighted contact networks,” in *Bionetics*, pp. 201–208, 2007.
- [33] L. Sander, C. Warren, and I. Sokolov, “Epidemics, disorder, and percolation,” in *Physica A*, vol. 325, pp. 1–8, 2003.
- [34] G. Chowell, J. M. Hyman, S. Eubank, and C. Castillo-Chavez, “Scaling laws for the movement of people between locations in a large city,” in *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, 2003.
- [35] Deviance Information Criterion. [Online]: <http://en.wikipedia.org/wiki/Devianceinformationcriterion>.
- [36] S. R. Chowdhury, C. Scoglio, and W. Hsu, “Simulative modeling to control the foot and mouth disease epidemic,” in *International Conference on Computational Science*, vol. In press, 2010.
- [37] FMD Bioportal [Online]: <https://fmdbioportal.ucdavis.edu>.

- [38] R. Kao, L. Danon, D. Green, and I. Kiss, “Demographic structure and pathogen dynamics on the network of livestock movements in great britain,” in *Proc. R. Soc. Lond. B*, vol. 273, 2006.
- [39] R. R. Kao, D. M. Green, J. Johnson, and I. Z. Kiss, “Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the uk,” in *J R Soc Interface*, vol. 4, p. 907916, 2007.
- [40] S. S. R. Abidi1 and A. Goh, *Applying knowledge discovery to predict infectious disease epidemics*. Springer Berlin / Heidelberg, 2006.
- [41] A. Hawksworth, C. Hansen, P. Good, M. Ryan, K. Russell, P. Kelley, and J. Gaydos, “Using autoregressive epidemic modeling to augment the existing department of defense (dod) febrile respiratory illness surveillance system at military training centers,” in *J Urban Health*, vol. 80, p. i38:Abstract, 2003.
- [42] C. Lagazio, E. Dreassi, and A. Biggeri, “A hierarchical bayesian model for spacetime variation of disease risk,” in *Statistical Modelling*, vol. 1, p. 1729, 2001.
- [43] X. Jiang and G. Wallstrom, “A bayesian network for outbreak detection and prediction,” in *AAAI*, 2006.
- [44] G. Cooper, D. H. Dash, W. Levander, J.D.and Wong, W. Hogan, and M. Wagner, “Uncertainty in artificial intelligence,” in *ICALP*, 2004.
- [45] M. Kobayashi, B. Dickey, T. Carpenter, and H. RE., “A dynamic optimal disease control model for foot-and-mouth disease:ii. model results and policy implications,” in *Prev Vet Med*, vol. 79, pp. 274–286, 2007.
- [46] L. Franci and S. Panigrahi, “Artificial neural network models of wheat leaf wetness,” in *Agricultural and Forest Meteorology*, vol. 88, pp. 57–65, 1997.

- [47] Y. Chtioui, S. Panigrahi, and L. Franci, “A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease,” in *Chemometrics and Intelligent Laboratory Systems*, vol. 48, pp. 47–58, 1999.
- [48] S. Alexandersen, Z. Zhang, A. I. Donaldson, and A. J. M. Garland, “The pathogenesis and diagnosis of foot-and-mouth disease,” in *Transactions of the ASAE*, vol. 40, pp. 247–252, 1997.
- [49] C. Viboud, B. Pierre-Yves, F. C. V. Alain-Jacques, and A. Flahault, “Prediction of the spread of influenza epidemics by the method of analogues,” in *American Journal Epidemiology*, vol. 158, pp. 996–1006, 2003.
- [50] T. Krishnamurti, V. Chakraborti, A. and Mehta, and M. A.V., “Experimental prediction of climate-related malaria incidence,” in [Online] <http://www.pitt.edu/super2/29011-30001/29021.ppt>, 2007.
- [51] O. Brit, P. Vounatsou, D. Gunawardena, G. Galappaththy, and P. Amerasinghe, “Models for short term malaria prediction in sri lanka,” in *Malaria Journal*, vol. 7, 2008.
- [52] D. Lai, “Monitoring the sars epidemic in china: time series analysis,” in *Journal of Data Science*, vol. 3, p. 279 293, 2005.
- [53] T. Abeku, S. Hay, S. Ochola, P. Langi, B. Beard, S. De Vlas, and J. Cox, “Malaria epidemic early warning and detection in african highlands,” in *Trends in Parasitology*, vol. 20, pp. 400–405, 2004.
- [54] B. M. Krueger, Robert F.and Hicks, C. J. Patrick, S. R. Carlson, and M. Iacono, William G.and McGue, “Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum,” in *Journal of Abnormal Psychology*, vol. 111, pp. 411–424, 2002.

- [55] M. Gui, “Advanced methods for prediction of animal-related outages in overhead distribution systems,” in *K-State Research Exchange [Online]*: <http://hdl.handle.net/2097/1510>, 2009.
- [56] Autoregressive model. [Online]: <http://en.wikipedia.org/wiki/Autoregressivemodel>.
- [57] I. Anderson, “Foot and mouth disease 2001: Lessons to be learned inquiry,” in *The Stationary Office, London*, vol. 187, 2002.
- [58] M. E. J. Newman and D. J. Watts, “Scaling and percolation in the small-world network model,” in *Phys. Rev. E*, vol. 60, pp. 7332–7342, American Physical Society, Dec 1999.
- [59] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Network robustness and fragility: Percolation on random graphs,” in *Phys. Rev. Lett.*, vol. 85, pp. 5468–5471, American Physical Society, Dec 2000.
- [60] R. Albert and A.-L. Barabási, “Topology of evolving networks: Local events and universality,” in *Phys. Rev. Lett.*, vol. 85, pp. 5234–5237, American Physical Society, Dec 2000.
- [61] M. J. Tildesley, N. J. Savill, D. J. Shaw, R. Deardon, S. P. Brooks, M. E. J. Woolhouse, B. T. Grenfell, and M. J. Keeling, “Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the uk,” in *Nature*, vol. 440, pp. 83–86, 2005.
- [62] S. Das and B. Panigrahi, “Multi-objective evolutionary algorithms,” in *Encyclopedia of Artificial Intelligence*, (Eds. J. R. Rabunal, J. Dorado and A. Pazos), Idea Group Publishing, vol. 3, pp. 1145 – 1151, 2008.
- [63] S. Das, B. K. Panigrahi, and S. Patnaik, “Nature-inspired algorithms for multi-objective optimization,” in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques*, (Eds. E. S. Olivas, J. D. Martin-Guerrero, M. Martinez, R. Magdalena, A. J. Serrano), IGI Global, vol. 1, pp. 95 – 104, 2009.

- [64] P. Koduru, Z. Dong, S. Das, S. M. Welch, and J. Roe, “Multi-objective evolutionary-simplex hybrid approach for the optimization of differential equation models of gene networks,” in *IEEE Transactions on Evolutionary Computation*, vol. 12, pp. 572 – 590, 2008.
- [65] 2009 H1N1 Flu [Online]: <http://www.cdc.gov/H1N1FLU>.

Appendix A

APPENDIX: TEMPORAL PREDICTIONS OF H1N1 IN USA

Background: Rapidly changing trends in the incidence of strain A-H1N1 have made long term predictions infeasible. Based on the weekly Influenza like Illness (ILI) incidence reports from the Centre of Disease Control and Prediction (CDC), we are able to predict the ILI at 10 different regions of the US and at a National level for 2 consequent weeks and seasonality of ILI for 4 consecutive weeks within a 95% confidence interval. Also, the rate of mortality is estimated for 6 major cities in the mid-western region of the US.

Methods: The selected temporal prediction models include: Time-Series, Neural Networks for ILI predictions, and Bayesian Networks, Monte-Carlo Simulations for mortality predictions. Weighted ILI incidence reports from weeks 40 in the year 2008 till week 41 of year 2009, and mortality reports from week 1 in 1997 till week 41 in 2009 are used as data. ILI trends are estimated using differenced time-series models, while ILI weekly incidences are predicted using neural networks. Bayesian Networks and Monte-Carlo Simulations predict the number of mortal casualties within a 95% confidence interval due to Pneumonia and Influenza.

Results, Discussion: While the time-series models have the lowest AICC/BIC values, neural network models are characterized by coefficient of determination (R^2) values in the range of (0.8261-0.9811) and symmetric mean absolute percentage error (SMAPE) in the range (3.492-12.289). ILI predictions at a National level for the US shows 8.21% and 8.07% ILI in week 42

and 43 in the year 2009 with an increasing trend till week 42, followed by a slight decreasing trend till week 46. Mortality in the Midwest is estimated to be in the range of (5.67-14.8) deaths per week for most major cities, while as many as 2.74 deaths per week are estimated in Kansas City till week 46 in the year 2009.

Temporal prediction models are very efficient in predicting the speed of spread of diseases, like the the strain A-H1N1 virus. Such models are simplistic in computation, involve very low parameterization and incur very low computational complexities. Besides, temporal prediction models are not only capable of predicting the time when a disease will transition from the pre-outbreak to the outbreak phase, but they also predict the mortality rate in the following time steps. This information is very useful to plan cost-optimal mitigation strategies such as travel restrictions and vaccinations with varying acuteness in different phases of the epidemic.

A.1 Predictions regarding Influenza Like Illness

Influenza-like illness (ILI), also known as acute respiratory infection (ARI) and flu-like syndrome, is a medical diagnosis of possible influenza or other illness causing a set of common symptoms. Symptoms commonly include fever, shivering, chills, malaise, dry cough, loss of appetite, body aches and nausea, typically in connection with a sudden onset of illness. In most cases, the symptoms are caused by cytokines released by immune system activation.

