The Critical Community Size and Extinctions for Livestock Diseases

Katherine Broadfoot[1], Henry Charlesworth[1], Neil Foster[1], Joe Hilton[1], Matt Keeling[2], and Simon Gubbins[3]

[1]Centre for Complexity Science, University of Warwick
[2]WIDER Centre, Mathematics Institute and School of Life Sciences, University of Warwick
[3]The Pirbright Institute, Pirbright Laboratory

Abstract

Foot-and-mouth disease (FMD) is an exotic livestock disease in the UK and much of Europe, that is endemic in many other regions. It is of interest to understand the persistence behaviour of FMD and determine the conditions under which it can become endemic. Influenced by work linking the persistence behaviour of measles to population size, we explore this connection for FMD. We use a stochastic compartmental model that operates at the level of individual animals to study the dynamics of FMD within a single farm. This model defines a continuous time Markov chain and this allows us to determine the complete ensemble of behaviour for single farms up to a population of 1,000. Above this size, we use direct stochastic simulation. For individual farms we begin to observe persistence behaviour in simulations of farms of size 500,000 and larger; to achieve persistence probabilities of 0.1 and 0.5, respective population sizes of approximately 1.5 million and 15 million are required. We then extend our compartmental model to capture the behaviour of FMD on networks of interacting farms, however this is slow for large networks of farms because we model the dynamics within the farms at an individual animal level. To combat this, we use the single farm ensemble results to inform a model which operates solely at the farm level without explicitly modelling the behaviour of cattle within the farms. We establish that this inter-farm model has similar qualitative behaviour to the individual level compartmental model and use it to investigate the persistence behaviour of FMD on large networks of farms. For farms arranged in a square lattice, we observe no long term persistence behaviour in simulations and for a random network we find that persistence is highly dependent on the contact rates between farms. We also find that FMD is more likely to persist in populations when they are divided into a network of farms as opposed to when they are homogeneously mixed.

1 Introduction

Foot-and-mouth disease (FMD) is a livestock disease which in the UK causes intermittent large-scale outbreaks with vast economic impact; the 2001 epidemic is estimated to have cost the UK economy £6-9 billion [1]. Although it is an exotic disease in the UK, meaning that it relies on reintroduction via external import to establish a new epidemic, it is endemic in many other regions [16, 17]. FMD is highly infective [4, 18] and we might therefore expect it to exhaust its local supply of susceptible individuals and then go extinct. It is therefore of practical interest to explain the endemic nature of FMD in certain territories by studying its persistence behaviour and reconciling this with its high level of infectivity.

Previous research on the persistence behaviour of measles suggests the existence of a *critical community size* [11]. This is a critical value of the population size, below which the disease goes extinct and above which it persists. It has also been suggested that the at-risk population size is the primary variable influencing the size of FMD epidemics [19]. With this in mind, our aim is to investigate the persistence of FMD in populations of cattle in relation to population size, and specifically to estimate its critical community size. Whilst the work on measles focused on single large populations, livestock populations are separated between

farms, each containing a relatively small number of animals. This spatial structure motivates the use of network models for interactions between farms, alongside more traditional models for interactions within the farm population. In particular, we are not solely interested in the relationship between population size and persistence, but also in the effect which the distribution of livestock has on persistence of FMD. For measles, persistence occurs in the local population when a small baseline level of infection is present outside of epidemic peaks. In the case of FMD, we are interested in persistence across a set of connected subpopulations, which occurs when the infection is continuously present within this collection of subpopulations.

2 Epidemics on a single farm

2.1 The SIR model

To study the dynamics of FMD in a single isolated cattle population (corresponding to a single farm or the population of cows in a geographic population under the assumption of homogeneous mixing), we take a compartmental SIR model [15] as our starting point. We denote by S the number of individuals who are susceptible to FMD, I the number infected with FMD, and R the number who have recovered and are immune to reinfection. The model incorporates the demographic processes of birth and death, but keeps population at the constant size $N := S + I + R$ by assuming that dead animals from each state are immediately replaced by new susceptibles. Analysis of survey data indicates that average herd size in the UK is somewhere between one and two hundred [3], and so, apart from when we wish to study the effect of varying herd size, we will usually take N to be 100. Deterministically, the dynamics are given by:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} + \mu I + \mu R, \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I - \mu I, \\ \frac{dR}{dt} &= \gamma I - \mu R, \end{aligned} \tag{1}$$

