International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Modeling and support tools for studying disease spread in livestock using networks

Joel Francis [a], **Greg Klotz, Neil Harvey and Deborah Stacey**

[a]*Computing and Information Science, University of Guelph, Guelph, Ontario, Canada*
(jfranc01@uoguelph.ca, gklotz@uoguelph.ca, neilharvey@gmail.com, dastacey@uoguelph.ca)

Abstract: The natural occurrence or intentional release of highly contagious agents of livestock disease can have serious consequences for any country's agricultural economy. Successful control and management of animal disease outbreaks require that adequate response strategies be developed beforehand. Disease spread simulation models are being used by veterinary epidemiologists to evaluate strategies for the control of disease spread. These models enable decision-makers and emergency preparedness personnel to explore many different scenarios to determine the effects of control measures such as vaccination, and study the likely size, duration and cost of outbreaks. This paper presents an overview of livestock disease spread modeling using the North American Animal Disease Spread Model. We show results from the unique network contact spread module of the model, as well as a sensitivity analysis that helps expose the differences between different contact spread network models.

Keywords: Biological system modeling; epidemiology; animal disease spread

1 INTRODUCTION

Modeling the spread of infectious disease in livestock is an important tool used by researchers and policy-makers to understand the magnitude of possible outbreaks and to prepare plans to use in an outbreak. The effectiveness of these plans to halt or limit an outbreak depends on the accuracy of the simulations thus requiring complex disease spread models. Using these information technology models enables researchers to better understand the spread and containment of disease, bringing outbreaks under control faster, saving the lives of animals, limiting financial hardship for farmers and restricting the environmental impact of the outbreak and its aftermath.

This paper presents experiments carried out using new modules developed for the North American Animal Disease Spread Model (NAADSM) Harvey et al. [2007], an advanced animal disease simulator developed at the University of Guelph and Colorado State University, in collaboration with the Canadian Food Inspection Agency (CFIA), the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), and the United States Department of Agriculture (USDA). The software is a high performance computing version of Spreadmodel PC, a Foot and Mouth Disease (FMD) simulator developed at the USDA by Schoenbaum and Terry Disney [2003]. The latest version of the software enables veterinary epidemiologists to collaborate on disease spread research by running simulations on a supercomputer through a Web interface. It enables disease spread simulation between herds connected through various contact networks, as well as between individuals in a herd. The model can be loaded with true connection data from a real population to model spread in a specifically connected network of farms. If a modeler does not have the actual network data for a real population they wish to model, they can load network characteristics from a similar population into NAADSM to build a reasonable representation of the target network.

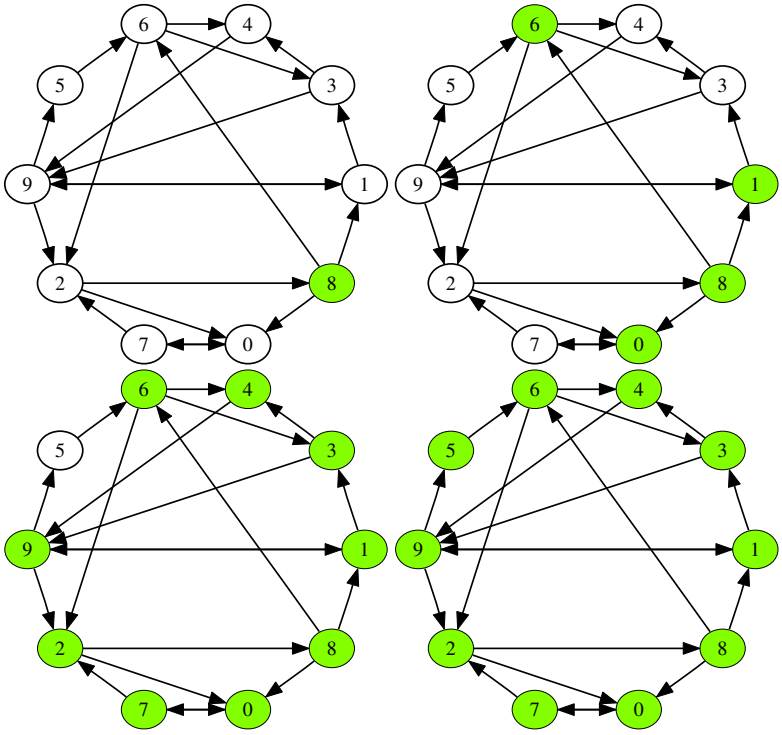

Figure 1: An example of disease spread using the SI model. Infectious herds are shown in green, with the initial infection in herd 8.