Outpatient Illness Surveillance if information on patient visits to health care providers for influenza-like illness is collected through the US Outpatient Influenza-like Illness Surveillance Network (ILINet). The Outpatient Influenza-like Illness Surveillance Network (ILINet) consists of more than 3,000 health care providers in all 50 states, the District of Columbia and the U.S. Virgin Islands reporting over 25 million patient visits each year. Each week, approximately 1,400 outpatient care sites around the country report data to CDC on the total number of patients seen and the number of those patients with influenza-like illness (ILI) by age group. For this system, ILI is defined as fever (temperature of 100F [37.8C] or greater) and a cough and/or a sore throat in the absence of a KNOWN cause other than influenza. Sites with electronic records use an equivalent

definition as determined by the state public health authorities [65]. The percentage of patient visits to health care providers for ILI reported each week is weighted on the basis of state population. This percentage is compared each week with the national baseline of 3.3%. The baseline is the mean percentage of patient visits for ILI during non-influenza weeks for the previous three seasons plus two standard deviations. Due to wide variability in regional level data, it is not appropriate to apply the national baseline to regional data, therefore, region specific baselines are calculated. [65] Regional baselines for the 2009-10 influenza season are:

1. Region 1 - Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
2. Region 2 - New Jersey, New York, Puerto Rico, and the U.S. Virgin Islands
3. Region 3 - Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia
4. Region 4 - Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee
5. Region 5 - Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin
6. Region 6 - Arkansas, Louisiana, New Mexico, Oklahoma, and Texas
7. Region 7 - Iowa, Kansas, Missouri, and Nebraska
8. Region 8 - Colorado, Montana, North Dakota, South Dakota, Utah, and Wyoming
9. Region 9 - Arizona, California, Guam, Hawaii, and Nevada
10. Region 10- Alaska, Idaho, Oregon, and Washington

A.2 Temporal Prediction Models

We have looked at 4 temporal Prediction models in our work. Since the ILI has a fluctuating trend and seasonality, we have used Time-Series Predictions and Feed-Forward Neural Networks to predict the ILI in upcoming 20 weeks following week 40 in the year 2009. Mortality due to Influenza and Pneumonia in the mid-western part of the United States is predicted using Bayesian Networks and Monte-Carlo simulations. The source of data, for predicting the number of fatal cases in the Mid-West is the mortality reports from CDC from the year 1997 till 2009.

A.2.1 Data and Statistics for ILI Predictions

We gathered ILI incidence data at all the 10 different regions in USA from the CDC public repository. The data ranges from week 40 in the year 2008 to week 39 in the year 2009. Based on the weekly incidence, the seasonality, trend and rate of fluctuation of the ILI curve was fitted against a time series (ARMA) model with lowest AICC/BIC and neural network models with lowest SMAPE and highest R^2 statistic.

The statistics SMAPE and R^2 are defined below: Symmetric mean absolute percentage error (*SMAPE*): It is an accuracy measure based on percentage (or relative) errors. A lower value of SMAPE signifies a well fitted model.

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|Y_i - X_i|}{(Y_i + X_i)} \quad (\text{A.1})$$

Coefficient of Determination (R^2):It is the proportion of variability in a data set that is accounted for by the statistical model. It provides a measure of how well future outcomes are likely to be predicted by the model. A closer value of this statistic to unity signifies a well fit model. When STOT is defined as the total sum of errors and SSE is the sum squared errors, the coefficient of

determination is computed as the following.

$$SSE = \sum_{i=1}^n (Y_i - X_i)^2 \quad (\text{A.2})$$

$$STOT = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (\text{A.3})$$

$$R^2 = 1 - \frac{SSE}{STOT} \quad (\text{A.4})$$

A.2.2 Temporal Prediction Model Description

1. Neural Network: A simple three-layered, feed forward, supervised, multi-layered perceptron is trained using 60% of the data set and the remaining 40% data is used for validation. The first input layer contains neurons with incidence records from the previous time step feeding into them. The output obtained from the neural network is compared to the incidence record of the current time step so that the weights on the links between the neurons of different layers can be correspondingly adjusted to minimize the root mean squared error. The above procedure is repeated on the training data sets, and at the end of the training phase we assume that the network has converged, and that the optimum weights and biases minimize the mean squared error. Eventually, the validation data sets are used to quantify the prediction efficiency of the network for the 11 independent data sets of National ILI, and the regional ILI for all the 10 different regions.

2. Time-Series: Autoregressive models account for the correlations between the numbers of infected livestock in every data record. The time series model can detect the seasonality and trend of the varying ILI, thus it predicts the time when ILI would have increasing and decreasing trends. However, the ILI trends have changed drastically since week 35 of the year 2009 ever since the strain A-H1N1 was declared to be the cause of swine flu pandemic. Thus we fit a differenced Auto Regressive Moving Average (ARMA) model with a lag of 7 so as to take into consideration all the data points from the initiation of the swine flu pandemic specifically for future predictions. An appropriate order of the ARMA(p,q) model is

chosen so as to minimize AICC and BIC values using the ITSM software.

3. Bayesian Network: A single-layer discrete Bayesian network is constructed with two input nodes representing the month type and the previous record level, and one output node representing the present record level, such that inputs and outputs are classified into discrete levels. Based on the mortality data due to influenza in the Mid-West, a conditional probability table and causal relationships are derived for each data set. Bayesian parameter estimation is carried out using Maximum Likelihood Estimation (MLE) and the mean expected output is compared with the actual data set for comparison.
4. Monte-Carlo Simulations: In the Bayesian Network implementation, we mathematically computed the mean expected output. Using the Monte-Carlo simulations, we experimentally evaluate the mean expected output by simulating 1000 instances of each input combination and probabilistically deciding on a random output level and a random output value in that particular output level. The mean of the output values from all the simulated 1000 instances is the mean expected output level. Also, we generate a 95% confidence interval from the 1000 simulated outputs corresponding to each input combination. This confidence interval is represented against the mean simulated expected output and the actual data for each data set.

A.3 Influenza Like Illness Predictions

The ILI predictions are achieved by time-series and neural network predictions fitted on the actual data. The detailed results are presented below.

A.3.1 Time Series Model

The following table Represents the various ARMA(p,q) models fitted for each of the 10 regional data sets and the National data set. The AICC and BIC is also tabulated. The order p,q are chosen

so as to yield the lowest AICC/BIC. Based on the time-series model fitted against the actual data, critical months are defined as the months when the virulence is greater than the acceptable baseline level. Suitable vaccination and control strategies need to be adopted in these critical months to prevent a stronger virulence in the following Influenza cycle that begins in spring 2010.

Region	City	Base ILI	ARMA(p,q)	AICC	BIC
1	Boston	1.2	(3,4)	74.366	74.366
2	NY	2.3	(2,5)	116.296	124.277
3	DC	3.0	(1,0)	141.913	141.865
4	Atlanta	2.3	(2,4)	98.848	102.87
5	Chicago	1.7	(4,0)	70.153	72.107
6	Dallas	4.6	(1,5)	168.854	170.95
7	Kansas City	1.8	(1,0)	131.04	133.039
8	Denver	1.3	(2,1)	69.00	72.47
9	SanFrancisco	2.8	(1,0)	144.609	144.146
10	Seattle	3.3	(4,2)	131.71	137.94
Nation		3.3	(1,1)	90.6501	93.615

The National ILI data set used for predicting the ILI till the second week of November 2009 is shown in Figure A.1. The data regarding incident ILI from week 40 in the year 2008 till week 41 in year 2009 is used to fit an ARMA model. Subsequent 4 predictions regarding the strength of ILI in the following weeks shows an increasing trend till week 42 and decreasing trends thereafter till week 46 in the year 2009 when the next influenza cycle sets in.

Time-Series analysis and estimation is useful in predicting the months when the virulence of the H1N1 virus is most critical, thus seeking immediate vaccination and mitigation. Differenced ARMA models with adequate lag to cover the critical time period wherein the swine flu onset was declared has been applied to predict the trends in ILI for 4 weeks following week 41 in the year 2009. The predicted ILI values and the ILI trends for 4 weeks is tabulated below.

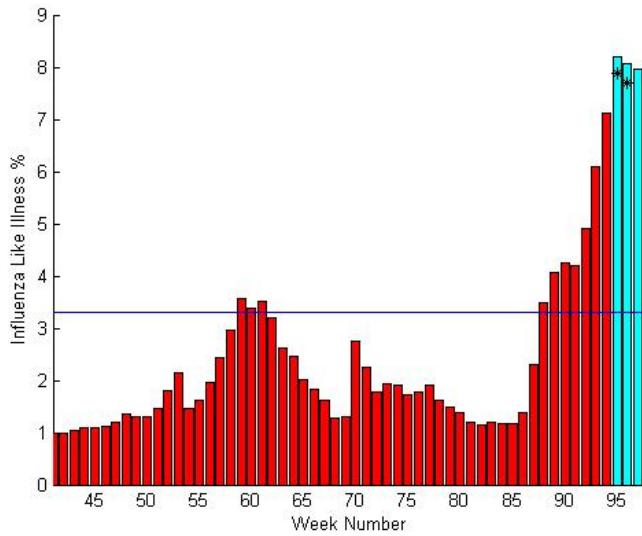


Figure A.1: Fluctuations in the National ILI from week 40 in the year 2008 to week 46 in the year 2009. The national baseline of ILI% is 3.3. Increasing trend of ILI is noticed till week 42 followed by decreasing trends till week 46 in the year 2010.

Region	Week42 predicted/actual	Week43 predicted/actual	Trend for next 4 weeks
1	4.72/4.94	5.471/5.6794	Increase(3)+Decrease(1)
2	6.1517/6.978	8.155/9.068	Increase(3)+Decrease(1)
3	9.67/10.113	11.36/10.724	Increase(2)+Decrease(2)
4	4.7878/4.878	5.097/4.9984	Decrease(1)+Increase(1)+Decrease(2)
5	8.7367/8.867	8.213/8.628	Increase(1)+Decrease(3)
6	10.39/11.1713	9.52/8.7658	Decrease(4)
7	11.98/11.7918	10.46/7.7745	Decrease(2)+Increase(2)
8	8.32/7.984	6.585/5.1685	Decrease(4)
9	6.85/6.32	6.398/7.3405	Decrease(4)
10	10.457/9.935	10.6032/9.4478	Decrease(2)+Increase(2)
Nation	8.2058/7.8735	8.073/7.7094	Increase(1)+Decrease(3)

Based on the weekly ILI trends thus predicted, the absolute ILI values can hereafter be estimated using Neural Networks. Since Neural Networks are capable of responding to quick fluctuations in data trends, thus once we are aware of the ILI trends, a well fit feed-forward neural network is successful in predicting absolute ILI values for consecutive 2 weeks.