where β is the contact rate, γ is the recovery rate and μ is the demographic death rate (we assume no mortality due to disease as this is rare in adult livestock [10]). Throughout this report the values $\beta = 5.26$, $\gamma = 1/3$ are used, as suggested by previous experimental and statistical work [4, 9]. We set $\mu = 0.001$, corresponding to an average lifespan of approximately three years. The basic reproductive ratio, R_0, is the average number of secondary infections from the primary infection of a single individual in a susceptible population [2]. This can be approximated from equations (1) as $R_0 = \beta/\gamma$ [12], giving a basic reproductive ratio of approximately 16 for our parameter values.

These dynamics can be modelled stochastically using a continuous time Markov chain with four events, corresponding to infection of susceptible individuals by infecteds, recovery of infected individuals, and deaths of infected and recovered individuals. The Markov chain is as follows:

$$\begin{aligned} (S, I, R) &\longrightarrow (S-1, I+1, R) \quad \text{with rate} \quad \frac{\beta SI}{N}, \\ (S, I, R) &\longrightarrow (S, I-1, R+1) \quad \text{with rate} \quad \gamma I, \\ (S, I, R) &\longrightarrow (S+1, I, R-1) \quad \text{with rate} \quad \mu R, \\ (S, I, R) &\longrightarrow (S+1, I-1, R) \quad \text{with rate} \quad \mu I. \end{aligned} \tag{2}$$

We can simulate this process using the Gillespie algorithm [13]. A typical realisation is shown in figure 1, where the initial number of susceptible individuals quickly become infected. After this initial burst of infection, the low number of susceptible individuals available for infection results in a rapid decrease in the number of infected.

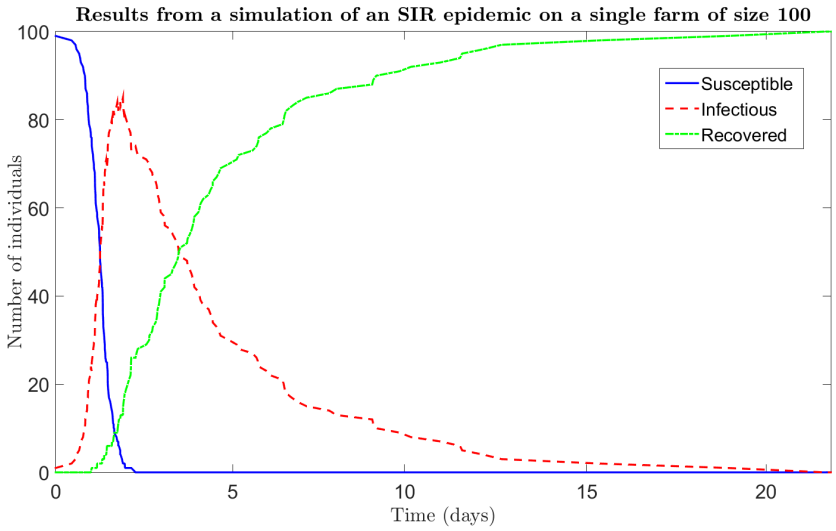

Figure 1: Results of a single realisation of the stochastic SIR model on a single farm of 100 individuals initialised with 1 infected and 99 susceptible individuals. The infection peaks very quickly and has died out after about 20 days.

Because the population size is kept constant, the number of recovered individuals at a given time is uniquely determined by the number of susceptible and infected individuals at that time; the problem can be reduced from three dimensions to two. The master equation for the process is given by

$$\begin{aligned}\frac{dp_{S,I}}{dt} &= \left[\frac{\beta(S+1)(I-1)}{N}\right] p_{S+1,I-1} + [\gamma(I+1)]p_{S,I+1} \\ &+ [\mu(N-(S-1)-I)]p_{S-1,I} + [\mu(I+1)]p_{S-1,I+1} \\ &- \left[\frac{\beta SI}{N} + \gamma I + \mu(N-S-I) + \mu I\right] p_{S,I},\end{aligned} \tag{3}$$

where $p_{S,I}$ is the probability of being in a state with S susceptible and I infected individuals. There are $\frac{1}{2}(N+1)(N+2)$ possible states for the system and, as in [14], $p_{S,I}$ can be mapped to the $f(S,I)$th entry of a one dimensional vector, $\boldsymbol{p}$, where $f(S,I) := NS - \frac{1}{2}(S(S-3)) + I + 1$. Equation (3) can then be written as