2 BACKGROUND

Complex epidemiological information management systems for livestock include EpiMAN (Sanson [1993]), InterSpread Plus (Stevenson et al. [2005]), and NAADSM (Harvey et al. [2007]). These models are widely used to study the outbreak and management of disease in livestock, as demonstrated by Colby and Johnson [2007], Dillon et al. [2007], Dubé et al. [2007], Keeling [2005], Kostova-Vassilevska [2004], Sanson [2004], Stevenson [2003], Yoon et al. [2006].

Disease spread models start with the principle that a population can be divided into a set of distinct classes depending on their experience with the disease. In the simplest case individuals can be either susceptible, infectious or recovered. This is known as the SIR model and shown by Newman [2002] and Woolhouse [2004]. Susceptible individuals have never come into contact with the disease and can catch the disease. When they catch the disease they move into the infectious state and spread the disease to others who are still susceptible. After some period of time the infectious individuals recover and move into the recovered state where they are presumed to be immune for life.

There are variations on the SIR model, such as susceptible-infectious (SI) and susceptible-infectious-susceptible (SIS) demonstrated by Hooper [2008]. In the SI model individuals can be either susceptible or infectious. Infectious individuals never recover and continue to spread the disease forever. NAADSM is generally used to model disease spread from herd to herd instead of individuals, A simple example of this is shown in Figure 1. In the SIS model individuals are either susceptible or infectious. After becoming infectious they recover but instead of becoming immune to the disease, individuals go back to the susceptible state. Introducing more states, such as vaccinated or quarantined is possible in NAADSM. This gives modelers more control over the parameters of their simulations at the cost of increased complexity.

3 CONTACT NETWORK MODELS

This section demonstrates a new feature of NAADSM that allows for different contact network models for how herds (or flocks) are *connected* to each other (shipping, markets, *etc.*). Subsection 3.1 highlights the effects of connecting farms using Random, Power-law and Scale-Free Network (SFN) structures to simulate livestock disease spread through direct contact (movement of infected animals between premises) using the simple SI model. Subsections 3.2 and 3.3 expand on these networks and introduce a new specialized network creation module which can be used to create contact networks for NAADSM. Subsection 3.4 looks at a sensitivity analysis of the parameters as they are affected by different contact networks.

3.1 Effects of connection structure on disease spread

Our experiments were designed to highlight the differences in the time required to infect all herds/flocks connected by different contact (or movement/shipping) networks, and to demonstrate the built in network generator developed for NAADSM. The disease spread parameters were kept simple: herds always infect the other herds they are connected to and an infectious herd never recovers. In the *original* NAADSM, the connection structure between herds (if exact information was not known) was modeled as a random network. We will refer to this as the *original* connection structure. Our new NAADSM connection network module allows for six more connection network types. These new contact networks were used for this experiment, each created so that herds connected to three other herds on average. The networks used were: 3 All (the baseline), Gaussian, Exponential, Exponential shifted (so each herd would connect to at least one other), Pareto and Scale-free (as described by Albert and Barabasi [1999]). These networks were not selected for realism, but to showcase the difference in the epidemics caused by the connections. In the baseline example, all herds are connected to three others, and in the Gaussian, most herds have three connections. The other networks are power-law; most herds have one connection (except Exponential, where most herds do not have any connections), but there are highly connected herds in the network that have dozens, or in the case of the Pareto and Scale-Free, hundreds of connections. Gaussian connections might be useful for modeling decentralized networks, for example farmers all trading with their neighbours, and power-law networks are more appropriate for networks that have large markets. Networks were created at the start of the simulation and stayed constant until the end of each run. The results shown in Figure 2 demonstrate that connection structure dramatically affects the average speed of outbreaks and should be considered when creating simulations.