A.3.2 Neural Network Model

Although Time-series models are very efficient in tracking the trend and seasonality of ILI patterns, they do not track sudden fluctuations in ILI levels. Thus, we perform Neural network based data fitting to check for sudden fluctuations in ILI levels that need to be detected. The following table summarizes the performance of a 7x3x1 multi-layered feed forward neural network. Seven neurons in the input layer signify the dependence of predicted values on ILI values of previous 7 weeks dating back till week 35 when swine flu was officially declared to be the cause in sudden increase in ILI values. The errors introduced and the coefficient of fitness along with the predictions for week 42 and 43 in year 2009 against the actual data has been tabulated below.

Region	SMAPE	R ²	Week42 predicted/actual	Week43 predicted/actual
1	9.1130	0.8261	5.5159/4.94	6.0274/5.6794
2	7.9295	0.9190	5.9797/6.978	8.4734/9.068
3	5.8042	0.8804	9.1842/10.113	11.722/10.724
4	6.7773	0.9673	4.4928/4.878	5.0607/4.9984
5	7.7287	0.8226	8.7615/8.867	8.4730/8.628
6	9.0832	0.9069	10.9539/11.1713	8.8192/8.7658
7	12.2891	0.9811	10.5773/11.7918	8.4496/7.7745
8	7.8263	0.9700	7.6120/7.984	5.0987/5.1685
9	5.9485	0.9056	7.678/6.324	7.0569/7.3405
10	10.3748	0.88171	9.833/9.9385	9.1054/9.4487
Nation	3.4919	0.9618	8.016/7.8735	7.9838/7.7094

National ILI fluctuations as predicted using Neural Networks is depicted in Figure A.2.

A.3.3 Comparison of Time Series and Neural Network Predictions

The following figures depict the variations in predictions for the same data set by time series and neural network analysis. We observe that time series fitted data has a gradual decreasing curve for the Seattle data set. Contrastingly, neural network predictions have steeper slopes. Both models are successful in representing the critical time phases when the ILI levels are sharply changing above the baseline levels.

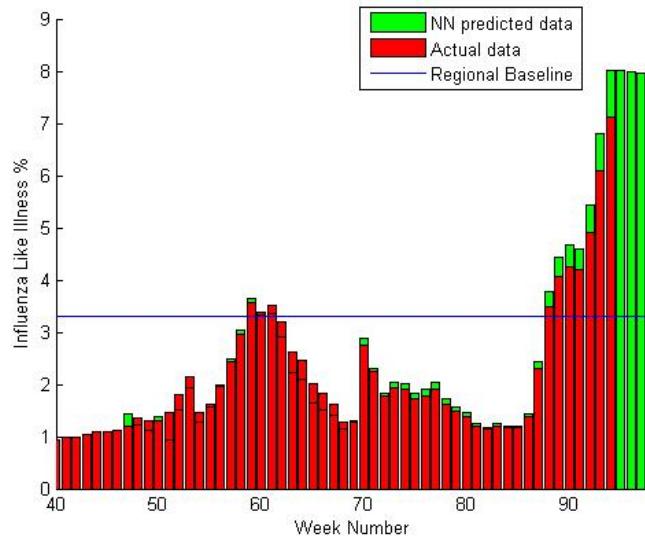


Figure A.2: National ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Neural Networks fitted against the real data. The national baseline of ILI% is 3.3. Unlike the Time-series prediction, the seasonality is not captured. Contrastingly, neural networks predict a steep decrease in the slope of ILI following week 42 in October 2009.

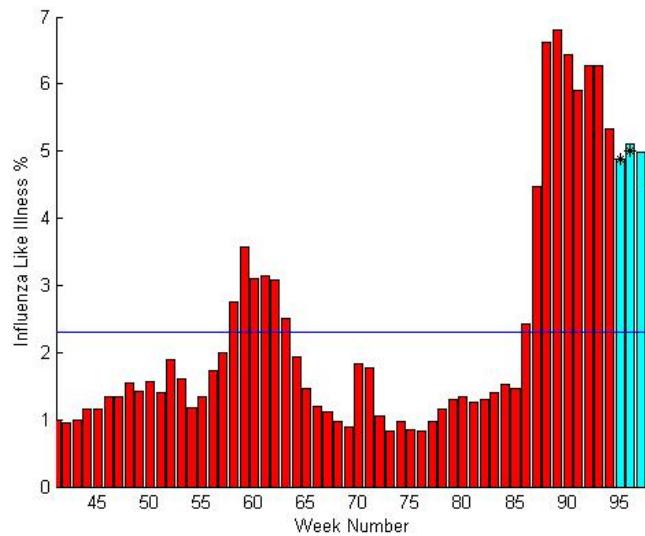


Figure A.3: Regional ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Time Series predictions for Atlanta. The regional baseline of ILI% is 2.3.

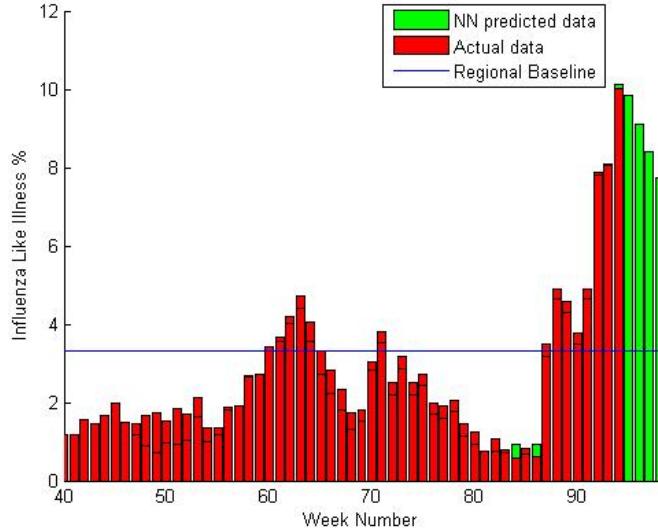


Figure A.4: National ILI from week 40 in the year 2008 to week 46 in the year 2009 predicted using Neural Networks fitted against the real data at Atlanta. The national baseline of ILI% is 2.3. This illustrates the dynamic fluctuations captured by neural networks.

A.4 Mortality Predictions

Data regarding the number of fatal cases due to Influenza from week 1, 1997 till week 41 , 2009 are used to track the expected number of fatal cases in the subsequent 4 weeks till the second week of November 2009. This is done using Bayesian Networks and Monte-Carlo simulations that predict the mean expected mortality and a 95% confidence interval for the expected fatal cases in each time step.

The following table presents predictions regarding the number of fatal cases in the Mid-Western region of the United States. For 6 different cities, the errors and goodness of fit are tabulated. The mean expected number of fatal cases along with the 95% confidence interval on the number of fatal cases is also presented week 42 and 43 for the year 2009.

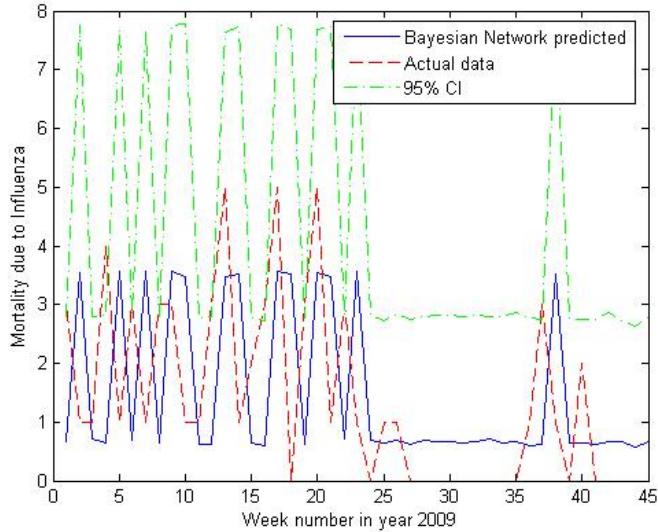


Figure A.5: Regional Mortality due to Pneumonia and Influenza from week 1 in the year 1997 to week 41 in the year 2009 taken from <http://www.cdc.gov/mmwr/>, is collected used to train up a Bayesian Network. Predictions regarding the expected mortality from week 42 in 2009 till week 46 in 2009, and a 95% confidence interval regarding the number of fatal cases has been depicted for a sample case of Kansas City.

City	Week 42/actual	Week 42 95% CI	Week 43/actual	Week 43 95%CI
Wichita	4.644/6	7.7908	12.5/6	9.9394
Kansas City	0.6327/1	2.67	0.6327/0	2.7415
Denver	8.0046/10	10.912	8.0046/9	9.9009
Tulsa	13.5576/9	17.32	11.5/10	14.81
Lincoln	5.6667/4	6.025	3.9016/3	5.667
Omaha	6.3037/9	9.7929	5.50/7	7.8817

Figure A.6 depicts the predictions thus produced.

Thus, we have temporal prediction models that incur a low-complexity to predict the acuteness of A-H1N1 influenza in USA in Fall 2009. Based on our results, the trends of influenza virus can be predicted and suitable measures can be taken to alleviate its acuteness at critical months. Future work would involve the implementation of a network based approach that will not only facilitate temporal predictions but also spatio-temporal predictions regarding the virulence of influenza in the United States.

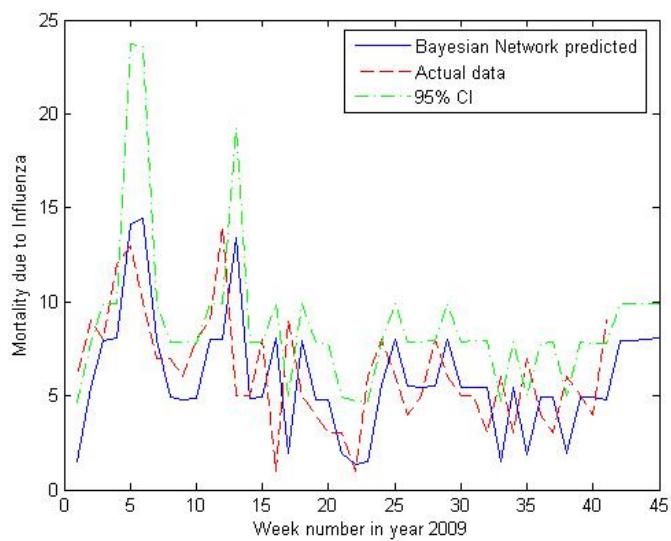


Figure A.6: Regional Mortality due to Pneumonia and Influenza from week 1 in the year 1997 to week 41 in the year 2009 taken from <http://www.cdc.gov/mmwr/>, is collected used to train up a Bayesian Network. Predictions regarding the expected mortality from week 42 in 2009 till week 46 in 2009, and a 95% confidence interval regarding the number of fatal cases has been depicted for a sample case of Denver.