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{p}Q, \tag{4}$$

where Q is the generator of the process. This is the Kolmogorov forward equation for the chain [8].

Solving the master equation forward in time gives the probability distribution for the state of the system as a function of time, $p_{S,I}(t)$. A variety of results can be obtained directly from the distribution. First of all, we can calculate the rate of extinction as a function of time using the following formula:

$$\frac{d}{dt}\sum_{S=0}^{N} p_{S,0}(t). \tag{5}$$

The dashed curves in figure 2 give the results obtained from this formula for cattle populations of size 10, 100 and 1,000. The distributions for all sizes are bimodal, with an initial peak corresponding to the epidemic failing to establish itself, and a second peak corresponding to extinction following a full epidemic. Whilst failure to establish an epidemic is consistent for all three farm sizes, larger farms will on average experience longer epidemics, with more variation in duration than smaller farms. The solid lines are results gained from direct stochastic simulation of the system. The simulation results agree closely with the exact ensemble

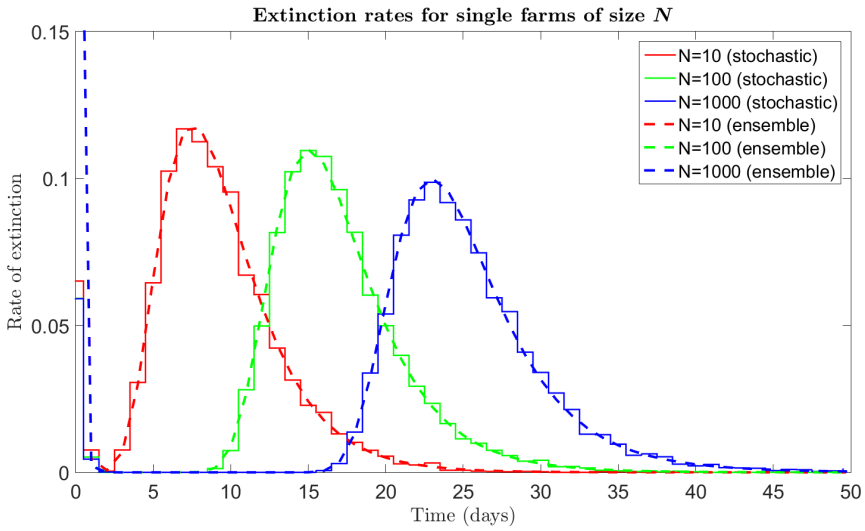

Figure 2: Extinction rates for farms of sizes $N = 10, 100$ and $1,000$, obtained through both ensemble methods and stochastic simulations, initialised with one infected and N-1 susceptible individuals. There is a high peak at t=0, due to the chance of the first event being the loss of the single initial infected individual. The larger the farm the later the second peak occurs in the extinction rates.

results. For farms with very large livestock populations, the master equations are not amenable to numerical analysis, and the agreement between direct simulation and the ensemble results justifies the use of simulation averages to approximate ensemble behaviour. For a farm of size N there are $\frac{1}{2}(N+1)(N+2)$ differential equations to solve, making direct simulation the more efficient method for obtaining results in very large systems.

Another helpful way to study the persistence of the disease is to plot the survival curve, given by

$$1 - \sum_{S=0}^{N} p_{S,0}(t). \tag{6}$$

This gives the probability that the infection survives up to time t. For small population sizes we can use the time evolution of the master equations to obtain the survival curve. For larger farms, we use direct simulation, the results of which are plotted in figure 3. The curves have a flat initial trajectory corresponding to the low probability of extinction in the early stages of the epidemic (provided the epidemic initially establishes itself), followed by a steep descent corresponding to extinction following an epidemic. For smaller farms, the probability that complete extinction occurs after the epidemic is close to one, and the survival curve reaches zero. This is the case for the three farm sizes for which we plotted the extinction rates in figure 2. For larger farms, starting around half a million cattle, the probability of survival levels out above zero following the period corresponding to an initial epidemic. This behaviour corresponds to some observing persistence of FMD following the initial epidemic in these simulations.