3.2 Euclidean Distance Modulated Scale-Free Networks

This section introduces a new external module for NAADSM which has the ability to create specialized contact networks that allow NAADSM to use more specialized versions of Scale-Free networks. The notion of geographic travel distance may be important when creating Scale-Free Networks, as modelers may want to create a structure where shipping and receiving takes place between premises that are close to each other or they may want to model shipping and receiving mostly between premises that are far apart. The latest module of NAADSM includes SFNs, but to enable better visualization of distance based disease spread, a modified SFN creation algorithm described by Manna and Sen [2002] has been added. This algorithm includes Euclidean distance as a factor when deciding the probability of attachment.

The algorithm creates the network as follows:

- In a two-dimensional plane, we consider a set of randomly distributed points such that $x <$ upper bound for x and $y <$ upper bound for y.
- The network grows by introducing at each time step "t" one node with randomly chosen co-ordinates. The attachment probability for an incoming node to connect to one of the

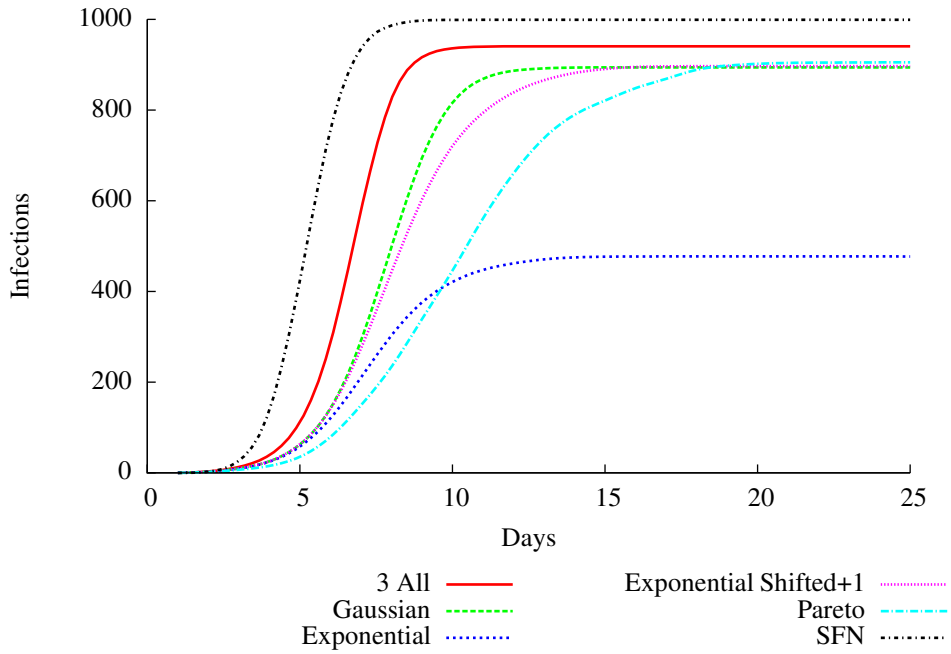

Figure 2: Comparison of average number of infections from disease spread simulation over six types of networks.

existent nodes "i" in the network is given by:

$$\Pi_i(t) \sim k_i(t)\ell^{\alpha} \tag{1}$$

where ℓ is the Euclidean distance between an incoming and existing node and α is a parameter that can be varied. Smaller values of α allow nodes with smaller distance to contribute whereas higher values of α allow nodes further away to contribute.

3.3 Directed Scale-Free Networks

The first definitions of SFNs focused on highly connected hubs but did not consider connection directionality, an important part of disease spread modeling. Our external add-on to NAADSM also creates directed SFNs, as introduced by Krapivsky et al. [2001]. This is needed when modeling industries that display *vertical integration*. In the pork rearing industry, for example, breeders ship piglets to growers once they reach a certain weight and they are shipped on to markets from there. There is no traversing backwards.

This software offers an extension to growing network models that are based on preferential attachment by integrating link directionality. The resulting network structure has power law in- and out-degree distributions. The algorithm follows a two phase approach until the required network size is reached:

- With probability p, a node is added to the network and connected to a node based on its in-degree
- With probability $1 - p$, a link is created between two nodes already existing in the network, considering the out-degree of the originating node and the in-degree of the target node.