Appendix B

Tables

Table B.1: Input Level Classification

InputLevel(I)	Range of infected animals at (t-1)
1	0-10
2	11-30
3	31-110
4	111-250
5	> 250

Table B.2: Output Level Classification

OutputLevel(O)	Range of infected animals at (t-1)
1	0-5
2	6-10
3	11-15
4	16-30
5	31-55
6	56-110
7	111-125
8	126-250
9	251-700
10	> 700

Table B.3: *The Conditional Probability Table with 10 input states*

Output Level	1	2	3	4	5	6	7	8	9	10
Input Level 1	0.7156	0.2796	0.0047	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 2	0.0000	0.0000	0.3704	0.6296	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 3	0.0000	0.0000	0.0000	0.0000	0.6076	0.3924	0.0000	0.0000	0.0000	0.0000
Input Level 4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2222	0.7778	0.0000	0.0000
Input Level 5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7857	0.2143
Input Level 6	0.7628	0.2372	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 7	0.0000	0.0000	0.3196	0.6804	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Input Level 8	0.0000	0.0000	0.0000	0.0000	0.4463	0.5537	0.0000	0.0000	0.0000	0.0000
Input Level 9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2245	0.7755	0.0000	0.0000
Input Level 10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6829	0.3171

Table B.4: Neural Network Results for Data set I

Node ID	AIC	MSE	BIC
1	224.1337577	2.890000000000000e-09	31.58729417
2	206.7508321	1.400000000000000e-09	30.85553618
3	257.7288683	1.170000000000000e-08	34.71410641
4	208.1182038	1.480000000000000e-09	30.92747171
5	229.3940190	3.600000000000000e-09	33.58943458
6	394.4564164	3.490000000000000e-06	40.32940108
7	224.8722713	2.980000000000000e-09	33.10237404
8	217.2251261	2.170000000000000e-09	32.76683601
9	218.6520511	2.300000000000000e-09	32.11343213
10	221.2831838	2.570000000000000e-09	31.69055957
11	217.7942192	2.220000000000000e-09	32.63265895
12	227.0195456	3.260000000000000e-09	31.70080854
13	261.1380400	1.350000000000000e-08	34.71350246
14	387.7764560	2.650000000000000e-06	40.15273409
15	279.5660109	2.910000000000000e-08	35.55355512
16	209.4266550	1.570000000000000e-09	32.16366958
17	230.8708340	3.830000000000000e-09	33.49630834
18	271.1111378	2.050000000000000e-08	34.61638466
19	227.8758420	3.380000000000000e-09	32.80395615
20	323.2545567	1.800000000000000e-07	37.45442799
21	219.3957971	2.370000000000000e-09	32.12475263
22	261.5188008	1.370000000000000e-08	34.92284893
23	270.9708151	2.040000000000000e-08	35.31250038
24	177.4448611	4.130000000000000e-10	30.15960391
25	263.3431666	1.480000000000000e-08	34.83023019
26	219.3262720	2.370000000000000e-09	33.16145908
27	392.5468051	3.230000000000000e-06	40.32016628
28	202.5680825	1.180000000000000e-09	32.00359765
29	215.7278377	2.040000000000000e-09	31.28547181
30	334.3762935	2.860000000000000e-07	37.79006599
31	222.0628659	2.650000000000000e-09	32.24883452
32	374.2397322	1.510000000000000e-06	39.59401994
33	362.4627417	9.210000000000000e-07	39.10218791
34	227.1952718	3.290000000000000e-09	33.47678226
35	241.4265751	5.950000000000000e-09	34.09141021
36	229.5297183	3.620000000000000e-09	33.48177946
37	326.8323210	2.090000000000000e-07	37.63204011
38	381.8064600	2.060000000000000e-06	39.94067872
39	315.5176312	1.300000000000000e-07	37.17830469
40	259.8759379	1.280000000000000e-08	34.7082879

Table B.5: Neural Network Results for Data set I continued

Node ID	AIC	MSE	BIC
41	216.6261627	2.12000000000000e-09	32.73935195
42	248.0644971	7.84000000000000e-09	34.23486778
43	242.6974709	6.27000000000000e-09	34.03997617
44	399.0419674	4.23000000000000e-06	40.36881870
45	443.1732615	2.66000000000000e-05	42.02182224
46	308.9748876	9.92000000000000e-08	36.53733949
47	263.6001105	1.50000000000000e-08	35.01455510
48	314.1111068	1.23000000000000e-07	37.02735931
49	306.7959711	9.06000000000000e-08	36.80488348
50	228.0494283	3.41000000000000e-09	33.37184031
51	225.3751855	3.05000000000000e-09	31.68155830
52	245.5672990	7.07000000000000e-09	34.24208628
53	286.1367129	3.83000000000000e-08	35.61126285
54	302.8752328	7.69000000000000e-08	36.62799111
55	396.5768445	3.82000000000000e-06	39.91148552
56	353.1351314	6.25000000000000e-07	38.59055351
57	202.5553462	1.18000000000000e-09	32.19149734
58	223.6190503	2.83000000000000e-09	31.55844009
59	413.1050166	7.60000000000000e-06	40.93466218
60	221.7632398	2.62000000000000e-09	32.23737911
61	345.8857140	4.62000000000000e-07	38.41992927
62	231.9052773	4.00000000000000e-09	33.65743464
63	246.0612679	7.21000000000000e-09	34.28147819
64	233.5262480	4.28000000000000e-09	33.73999848
65	291.2383775	4.74000000000000e-08	36.02380071
66	211.6540353	.72000000000000e-09	31.05933338
67	217.9924804	2.24000000000000e-09	32.97838077
68	250.8282729	8.80000000000000e-09	34.47471707
69	197.3049034	9.46000000000000e-10	30.46193868
70	409.0071742	6.41000000000000e-06	40.92371586
71	337.4947849	3.26000000000000e-07	38.07600794
72	238.8501245	5.34000000000000e-09	33.95905006
73	239.0828047	5.39000000000000e-09	33.88127932
74	393.5477697	.36000000000000e-06	39.939539040
75	388.8241796	2.76000000000000e-06	39.99168510
76	227.9006083	3.38000000000000e-09	32.53412786
77	245.2879897	6.98000000000000e-09	34.25242663
78	190.7760951	7.21000000000000e-10	31.06364562
79	301.6659742	7.32000000000000e-08	36.48366243

Table B.6: Neural Network Results for Data set S

Node ID	AIC	MSE	BIC
1	222.421180800000	2.690000000000000e-09	31.5079718200000
2	225.299151400000	3.040000000000000e-09	31.6319569600000
3	267.569074300000	1.770000000000000e-08	35.1258423900000
4	222.493183000000	2.700000000000000e-09	32.7304812500000
5	232.158287900000	4.040000000000000e-09	33.6735291300000
6	389.258975700000	2.810000000000000e-06	40.1925920800000
7	261.073206200000	1.350000000000000e-08	34.7522601300000
8	228.828502700000	3.520000000000000e-09	33.3035534500000
9	234.260409000000	4.410000000000000e-09	33.7290495800000
10	217.929004400000	2.230000000000000e-09	31.7739595800000
11	208.546357600000	1.510000000000000e-09	32.6849451300000
12	205.226232000000	1.320000000000000e-09	30.9237602100000
13	251.872861300000	9.190000000000000e-09	34.3742528100000
14	382.534180600000	2.130000000000000e-06	39.9219244500000
15	284.267026500000	3.540000000000000e-08	35.6306658400000
16	227.138278100000	3.280000000000000e-09	33.4830471100000
17	250.579898500000	8.710000000000000e-09	34.4630490200000
18	241.492194500000	5.960000000000000e-09	34.0619904400000
19	205.711050000000	1.340000000000000e-09	32.5589845300000
20	292.582446900000	5.010000000000000e-08	36.1790629900000
21	202.722115000000	1.190000000000000e-09	31.1460435300000
22	281.346840000000	3.140000000000000e-08	35.5018447500000
23	342.262841900000	3.970000000000000e-07	37.9612383500000
24	210.937209900000	1.670000000000000e-09	32.1281285700000
25	293.589389100000	5.230000000000000e-08	36.2600018600000
26	218.140856500000	2.250000000000000e-09	32.5569435800000
27	389.052255500000	2.790000000000000e-06	40.1455163300000
28	216.276939800000	2.090000000000000e-09	32.0377995900000
29	227.657550600000	3.350000000000000e-09	31.7578845500000
30	350.981517300000	5.710000000000000e-07	38.4347686900000
31	236.816669200000	4.910000000000000e-09	33.7185127700000
32	367.235852900000	1.120000000000000e-06	39.2937817800000
33	356.579559300000	7.210000000000000e-07	38.8623938400000
34	221.563875300000	2.600000000000000e-09	33.0584453000000
35	257.316684600000	1.150000000000000e-08	34.6342013200000
36	242.146487000000	6.130000000000000e-09	33.9792890200000
37	330.050545700000	2.390000000000000e-07	37.7618001600000
38	395.577277200000	3.660000000000000e-06	40.5087392500000
39	312.932138900000	1.170000000000000e-07	37.0708499500000
40	256.332616900000	1.110000000000000e-08	34.5464769200000