Using direct simulation we can plot the probability of persistence as a function of population size. We specifically plot the probability of survival past $1,000$ days, which is significantly longer than the duration of epidemics we observe in simulation, and so survival past this time is a good indicator of long-term persistence. This plot is given in figure 4. We begin to observe persistence in our simulations for farms which have a cattle population of greater than approximately 500,000. For population sizes greater than approximately 1.5 million the probability of persistence is greater than 0.1 and for populations greater than approximately 15 million the probability of persistence is greater than 0.5.

It is clear from figure 4 that any individual farm will be too small for FMD to persist within it. Comparing these results to data on cattle populations in different countries can give us an estimate for the probability of persistence given the (unrealistic) assumption of homogeneous interaction within the country's cattle population. The cattle population of the UK is approximately 9.7 million [5], meaning that under free mixing of cattle we would expect the probability of FMD establishing itself as an endemic disease after being imported to the UK to be around 10%. India, Brazil and China all have cattle populations in the hundred millions [6], and would thus stand a high chance of supporting endemic FMD under homogeneous mixing of cattle. Indeed, China and India both have a very high incidence of FMD amongst cattle and are subject to outbreaks throughout the year [19]. Brazil has a lower incidence of FMD and experiences only sporadic outbreaks, a fact which has been attributed to South America's relative geographic isolation and the FMD free status of much of North America [19]. These predictions are the result of the very specific assumption of homogeneous mixing, ignoring any spatial structure which the population may possess. In practice, cattle movements should be far from homogeneous. To gain more realistic insight into persistence, we need to take into account the fact that real world cattle populations are distributed spatially across farms, with lower levels of interaction between farms than within them.

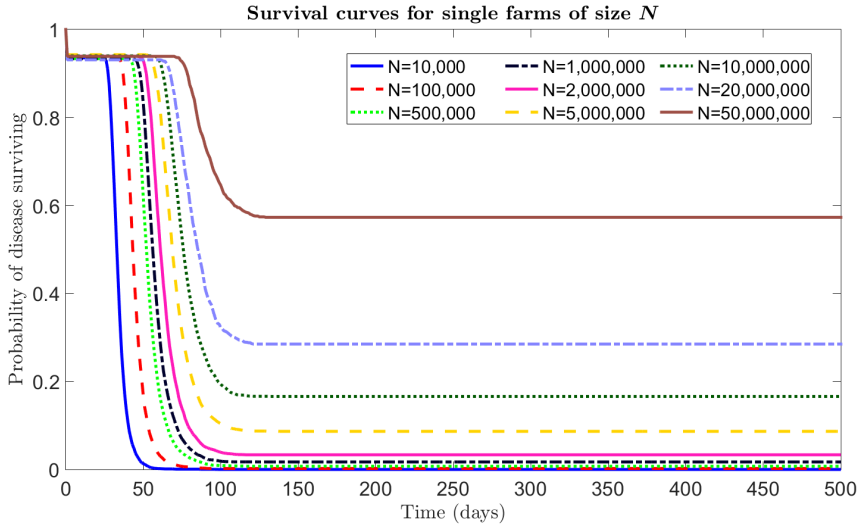

Figure 3: Survival curves for farms of varying sizes, obtained through stochastic simulations.

2.2 Imports of infection

So far we have considered individual isolated populations of cattle. In practice, cattle populations are distributed across farms, with different levels of movement and interaction between and within individual farms. To begin with, we consider a fully coupled system of identical farms; for any farm in this system the rate of recovery and the rates relating to demographic processes on the farm are the same as in (2), but the rate of infection is now given by $\frac{\beta SI}{N} + \sigma \frac{\beta S\bar{I}}{N}$, where $\bar{I}$ is the average number of infected cattle on a farm and σ is a constant corresponding to the level of interaction between the farms.