This software generates a connection file, which can be imported into NAADSM for simulations.

3.4 Sensitivity Analysis

To investigate the impact of changing the network structure, a fractional factorial experiment was designed with the following input parameters as factors: population density (A), direct contact movement rate (B), indirect contact movement rate (C), probability of infection given exposure by direct contact (D), probability of infection given exposure by indirect contact (E), disease parameters (F), day at first detection (G), probability of trace success (H), detected units before vaccination begins (I), radius of vaccination ring (J), radius of destruction ring (K), delay to vaccine immunity (L), and trace period (M).

The disease parameters *(F)* were chosen to simulate a foot-and-mouth outbreak. Each factor had a lower level and an upper level. An analysis of variance (ANOVA) was conducted using SAS, where the output variables were analysed under the influence of the various factors. The output variables of interest were *Time-to-end-outbreak, Cumulative-infected, Total-detected, Total-vaccinated and Total-destroyed*. For the sake of brevity we focus on the Time-to-end-outbreak. The results give us a clear indication about the factors that can be varied to explore the possible variations in disease spread.

Each model behaved fairly differently due to the inherent properties of the network. To emphasize the effects of each factor and its ordering, we chose to use *Pareto Plots*. A Pareto plot is an ordered histogram that communicates the importance of the effects of each factor on that particular response variable. It orders the difference of the means for each paramter at their respective setting (high or low). A few observations of interest (Figures 3, 4, 5):

- The effect of the disease parameters (F) are more pronounced in the Scale-Free models than compared to the original random network model. A change in the setting from high to low has a statistically significant impact on the newer models.
- The direct and indirect contact movement (C and D) rate is more important in the original and distance based Scale-Free model than compared to the directed Scale-Free model.
- The effect of changing the population density (A) is statistically significant in the case of the distance based Scale-Free network
- The vaccination ring (J) has a small but significant impact only in the case of the directed Scale-Free network
- The effect of tracing (H) is only significant in the original model and has no effect on the Scale-Free models

In terms of the two factor interactions, we see a number of significant interactions for all the models (Figure 6). We note that though some factors were not present as main effects, they happen to appear as significant two factor interactions. We note the significance of AL (population density-delay to vaccine immunity) for the original model and HL (probability of trace success-delay to vaccine immunity) for the distance based scale-free model.