Table B.7: Neural Network Results for Data set S continued

Node ID	AIC	MSE	BIC
41	219.511720800000	2.390000000000000e-09	32.0129206400000
42	256.521908400000	1.120000000000000e-08	34.6807209000000
43	206.239056000000	1.370000000000000e-09	32.6249155000000
44	274.510756500000	2.360000000000000e-08	35.4598988300000
45	442.575666400000	2.590000000000000e-05	41.9172104400000
46	272.010325300000	2.130000000000000e-08	35.3645232300000
47	254.092118300000	1.010000000000000e-08	34.4959112800000
48	299.785744900000	6.760000000000000e-08	36.0827405300000
49	296.159037300000	5.820000000000000e-08	36.3068620100000
50	260.228644100000	1.300000000000000e-08	34.7574256700000
51	214.655431300000	1.950000000000000e-09	31.6003407200000
52	248.312713000000	7.920000000000000e-09	34.2976271400000
53	288.543996300000	4.230000000000000e-08	35.7365922600000
54	296.791765400000	5.970000000000000e-08	36.3982470100000
55	426.200889400000	1.310000000000000e-05	41.4820284000000
56	348.695443300000	5.190000000000000e-07	38.5121861500000
57	218.990222600000	2.330000000000000e-09	32.8088684200000
58	209.686143200000	1.580000000000000e-09	30.9773005400000
59	428.872160200000	1.470000000000000e-05	41.4976476100000
60	224.620168200000	2.950000000000000e-09	32.7073127100000
61	377.056112200000	1.690000000000000e-06	39.7315401700000
62	231.855318900000	3.990000000000000e-09	33.6753300500000
63	250.891606200000	8.820000000000000e-09	34.4426412200000
64	242.493736200000	6.220000000000000e-09	34.1279932600000
65	315.646687600000	1.310000000000000e-07	37.1809922900000
66	219.283512500000	2.360000000000000e-09	31.3781361300000
67	195.784383800000	8.880000000000000e-10	31.0818523600000
68	230.847446300000	3.830000000000000e-09	33.6507200200000
69	219.497510200000	2.380000000000000e-09	31.4021813300000
70	428.414652500000	1.440000000000000e-05	41.8670436800000
71	323.562371700000	1.820000000000000e-07	37.4775765500000
72	240.930575800000	5.820000000000000e-09	34.0018444500000
73	241.010967700000	5.840000000000000e-09	33.9364262500000
74	300.945469900000	7.100000000000000e-08	36.5235056000000
75	405.203786600000	5.470000000000000e-06	40.5031008600000
76	223.138811200000	2.780000000000000e-09	32.8260628900000
77	248.544615400000	8.000000000000000e-09	34.3252580300000
78	223.188682100000	2.780000000000000e-09	32.9180699000000
79	285.033232000000	3.660000000000000e-08	35.7453073900000

Table B.8: AR Model Results for Data set I

Node ID	p	sample size	AIC	MSE	BIC
1	1	3	-134.550529700000	1.34000000000000e-07	-13.7686240800000
2	1	5	-128.959958100000	1.49000000000000e-08	-15.7177337900000
3	1	5	-113.817833300000	0.000188847000000000	-6.12106761700000
4	2	4	-138.497522600000	1.40000000000000e-07	-10.9856199000000
5	1	3	-125.298709500000	1.60000000000000e-05	-8.42470244600000
6	1	4	-98.4800436900000	0.0030549130000000	-3.16148944200000
7	1	3	-155.911164600000	9.38000000000000e-06	-8.82605455800000
8	1	7	-140.543827900000	5.77000000000000e-06	-9.88639693200000
9	1	4	-130.571946100000	3.01000000000000e-06	-10.4141743200000
10	1	5	-129.045002600000	7.28000000000000e-07	-11.6069071200000
11	1	5	-117.443717300000	0.000199478000000000	-5.97423369500000
12	1	3	-146.525605500000	3.96000000000000e-08	-15.0177639900000
13	1	3	-140.131055100000	0.000311521000000000	-5.69222186900000
14	1	8	-103.895332800000	0.0026976950000000	-3.81314290000000
15	1	6	-113.126996200000	0.0042235960000000	-3.04215894300000
16	2	3	-187.582441500000	1.87000000000000e-05	-5.78123128000000
17	1	8	-121.956773800000	1.66000000000000e-05	-8.84490530600000
18	1	4	-110.193564500000	0.000972553000000000	-4.57570290800000
19	1	4	-151.010094700000	2.30000000000000e-05	-8.41158036800000
20	1	6	-101.952334000000	0.0014328860000000	-4.25272683200000
21	1	3	-145.744690900000	2.57000000000000e-06	-10.1026688300000
22	1	5	-133.387681000000	8.06000000000000e-05	-6.86744180100000
23	1	3	-115.797671100000	0.000299624000000000	-5.59408323100000
24	2	3	-119.999631100000	1.61000000000000e-05	-5.71026022600000
25	1	5	-139.329520700000	0.000366090000000000	-5.68560046100000
26	1	11	-126.827812400000	7.97000000000000e-06	-9.97607868000000
27	1	9	-83.2155878700000	0.0077947880000000	-2.83594808800000
28	1	5	-133.329741600000	1.11000000000000e-06	-11.3302337100000
29	1	3	-159.712054200000	2.03000000000000e-07	-12.5873953300000
30	1	11	-60.6181928000000	0.0298168340000000	-1.74880210500000
31	1	4	-152.457823000000	2.41000000000000e-06	-10.2634172000000
32	1	10	-86.9556004200000	0.00518964400000000	-3.38062628500000
33	1	4	-127.941716200000	0.000854371000000000	-4.43436898500000
34	1	6	-113.582647000000	0.000109818000000000	-6.96953683200000
35	1	3	-170.937155600000	3.00000000000000e-05	-7.62916455200000
36	1	5	-133.590032200000	1.70000000000000e-05	-8.63921654400000
37	1	3	-119.992413700000	0.000825446000000000	-4.28076418100000
38	1	11	-99.6090798400000	0.00546837600000000	-3.45103863900000
39	1	11	-82.5993198200000	0.0102406100000000	-2.82540738300000
40	1	5	-125.332303900000	0.000681772000000000	-4.73531405900000

Table B.9: AR Model Results for Data set I continued

Node ID	p	sample size	AIC	MSE	BIC
41	1	3	-138.900877100000	0.00105075000000000	-4.04131245700000
42	1	5	-125.921784400000	3.98000000000000e-05	-7.68642798000000
43	1	3	-173.413669800000	3.97000000000000e-05	-7.32013707000000
44	1	10	-74.7152683200000	0.0154658340000000	-2.28733965200000
45	1	3	-52.4201664600000	0.0186144180000000	-1.26070202500000
46	1	6	-112.940039400000	0.00167765600000000	-4.19184656300000
47	1	4	-149.495483100000	4.32000000000000e-05	-7.50290890000000
48	1	11	-84.0563710300000	0.0139809120000000	-2.51436535000000
49	1	7	-112.337860700000	0.00125888900000000	-4.57921016400000
50	1	3	-133.571676600000	0.000298842000000000	-5.34407456200000
51	1	8	-174.537358900000	2.73000000000000e-07	-12.9511142700000
52	1	3	-144.821010900000	0.000107139000000000	-6.78518603100000
53	1	3	-110.416504700000	0.000297662000000000	-5.30008328000000
54	1	3	-166.222502200000	8.13000000000000e-05	-6.63071308300000
55	1	6	-75.7259814200000	0.0095637900000000	-2.26471092600000
56	1	11	-102.312093900000	0.00217361200000000	-4.37097872200000
57	1	6	-125.268673000000	1.61000000000000e-06	-10.9444456300000
58	2	3	-145.430912800000	1.57000000000000e-08	-12.5738254600000
59	1	11	-65.8192767900000	0.0172775420000000	-2.39023651600000
60	1	3	-115.174443800000	0.000245076000000000	-5.52672038500000
61	1	6	-91.4205522800000	0.00295930100000000	-3.61220870600000
62	1	3	-123.851986500000	0.000206414000000000	-5.68778656600000
63	1	4	-116.479273200000	0.000304661000000000	-5.53981603200000
64	1	3	-125.997368300000	0.000204107000000000	-5.69985058400000
65	1	4	-128.272221900000	0.000313386000000000	-5.69881455300000
66	1	5	-134.044236000000	9.84000000000000e-09	-16.4993644700000
67	1	5	-107.652488100000	8.21000000000000e-06	-9.15502728600000
68	1	4	-119.407718300000	0.000428876000000000	-5.08173907400000
69	1	7	-118.923216200000	1.37000000000000e-08	-16.5094730400000
70	1	10	-66.2689391300000	0.0205315550000000	-2.02022318700000
71	1	8	-92.8503898400000	0.00266577500000000	-3.78950753400000
72	1	3	-146.886798600000	4.17000000000000e-05	-7.27059868200000
73	1	3	-116.561419000000	4.49000000000000e-05	-7.28734596100000
74	1	4	-88.6553702600000	0.00391419000000000	-3.03231423100000
75	1	11	-79.1210601000000	0.0131508870000000	-2.61948960000000
76	1	3	-122.903085300000	6.27000000000000e-05	-6.85971234100000
77	1	7	-110.219069700000	0.000827712000000000	-4.81871675800000
78	1	6	-120.116984200000	0.000328691000000000	-5.76883452900000
79	1	10	-103.338448000000	0.001186711000000000	-4.88638042300000