These rates define a new Markov chain for a single farm in the system. As for the single population case, we can write down the Kolmogorov forward equation (4) for this chain and use the generator to calculate its stationary distribution. The stationary distribution, $\boldsymbol{p}^*$, can be calculated by finding the unique left eigenvector of Q corresponding to the eigenvalue $\lambda_1 = 0$ [14]. We assume that our system contains a large (infinite) number of farms with homogeneous interactions between them, so all farms experience the same interactions and evolve to the same stationary distribution. Therefore the stationary distribution of a single farm in the system can be used to calculate a new mean value of infected cattle per farm across the whole system and

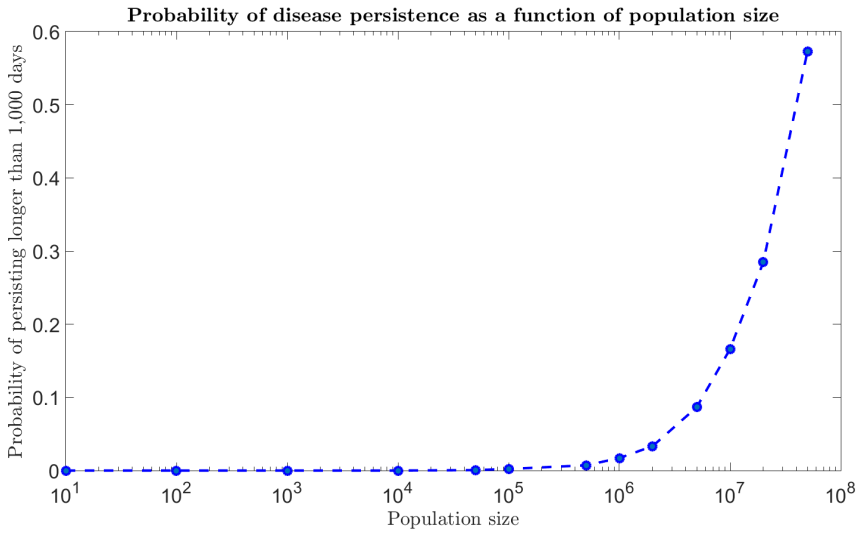

Figure 4: Probability of persistence past the $1,000$ day mark for homogeneously distributed cattle populations of varying sizes.

hence we can use an iterative procedure to find an equilibrium mean level of infection across the entire system.

We begin by choosing an estimate for $\bar{I}$ and finding the stationary distribution of the resulting Markov chain for a single farm in the system. This gives us a probability distribution from which we can calculate the mean number of infected cattle. If this is suitably close (within machine precision) to $\bar{I}$, then we accept this value as the equilibrium. Otherwise, replace $\bar{I}$ with the newly calculated mean, and define a new Markov chain using this choice of $\bar{I}$. Iterating this procedure, we obtain a stable value of $\bar{I}$, and take this to be the mean number of infected individuals on each farm across our homogeneous network.

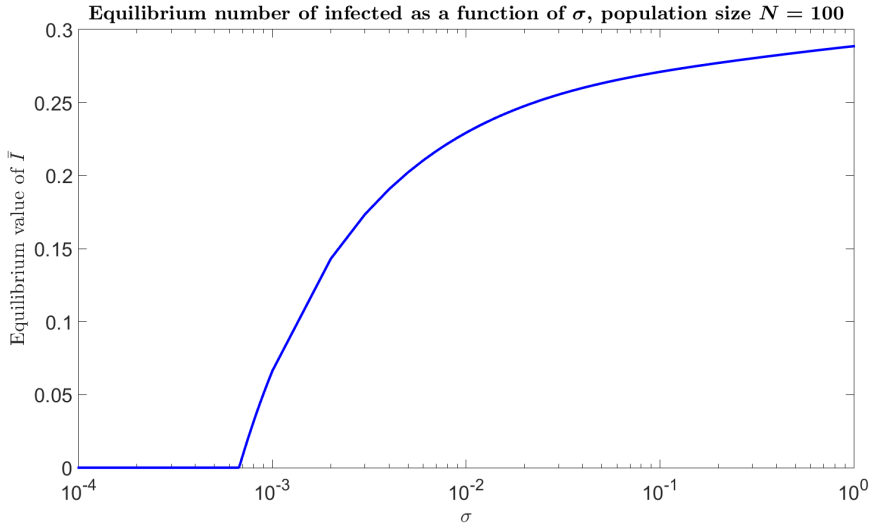

Figure 5: Average number of infected cattle per farm, $\bar{I}$, as a function of contact rate, σ, obtained through the iterative procedure defined in the body of the text. For low levels of contact we see global extinction of FMD, whilst for larger values of σ we observe persistence.