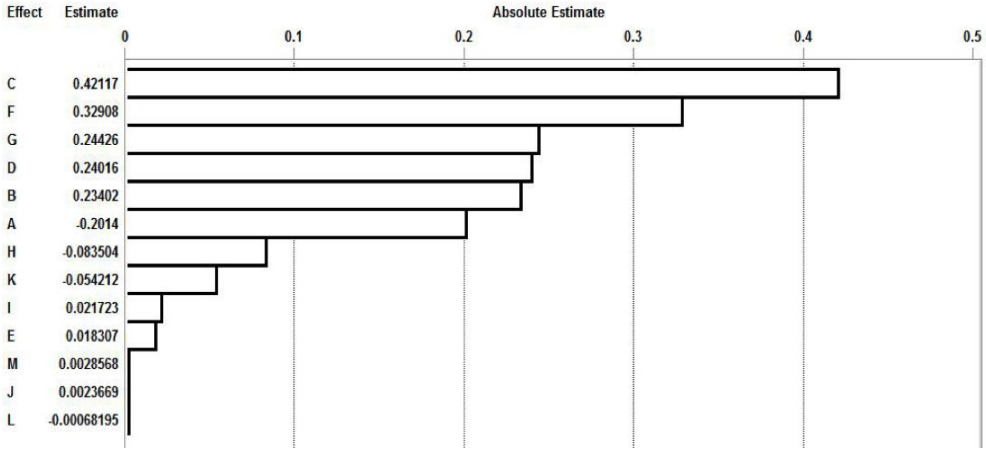

Figure 3: Original Model: Pareto Plot

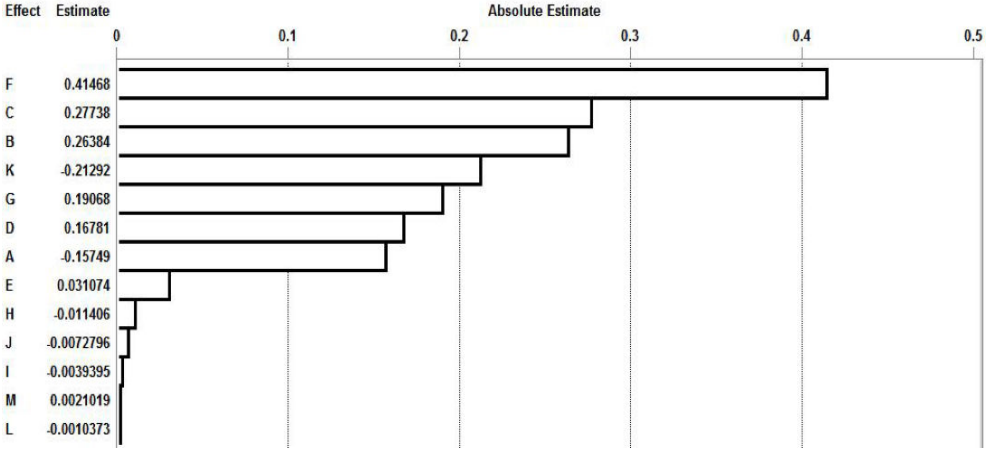

Figure 4: Distance Scale-Free Model: Pareto Plot

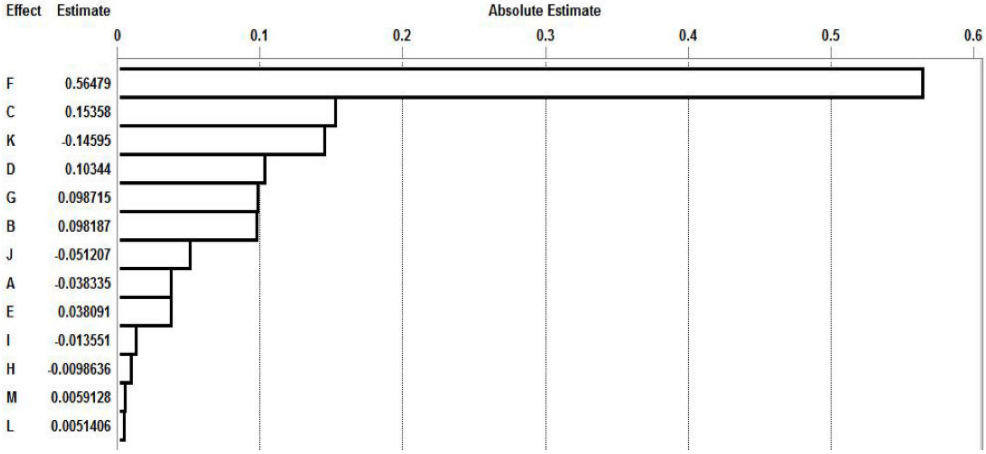

Figure 5: Directed Scale-Free Model: Pareto Plot

Effect	Estimate	Std Error	t Ratio	P Value
A	-0.2014	0.023181	-8.688	<.0001
B	0.23402	0.023181	10.095	<.0001
C	0.42117	0.023181	18.169	<.0001
D	0.24016	0.023181	10.36	<.0001
F	0.32908	0.023181	14.196	<.0001
G	0.24426	0.023181	10.537	<.0001
H	-0.083504	0.023181	-3.6022	0.0004
K	-0.054212	0.023181	-2.3386	0.0202
A*F	0.082268	0.023181	3.5489	0.0005
A*I	0.078237	0.023181	3.375	0.0009
A*L	0.063701	0.023181	2.748	0.0065
B*C	-0.3441	0.023181	-14.844	<.0001
B*D	-0.31245	0.023181	-13.479	<.0001
B*F	0.067059	0.023181	2.8928	0.0042
B*G	-0.083838	0.023181	-3.6167	0.0004
C*D	-0.3491	0.023181	-15.06	<.0001
C*E	-0.13837	0.023181	-5.9692	<.0001
C*F	0.14518	0.023181	6.263	<.0001
C*G	-0.23431	0.023181	-10.108	<.0001
C*H	0.087307	0.023181	3.7663	0.0002
C*K	0.059016	0.023181	2.5458	0.0116
D*F	0.09291	0.023181	4.008	<.0001
E*F	0.05873	0.023181	2.5335	0.0120
E*J	0.25395	0.023181	10.955	<.0001
G*H	0.090655	0.023181	3.9107	0.0001