Table B.10: AR Model Results for Data set S

Node ID	p	sample size	AIC	MSE	BIC
1	1	8	-100.825427700000	1.95000000000000e-08	-16.3231448500000
2	1	3	-106.039121000000	3.29000000000000e-08	-14.4101577100000
3	1	5	-100.262636500000	0.000279333000000000	-5.81594202900000
4	1	9	-107.233952200000	4.81000000000000e-06	-10.2891582100000
5	1	5	-105.898028800000	0.000133955000000000	-6.61343902300000
6	1	3	-95.4030449800000	0.0021812740000000	-3.44463591300000
7	1	8	-112.731005400000	0.000153317000000000	-6.65297615900000
8	2	3	-151.922626400000	5.68000000000000e-05	-4.64444613200000
9	1	10	-108.175825500000	1.53000000000000e-05	-9.26902213500000
10	1	4	-117.744022700000	1.21000000000000e-06	-11.1584645600000
11	1	7	-105.281376300000	2.30000000000000e-05	-8.41391332300000
12	1	3	-132.095532400000	1.19000000000000e-06	-11.3504880100000
13	1	3	-145.598418100000	5.36000000000000e-05	-7.15051723900000
14	1	7	-83.9966998300000	0.0020221570000000	-3.95898844500000
15	1	7	-99.3132698400000	0.0050117470000000	-3.00427970400000
16	1	3	-142.281527200000	6.69000000000000e-05	-7.27037417800000
17	1	5	-97.9932371700000	0.000845043000000000	-4.83393528700000
18	1	5	-104.549892900000	0.000122500000000000	-6.80209424600000
19	1	4	-96.4828930900000	0.000532410000000000	-5.33766279100000
20	1	3	-104.915032700000	0.000902624000000000	-4.59785640800000
21	1	4	-144.606042600000	1.09000000000000e-06	-11.0671496500000
22	1	3	-86.7857231600000	0.000110044000000000	-6.36936777900000
23	1	11	-84.6945825400000	0.00189113800000000	-4.56706323400000
24	1	4	-120.792488900000	2.14000000000000e-05	-8.25764578300000
25	1	5	-98.1978400200000	0.0013822150000000	-4.44422993800000
26	1	3	-127.745831900000	5.02000000000000e-06	-9.48959657400000
27	1	8	-81.6324597800000	0.0044294740000000	-3.26567426700000
28	1	7	-94.4351232300000	5.81000000000000e-07	-12.0669712200000
29	1	3	-109.230243900000	1.78000000000000e-07	-12.7772590600000
30	1	11	-62.1588540300000	0.0291111730000000	-1.77307013000000
31	2	3	-140.454738300000	2.44000000000000e-05	-5.02638970700000
32	1	3	-85.8730150500000	0.00455864400000000	-2.60957114100000
33	1	3	-113.153131500000	0.000610741000000000	-4.60978729800000
34	1	5	-93.8331131300000	2.42000000000000e-05	-8.50616461600000
35	1	3	-92.3919063700000	4.12000000000000e-05	-7.45116920800000
36	1	3	-123.278393300000	0.000285754000000000	-5.91026942100000
37	1	3	-96.7425770800000	0.00148526900000000	-4.21936952800000
38	1	10	-95.6462646400000	0.00430881200000000	-3.55078085300000
39	1	3	-81.8374520000000	0.00557288000000000	-2.37550287800000
40	1	5	-116.110933400000	0.000373962000000000	-5.33301168700000

Table B.11: AR Model Results for Data set S continued

Node ID	p	sample size	AIC	MSE	BIC
41	1	4	-133.221363800000	6.90000000000000e-05	-6.90404707200000
42	1	4	-120.045563900000	0.000360291000000000	-5.33533416400000
43	1	3	-129.998421100000	1.97000000000000e-06	-10.3389231100000
44	1	10	-79.0706943500000	0.0114857620000000	-2.58669587800000
45	1	3	-51.9536851700000	0.0178807720000000	-1.33573858600000
46	2	3	-102.147155500000	0.000116161000000000	-3.61965594400000
47	1	5	-116.096038600000	1.65000000000000e-05	-8.45770268000000
48	1	4	-86.6307685100000	0.0104233690000000	-1.87453155500000
49	1	7	-96.0288296900000	0.0014383840000000	-4.28194474600000
50	1	3	-114.439013500000	0.000284239000000000	-5.41261229600000
51	1	4	-117.960247100000	1.25000000000000e-06	-10.9214099700000
52	1	4	-117.926533200000	0.000108315000000000	-6.77980123600000
53	1	3	-145.895105100000	0.000335702000000000	-5.20519348100000
54	2	3	-142.416486200000	6.07000000000000e-05	-4.13410521900000
55	1	3	-67.4311099400000	0.0106429240000000	-1.76933672800000
56	1	5	-102.577855700000	0.000495480000000000	-5.06397988100000
57	1	4	-93.6342017600000	4.62000000000000e-06	-9.84783685200000
58	1	3	-137.045138800000	9.60000000000000e-09	-16.1138345000000
59	1	11	-52.6263570500000	0.0261052140000000	-1.97624316700000
60	1	3	-106.109606900000	0.000313526000000000	-5.26845054500000
61	1	4	-89.2187157500000	0.00292173500000000	-3.20063133700000
62	1	3	-125.185385500000	0.000197105000000000	-5.70891880200000
63	1	3	-123.647152800000	0.000238831000000000	-5.60386108200000
64	1	3	-113.568637900000	7.80000000000000e-05	-6.82659664700000
65	1	9	-98.7102510800000	0.00129983700000000	-4.68974949600000
66	1	3	-100.099782900000	1.00000000000000e-08	-15.9740701300000
67	2	4	-117.529900000000	1.46000000000000e-06	-8.11747690400000
68	1	5	-111.587689300000	5.09000000000000e-05	-7.35401495600000
69	1	3	-117.639769900000	7.48000000000000e-08	-13.5851617900000
70	1	4	-63.8398594300000	0.0128878870000000	-1.90530971600000
71	1	4	-98.3268746200000	0.0021264850000000	-3.47642486200000
72	2	3	-125.565441500000	5.24000000000000e-05	-4.25861960000000
73	1	5	-95.2819979000000	4.00000000000000e-05	-7.65316103500000
74	1	3	-141.212397400000	0.000469040000000000	-5.33633575800000
75	1	11	-62.1663160500000	0.0268512390000000	-1.93215911200000
76	1	8	-105.212936800000	3.28000000000000e-05	-8.19678293500000
77	1	6	-93.7619299000000	0.000524535000000000	-5.16455873200000
78	1	5	-126.574839400000	5.40000000000000e-06	-9.61963940800000
79	1	5	-97.3090942900000	0.000893458000000000	-4.50852673400000

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	6.310552	0.947369	0	0	0	0	0	0	0	0	0
*10^-4	3	3.771548	1.780645	1.040586	0	0	0	0	0	0	0	0
	4	3.843783	0.950913	1.374315	0.762992	0	0	0	0	0	0	0
	5	4.484188	1.220665	1.173147	0.800148	0.822272	0	0	0	0	0	0
	6	2.433716	3.388474	1.808123	2.521226	2.273526	1.195965	0	0	0	0	0
	7	5.401291	0.921287	1.11247	1.74463	1.115324	2.284228	1.405691	0	0	0	0
	8	7.431936	1.553089	1.554445	1.295002	0.81395	0.881021	0.897449	0.70846	0	0	0
	9	4.035966	0.667553	2.0221	1.129283	1.020675	0.83553	0.73364	0.726408	0.511091	0	0
	10	3.822984	1.926419	0.906339	1.035606	1.396592	0.807239	0.838998	3996.012	0.891423	2998.165	0
	11	1.698598	1.816232	0.791196	0.678329	0.69399	0.634915	0.721658	0.580931	0.92904	4580.282	3998.067

(a) Standard deviation of RMSE for I_t .

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	0.062173	0.062413	0	0	0	0	0	0	0	0	0
*10^-4	3	0.066589	0.066884	0.067132	0	0	0	0	0	0	0	0
	4	0.062278	0.062401	0.062444	0.062359	0	0	0	0	0	0	0
	5	0.062423	0.062355	0.062373	0.06229	0.062183	0	0	0	0	0	0
	6	0.065241	0.065498	0.065429	0.065314	0.06503	0.064803	0	0	0	0	0
	7	0.065512	0.065376	0.06527	0.065101	0.064964	0.064771	0.064478	0	0	0	0
	8	0.051403	0.049826	0.049699	0.049693	0.049584	0.049464	0.049441	0.049339	0	0	0
	9	0.051096	0.049931	0.049787	0.049651	0.049592	0.049467	0.049413	0.049352	0.049233	0	0
	10	0.053941	0.052719	0.052587	0.052439	0.05236	0.052242	0.052066	0.241486	0.051873	0.146603	0
	11	0.053802	0.052762	0.052484	0.052428	0.052333	0.05219	0.05206	0.051911	0.051819	0.336175	0.241253

(b) Average RMSE for I_t

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	5.948035	0.729169	0	0	0	0	0	0	0	0	0
*10^-4	3	4.48631	2.498337	0.873248	0	0	0	0	0	0	0	0
	4	3.369987	0.836054	1.291679	0.902712	0	0	0	0	0	0	0
	5	3.889623	1.671172	1.433332	1.216624	0.950433	0	0	0	0	0	0
	6	2.760118	4.260988	2.271009	3.295078	2.941242	1.748568	0	0	0	0	0
	7	5.841709	1.5541	1.370496	2.086569	1.622629	3.174777	2.137844	0	0	0	0
	8	7.748699	1.847479	2.164124	1.704097	1.123234	1.133637	1.223308	0.956703	0	0	0
	9	4.676097	0.963664	2.526879	1.635854	1.340276	1.096209	1.063275	0.932634	0.681954	0	0
	10	3.94085	2.483713	1.276139	1.387756	2.005633	1.067103	1.360633	3996.594	1.266113	2998.908	0
	11	1.55022	2.06739	1.116844	0.837765	0.977891	0.698742	0.923908	0.678229	1.174378	4581.556	3999.21

(c) Standard deviation of RMSE for I_v .

Table B.12: RMSE analysis for best Neural network structure.

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	0.062029	0.06238	0	0	0	0	0	0	0	0	0
	3	0.066233	0.066538	0.066934	0	0	0	0	0	0	0	0
	4	0.062183	0.062363	0.062401	0.062418	0	0	0	0	0	0	0
	5	0.06229	0.062262	0.062311	0.062249	0.062218	0	0	0	0	0	0
	6	0.065266	0.065353	0.06518	0.064931	0.064542	0.064205	0	0	0	0	0
	7	0.065449	0.065107	0.064959	0.06465	0.06443	0.064099	0.063644	0	0	0	0
	8	0.051505	0.049877	0.049771	0.049795	0.049681	0.049563	0.049573	0.049458	0	0	0
	9	0.051193	0.050035	0.049889	0.049748	0.049711	0.049575	0.049546	0.049487	0.04938	0	0
	10	0.053986	0.052741	0.052608	0.052409	0.052314	0.052162	0.051962	0.241365	0.051702	0.146391	0
	11	0.053771	0.052722	0.052429	0.052343	0.052232	0.052052	0.051907	0.051718	0.051604	0.33599	0.241033

(a) Average RMSE for I_v .