The result of calculating the equilibrium number of infected animals for varying values of σ for farms of

size 100 (recall that we motivated this choice of farm size in the previous section) is shown in figure 5. For low levels of contact between farms, the equilibrium number of infecteds is zero, corresponding to global extinction. As σ passes between 6×10^{-4} and 7×10^{-4} a phase transition occurs, and the equilibrium number of infecteds begins to increase with σ. These positive values of $\bar{I}$ correspond to equilibria in which FMD is present in the ensemble of farms. Thus for large homogeneously connected networks of farms, we can conclude that persistence behaviour is dependent on the strength of the interactions between the farms. This relationship supports the use of movement bans as a control measure for FMD, and indicates that persistence could be prevented if a very low level of inter-farm contact is achieved. However, it gives us no information on how long it takes to reach equilibrium for different values of σ, and so it may be that even when $\bar{I}$ is zero at equilibrium, the population experiences a large epidemic first. It is also worth bearing in mind that this methodology utilises a highly simplified view of the spatial structure of livestock populations, and so any conclusions we get from it should be treated very tentatively.

In addition to iteratively finding a consistent value for $\bar{I}$, the system can be solved forward in time, with the value for $\bar{I}$ being updated at each time step. However, this process is no longer linear and so must be written as

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{p}\left(Q + R\bar{I}\right), \tag{7}$$

where Q is the same generator as in equation (4) and $R\bar{I}$ contains the terms changing the infection terms from $\frac{\beta SI}{N}$ to $\frac{\beta SI}{N} + \sigma\frac{\beta S\bar{I}}{N}$. As the system is solved forward in time, $\bar{I}$ is calculated at each time step, updating the system of ODEs.

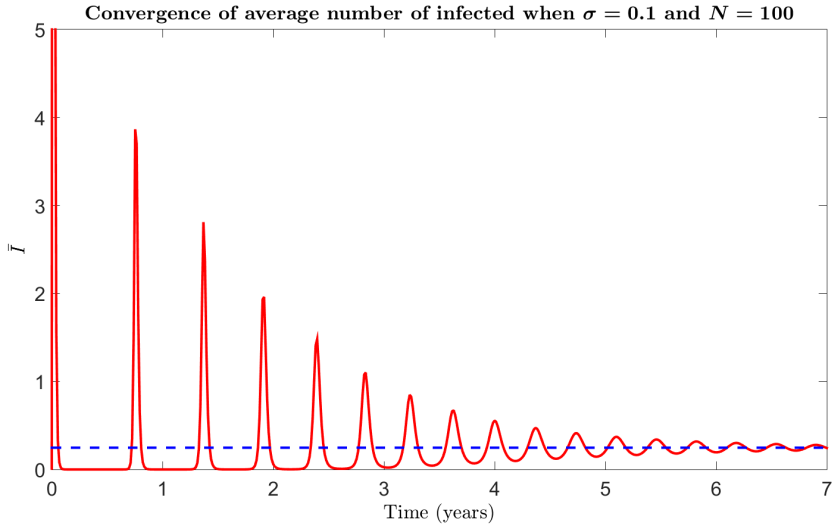

Figure 6: Convergence of the average number of infected individuals, shown in red, obtained through solving the system of ODEs forward in time numerically and updating the system at each time step, to the value obtained iteratively via stationary distribution (blue dashed line). The first peak in fact reaches $\bar{I} = 40.3296$ and converges through damped oscillations to an equilibrium value of $\bar{I} = 0.2477$.