(a) Original Model

Effect	Estimate	Std Error	t Ratio	P Value
A	-0.15749	0.026968	-5.8399	<.0001
B	0.26384	0.026968	9.7833	<.0001
C	0.27738	0.026968	10.285	<.0001
D	0.16781	0.026968	6.2226	<.0001
F	0.41468	0.026968	15.377	<.0001
G	0.19068	0.026968	7.0705	<.0001
K	-0.21292	0.026968	-7.8951	<.0001
A*B	0.090737	0.026968	3.3646	0.0009
A*C	0.1965	0.026968	7.2864	<.0001
A*F	-0.06185	0.026968	-2.2934	0.0227
A*G	0.069615	0.026968	2.5814	0.0105
A*K	-0.13908	0.026968	-5.1571	<.0001
B*C	-0.27532	0.026968	-10.209	<.0001
B*D	-0.18462	0.026968	-6.8459	<.0001
B*F	0.17195	0.026968	6.3759	<.0001
B*G	-0.080655	0.026968	-2.9907	0.0031
B*K	0.067264	0.026968	2.4942	0.0133
C*D	-0.2615	0.026968	-9.6967	<.0001
C*F	0.27417	0.026968	10.167	<.0001
C*G	-0.067814	0.026968	-2.5146	0.0126
D*F	0.11272	0.026968	4.1796	<.0001
E*F	0.064656	0.026968	2.3975	0.0173
E*J	0.12461	0.026968	4.6208	<.0001
F*G	0.062567	0.026968	2.32	0.0212
G*K	0.058518	0.026968	2.1699	0.0310
H*L	0.12889	0.026968	4.7793	<.0001

(b) Distance Scale-free Model

Effect	Estimate	Std Error	t Ratio	P Value
A	-0.038335	0.015319	-2.5025	0.0130
B	0.098187	0.015319	6.4096	<.0001
C	0.15358	0.015319	10.026	<.0001
D	0.10344	0.015319	6.7527	<.0001
E	0.038091	0.015319	2.4866	0.0136
F	0.56479	0.015319	36.869	<.0001
G	0.098715	0.015319	6.4441	<.0001
J	-0.051207	0.015319	-3.3428	0.0010
K	-0.14595	0.015319	-9.5276	<.0001
A*D	0.034305	0.015319	2.2394	0.0261
A*K	-0.095982	0.015319	-6.2657	<.0001
B*C	-0.18311	0.015319	-11.953	<.0001
B*D	-0.20678	0.015319	-13.499	<.0001
B*E	-0.060276	0.015319	-3.9348	0.0001
B*F	0.1732	0.015319	11.307	<.0001
B*G	-0.045678	0.015319	-2.9819	0.0032
B*J	0.064409	0.015319	4.2046	<.0001
B*K	0.039199	0.015319	2.5589	0.0112
C*D	-0.18677	0.015319	-12.192	<.0001
C*E	-0.091423	0.015319	-5.9681	<.0001
C*F	0.27994	0.015319	18.274	<.0001
C*G	-0.066536	0.015319	-4.3434	<.0001
C*J	0.061223	0.015319	3.9966	<.0001
C*K	0.082573	0.015319	5.3903	<.0001
D*E	-0.053832	0.015319	-3.5142	0.0005
D*F	0.16456	0.015319	10.742	<.0001
D*G	-0.053532	0.015319	-3.4945	0.0006
D*J	0.065091	0.015319	4.2491	<.0001
E*F	0.038082	0.015319	2.486	0.0137
E*J	0.16379	0.015319	10.692	<.0001
F*G	0.054359	0.015319	3.5485	0.0005
F*J	0.045984	0.015319	3.0018	0.0030

(c) Directed Scale-free Model

Figure 6: ANOVA Tables

4 CONCLUSIONS

This paper gave examples of livestock disease spread simulation using NAADSM, highlighting the model's new ability to simulate various connection/contact structures. The effects of different connection structures on the spread of disease in animals was demonstrated, showing that connection structure has a significant effect on the speed of the outbreak and the number of herds infected. This shows that it is worthwhile to conduct a survey of the farms and markets being modeled in order to understand their connection structure and use software that can use these connections to run simulations.