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	1.713061	0.400871	0	0	0	0	0	0	0	0	0
*10^-3	3	2.806966	0.599809	0.292405	0	0	0	0	0	0	0	0
	4	1.600388	0.489289	0.195729	0.05471	0	0	0	0	0	0	0
	5	1.324842	0.5833	0.185242	0.05588	0.034803	0	0	0	0	0	0
	6	1.98665	0.493961	0.241961	0.070448	0.046309	0.060303	0	0	0	0	0
	7	0.854338	0.287385	0.095285	0.069083	0.058943	0.085812	0.066988	0	0	0	0
	8	0.962515	0.654814	0.398105	0.375922	0.211389	0.198211	0.13068	0.226303	0	0	0
	9	0.550287	0.630244	0.493983	0.238393	0.166391	0.190363	0.096626	0.104979	0.084559	0	0
	10	0.812954	0.399734	0.366642	0.346248	0.237842	0.058644	0.043083	0.252233	0.052933	0.626025	0
	11	0.781274	0.47561	0.313265	0.303514	0.172357	0.051755	0.16418	0.095505	0.049768	0.430119	413.2222

(b) Standard deviation of RMSE for S_t .

Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	0.042772	0.037386	0	0	0	0	0	0	0	0	0
	3	0.043738	0.039488	0.039082	0	0	0	0	0	0	0	0
	4	0.039551	0.03605	0.034869	0.034688	0	0	0	0	0	0	0
	5	0.038238	0.035644	0.034912	0.034757	0.034701	0	0	0	0	0	0
	6	0.037893	0.034361	0.033469	0.033189	0.033123	0.033051	0	0	0	0	0
	7	0.037156	0.034042	0.033435	0.033227	0.03317	0.033103	0.033083	0	0	0	0
	8	0.028614	0.023784	0.020721	0.019397	0.018657	0.01831	0.018047	0.017866	0	0	0
	9	0.028194	0.023567	0.020752	0.019267	0.018538	0.018359	0.017903	0.017838	0.017677	0	0
	10	0.03111	0.025245	0.022558	0.021452	0.020699	0.020207	0.019983	0.01998	0.019861	0.020081	0
	11	0.030857	0.025084	0.022463	0.021383	0.020604	0.020155	0.020064	0.019902	0.019834	0.020054	0.215952

(c) Average RMSE for S_t .

Table B.13: RMSE analysis for best Neural network structure continued.

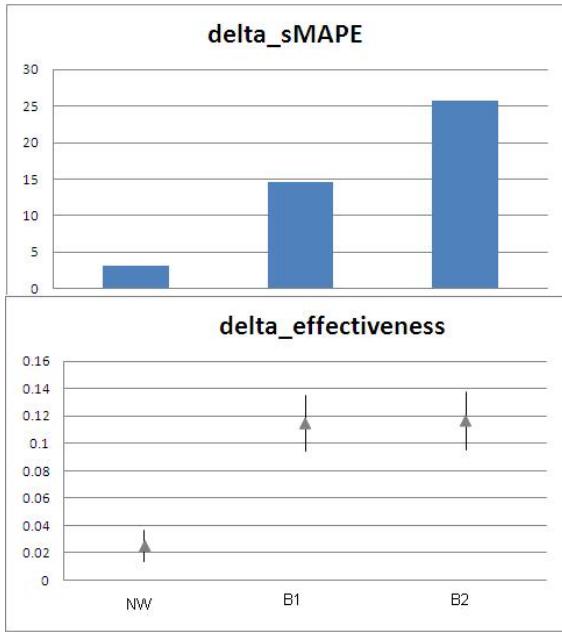
Layer	Hidden ne	1	2	3	4	5	6	7	8	9	10	11
Input neu	2	1.498971	0.596276	0	0	0	0	0	0	0	0	0
*10^-3	3	2.193729	0.608541	0.267397	0	0	0	0	0	0	0	0
	4	1.862968	0.578421	0.179852	0.071873	0	0	0	0	0	0	0
	5	1.693829	0.268808	0.167048	0.096516	0.096606	0	0	0	0	0	0
	6	1.028045	0.334211	0.21463	0.137801	0.073177	0.085833	0	0	0	0	0
	7	1.720591	0.386386	0.089197	0.104419	0.092843	0.130227	0.137562	0	0	0	0
	8	0.824314	0.730889	0.629799	0.508851	0.212442	0.161741	0.212383	0.134998	0	0	0
	9	0.88074	0.440748	0.326809	0.36627	0.186122	0.213512	0.197475	0.207808	0.19503	0	0
	10	0.543959	0.451378	0.212871	0.387977	0.143649	0.115	0.17336	0.111604	0.098648	309.1754	0
	11	0.460607	0.364565	0.270063	0.130202	0.183241	0.137091	0.053968	0.088344	309.7613	1.215907	413.1375

(a) Standard deviation of RMSE for S_v .

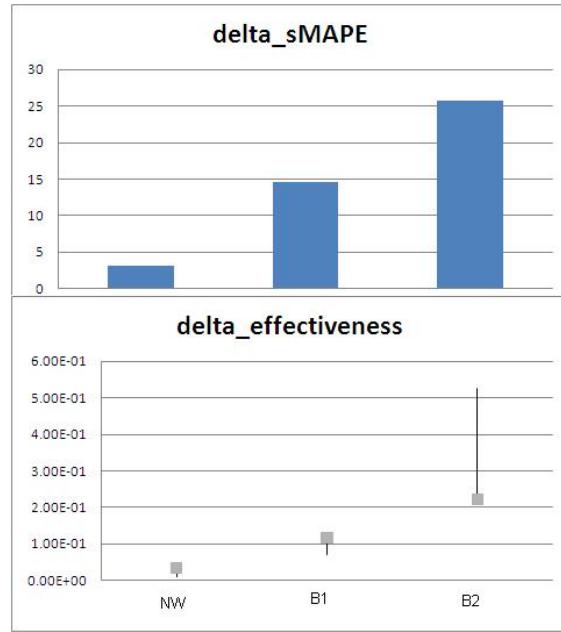
Layer	Hidden	1	2	3	4	5	6	7	8	9	10	11
Input	2	0.041605	0.037484	0	0	0	0	0	0	0	0	0
	3	0.042563	0.039249	0.038832	0	0	0	0	0	0	0	0
	4	0.039436	0.036132	0.034661	0.034444	0	0	0	0	0	0	0
	5	0.039248	0.035436	0.034734	0.034591	0.034454	0	0	0	0	0	0
	6	0.037464	0.034351	0.033331	0.032992	0.03284	0.032798	0	0	0	0	0
	7	0.03693	0.034243	0.033273	0.033054	0.033005	0.032898	0.032855	0	0	0	0
	8	0.028973	0.02388	0.021138	0.019671	0.018648	0.018313	0.018147	0.017891	0	0	0
	9	0.028317	0.023185	0.020704	0.019378	0.01869	0.018338	0.018078	0.017886	0.017881	0	0
	10	0.030508	0.025244	0.022643	0.021403	0.020516	0.020301	0.020138	0.019976	0.019958	0.119827	0
	11	0.030417	0.025062	0.022412	0.021048	0.020689	0.020211	0.01999	0.019985	0.118111	0.020401	0.21606

(b) Average RMSE for S_v .

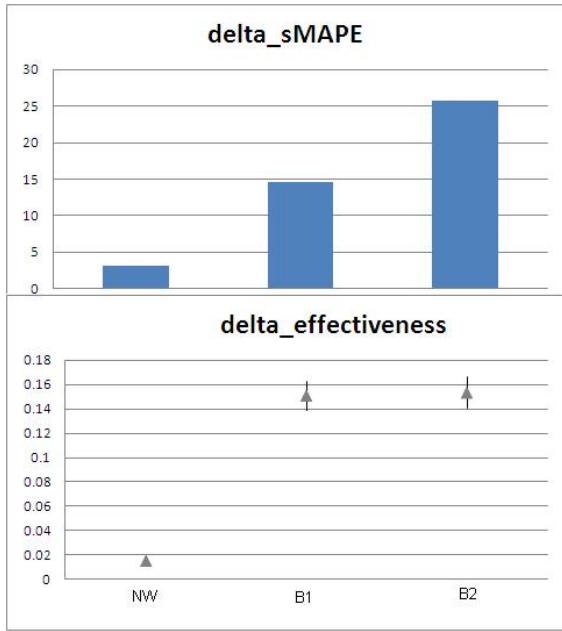
Table B.14: RMSE analysis for best Neural network structure continued.



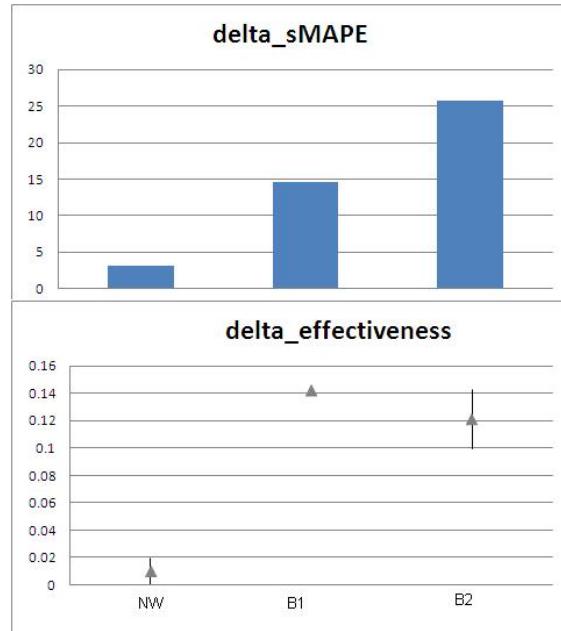
(a) Mitigation Strategy 1



(b) Mitigation Strategy 2



(c) Mitigation Strategy 3



(d) Mitigation Strategy 4

Figure B.1: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Iran. The prediction percentage error between the baseline models and model W($\Delta sMPAE$) is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.