Figure 6 shows the real-time evolution of $\bar{I}$ when $\sigma = 0.1$, initialised with $p_{100,0} = 0.99$, $p_{99,1} = 0.01$ and $p_{S,I} = 0$ otherwise, giving an initial average number of infected of $\bar{I} = 0.01$. From figure 5, we expect convergence of $\bar{I}$ to a non-zero value determined by the iterative procedure explained above, as σ is above its critical value. The value of $\bar{I}$ increases from its initial value then rapidly decreases until it is very close to zero. However, in less than a year, it again increases and continues to oscillate around the non-zero equilibrium value until converging to this value after approximately seven years.

3 Models of networks of interacting farms

We now extend our stochastic compartmental model to describe epidemics on networks of connected farms. Denoting the state of farm i by (S_i, I_i, R_i), we define a Markov chain with the following transitions:

$$\begin{aligned}
(S_i, I_i, R_i) &\longrightarrow (S_i - 1, I_i + 1, R_i) \quad \text{with rate} \quad \frac{\beta S_i I_i}{N_i} + \sum_{j \neq i} \sigma_{ji} \frac{\beta S_i I_j}{N_j}, \\
(S_i, I_i, R_i) &\longrightarrow (S_i, I_i - 1, R_i + 1) \quad \text{with rate} \quad \gamma I_i, \\
(S_i, I_i, R_i) &\longrightarrow (S_i + 1, I_i, R_i - 1) \quad \text{with rate} \quad \mu R_i, \\
(S_i, I_i, R_i) &\longrightarrow (S_i + 1, I_i - 1, R_i) \quad \text{with rate} \quad \mu I_i,
\end{aligned} \tag{8}$$

with i ranging over the complete set of farms. Here σ_{ij} is the (i,j)th entry of the weighted adjacency matrix for the network, corresponding to the level of contact which farm j receives from farm i. We will specifically consider the cases of a ring lattice, a square lattice, a complete graph, and a random graph generated by placing random points in space and connecting each point to a set number of its nearest neighbours.

Whilst it is still possible to simulate this chain using the Gillespie algorithm, the large number of events for multiple farm systems means that the time step between events becomes very small. As the rates have to be recalculated after each time step, it becomes inefficient to simulate systems of many farms this way. As an alternative, we use a modification of the Gillespie algorithm known as the τ-leap method [7]. This method assumes the rates of each event in the process to be constant over the time interval $[t, t+\tau)$ for some choice of τ, so that the number of times an event with associated rate λ_i at time t happens in this interval is distributed according to a Poisson distribution with parameter $\tau\lambda_i$. We choose the number of occurrences of each event during the interval $[t, t+\tau)$ from the appropriate distribution and then update the system accordingly. We thus only need to calculate the system's rates at each constant time step, allowing for more efficient simulation of large networks of farms.

Figure 7 shows simulation results for four different network structures. In each case we scale the $\sigma_{i,j}$ values according to the out-degree of the nodes, resulting in comparable total interaction across the network in all cases. We see from the figure that the network structure plays a role in how the infection spreads, spreading more slowly over networks with fewer connections between nodes.

3.1 Inter-farm interactions

Direct simulation of the internal dynamics of every farm on a network is computationally intensive, even with the τ-leap algorithm, simulation of large numbers of farms is too slow to be feasible. We would therefore like to model the interactions between farms at the farm level, without explicit simulation of the internal dynamics on each farm. We take herd size to be $N = 100$ for all farms, and explore the dependence of global persistence on the number and distribution of such farms.

To this end, we refer back to the master equations (3) for a single isolated farm. We evolve the master equations forward in time to obtain the probability distribution of extinction times and average number of infected cattle as a function of time for a single farm of population 100, starting from one infected and for each possible initial number of susceptibles (ranging from 1 to 99). We can then define an infection process at the farm level as follows: at the moment of infection, we select an extinction time t_{end} according to the probability distribution defined by the extinction curve. Denoting the infected farm by i, each farm j is subject to a force of infection proportional to the average number of infected cattle on the farm during the epidemic,

$$\lambda_{ij} = \frac{\sigma_{ij}\beta S_j}{N t_{end}} \int_0^{t_{end}} I_i dt. \tag{9}$$