The NAADSM model (http://www.naadsm.org/) is free software available under the GPL. The features of importing and creating connections into NAADSM are currently being integrated into the model, and will be available for download shortly.

ACKNOWLEDGMENTS

This research was made possible by grants from the Poultry Industry Council of Canada, the Ontario Ministry of Agriculture and Food and Rural Affairs and by the United States Department of Agriculture.

REFERENCES

Albert, R. and A. Barabasi. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, 1999.

Colby, M. and Y. Johnson. Potential uses for geographic information system-based planning and decision support technology in intensive food animal production. *Animal Health Research Reviews*, 3(01):31–42, 2007.

Dillon, E., A. Matthews, and F. Thorne. Foot-and-Mouth Disease control costs compared: An Irish case study. 2007.

Dubé, C., M. Stevenson, M. Garner, R. Sanson, B. Corso, N. Harveyl, J. Griffin, J. Wilesmith, and C. Estrada. A comparison of predictions made by three simulation models of foot-and-mouth disease. *New Zealand Veterinary Journal*, 55(6):280–288, 2007.

Harvey, N., A. Reeves, M. Schoenbaum, F. Zagmutt-Vergara, C. Dubé, A. Hill, B. Corso, W. McNab, C. Cartwright, and M. Salman. The North American Animal Disease Spread Model: A simulation model to assist decision making in evaluating animal disease incursions. *Preventive Veterinary Medicine*, 82(3-4):176–197, 2007.

Hooper, R. Epidemic Modelling: An Introduction, 2008.

Keeling, M. Models of foot-and-mouth disease. *Proceedings: Biological Sciences*, 272(1569): 1195–1202, 2005.

Kostova-Vassilevska, T. On The Use Of Models To Assess Foot-And-Mouth Disease Transmission And Control. Technical report, UCRL-TR-205241, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2004.

Krapivsky, P., G. Rodgers, and S. Redner. Degree distributions of growing networks. *Physics Review Letters*, 86(123):5401–5404, 2001.

Manna, S. and P. Sen. Modulated scale-free network in euclidean space. *Physics Review*, 66(6), 2002.

Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1):016128, Jul 2002.

Sanson, R. *The development of a decision support system for an animal disease emergency*. Massey University,[Department of Veterinary Clinical Sciences], 1993.

Sanson, R. The Use of GIS in Epidemic Disease Response. *GIS and Spatial Analysis in Veterinary Science*, 2004.

Schoenbaum, M. and W. Terry Disney. Modeling alternative mitigation strategies for a hypothetical outbreak of foot-and-mouth disease in the united states. *Preventive Veterinary Medicine*, 58(1-2):25–52, 2003.

Stevenson, M. *The Spatio-temporal Epidemiology of Bovine Spongiform Encephalopathy and Foot-and-Mouth Disease in Great Britain*. PhD thesis, Ph. D. Thesis, Massey University, New Zealand, 2003.

Stevenson, M., R. Sanson, M. Stern, B. OLeary, G. Mackereth, M. Sujau, N. Moles-Benfell, and R. Morris. InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. *Technical paper for Research Project BER-60-2004. Ministry of Agriculture and Forestry, Biosecurity New Zealand, Wellington*, 48, 2005.

Woolhouse, M. Chapter 13 Mathematical Models of the Epidemiology and Control of Foot-and-Mouth Disease. *Foot and Mouth Disease: Current Perspectives*, 2004.

Yoon, H., S. Wee, M. Stevenson, B. OLeary, R. Morris, I. Hwang, C. Park, and M. Stern. Simulation analyses to evaluate alternative control strategies for the 2002 foot-and-mouth disease outbreak in the Republic of Korea. *Preventive Veterinary Medicine*, 74(2-3):212–225, 2006.