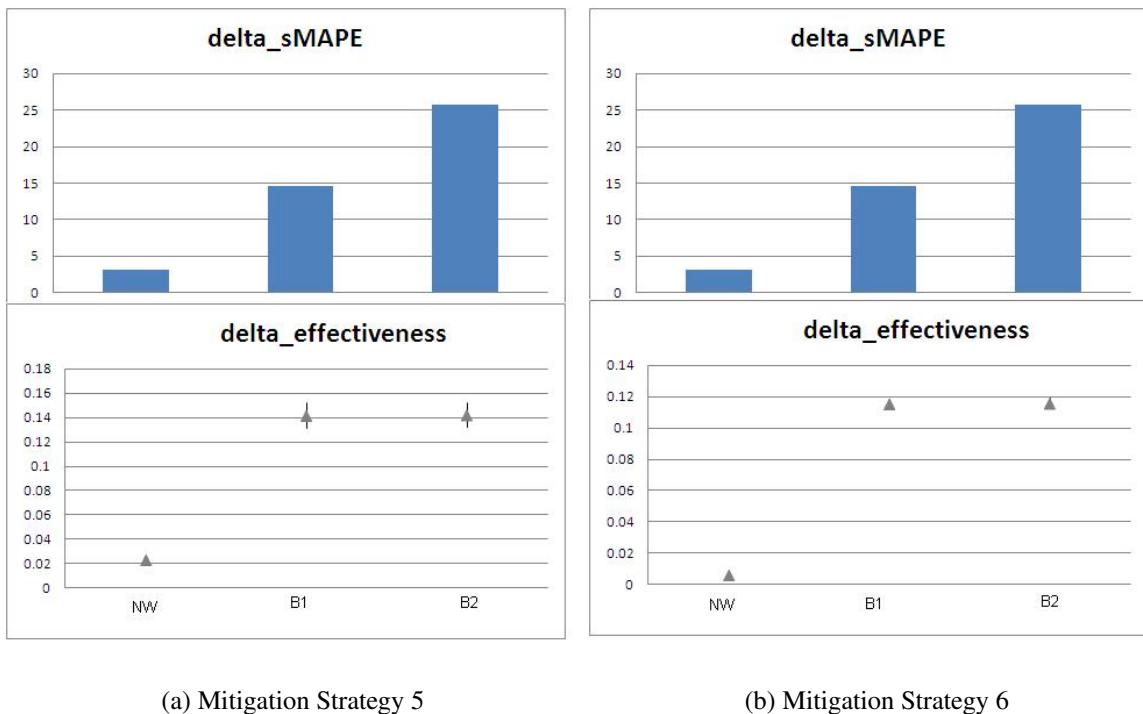
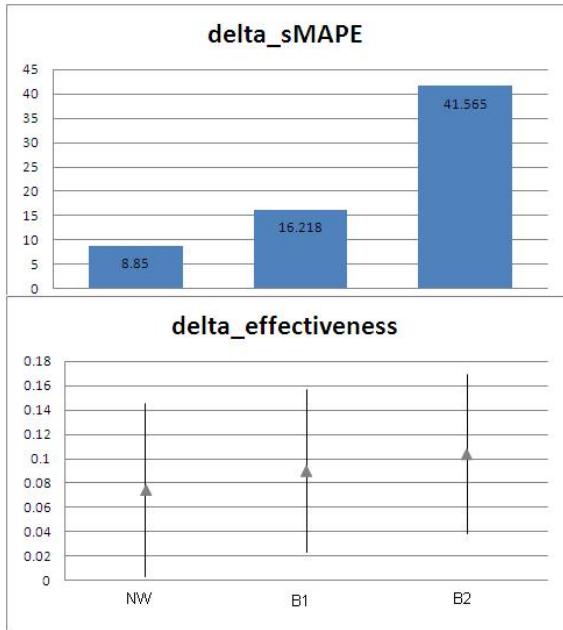
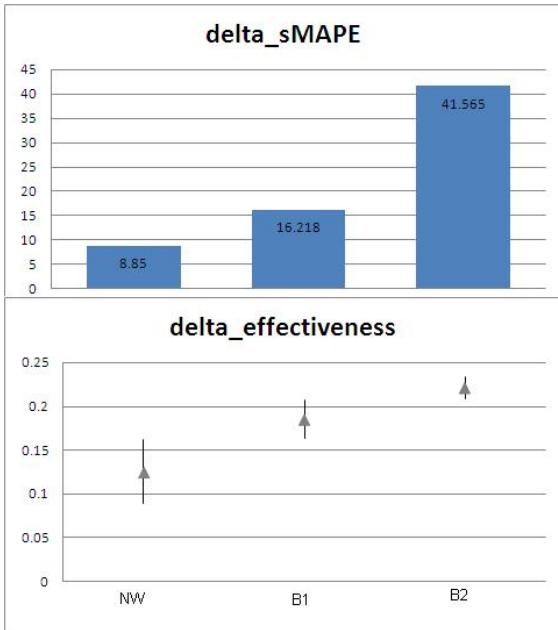


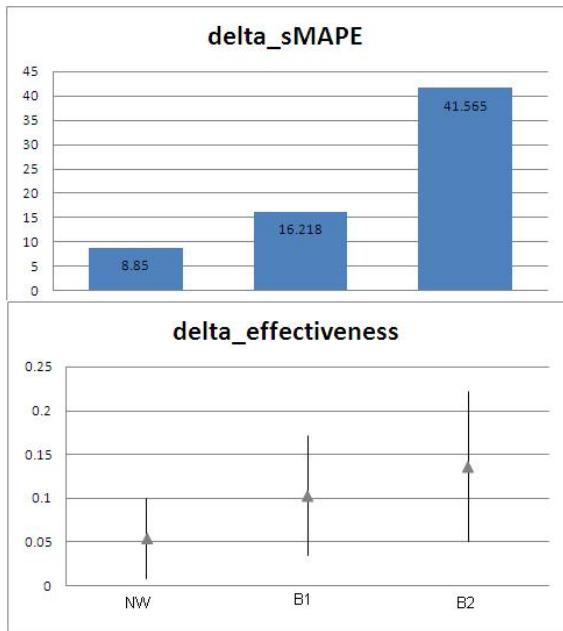
Figure B.2: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Iran.



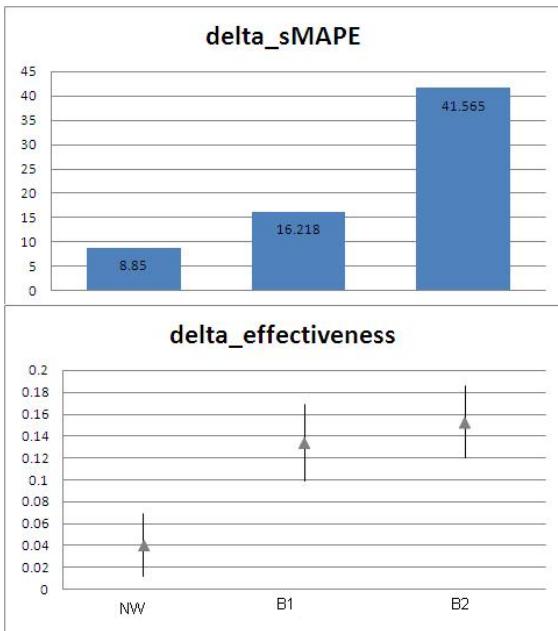
(a) Mitigation Strategy 1



(b) Mitigation Strategy 2



(c) Mitigation Strategy 3



(d) Mitigation Strategy 4

Figure B.3: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 1-4 in Thailand. The prediction percentage error between the baseline models and model W($\Delta sMPAE$) is compared with change in effectiveness of mitigation strategies. The unit of effectiveness is percentage per million US dollars.

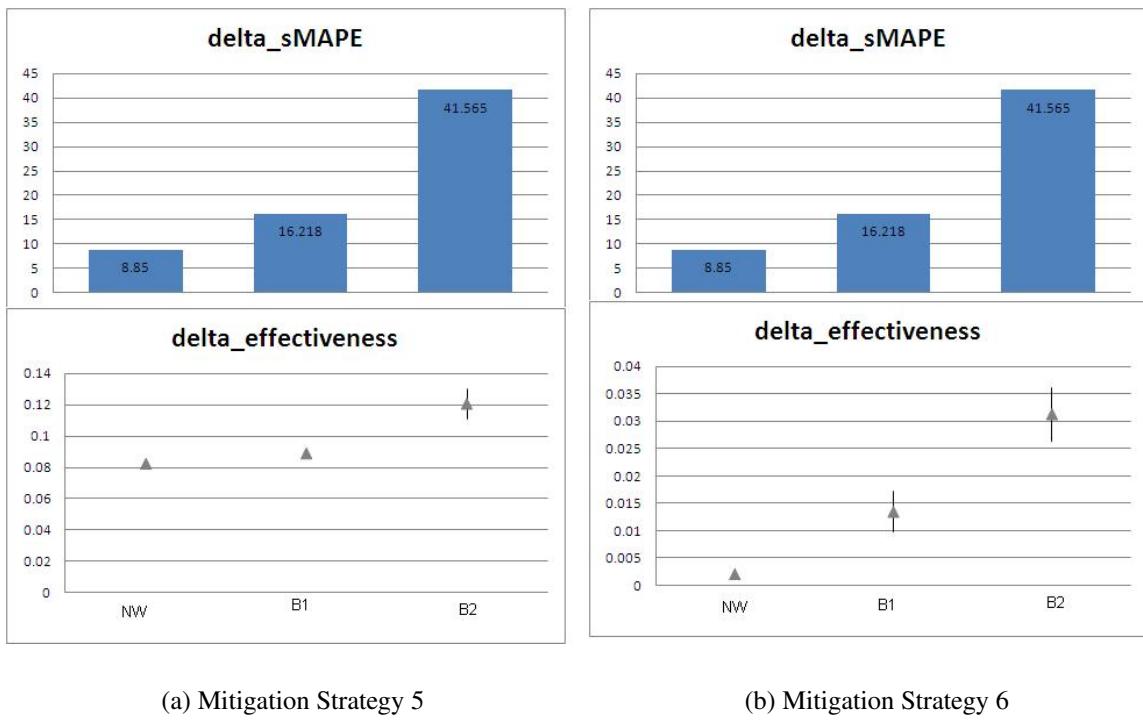


Figure B.4: Impact of prediction errors due to baseline models NW, B1, B2 on mitigation strategies 5,6 in Thailand.