Infection of farm j by farm i then occurs with probability $p = 1 - e^{-\lambda_{ij} t_{end}}$. We choose an infection time by normalising the average number of infecteds over time on farm i up to time t_{end} to give a probability

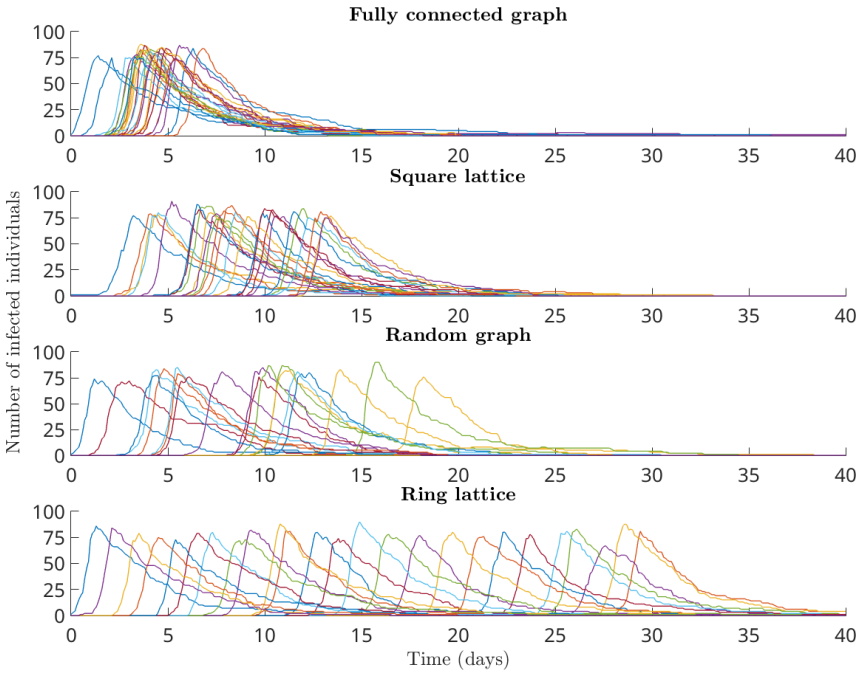

Figure 7: Direct simulation results of the stochastic SIR model (8) on different networks. For each network 25 farms of population 100 were simulated, initialised with a single infected individual on one farm, with all other individuals being susceptible. For the fully connected graph we have $\sigma_{i,j} = \frac{1}{24\times 100}$, for the square lattice $\sigma_{i,j} = \frac{1}{\deg(\mathrm{i})\times 100}$, for the random graph (connected to three nearest neighbours) $\sigma_{i,j} = \frac{1}{3\times 100}$ and for the ring lattice $\sigma_{i,j} = \frac{1}{100}$.

distribution (discretised so that our base unit of time is one day), such that the disease is most likely to be transmitted whenever the number of infecteds is highest. This is an approximation which should work well for low enough values of σ. It is also assumed that when the infection reaches a farm, unless the disease dies out immediately all individuals are infected, so that at t_{end} the population of farm i is entirely made up of individuals in the recovered class (see appendix A). Demographic turnover happens deterministically, so that recovered individuals are replaced by susceptibles until there are no individuals left in the recovered class.

This model allows us to explore a more ambitious range of population sizes and structures at the expense of fine details. The ring lattice and fully connected graph in figure 7 are extreme cases and are unlikely to be representative of a realistic population. We begin by considering a square lattice with periodic boundary conditions. Survival curves for lattices up to size 10×10 are shown in figure 8, calculated using both the explicit internal dynamics and the farm level model. Whilst there is reasonable qualitative agreement between the two models, the farm level model consistently overestimates the probability of survival up to a given time. This overestimate appears to increase with lattice size, meaning that whilst the farm level model can give us insights into the general relationship between number of farms and persistence, we should be wary of any specific numerical predictions from this model. The discrepancies between the two models is likely to be primarily because of the assumption that the time that a farm gets infected by a neighbouring farm depends only on the distribution of the number of infected on that farm. This means that when the probability of spreading is high right at the beginning of the infection this model still tends to introduce a delay until the average number of infecteds is higher. An improvement to this model which provides better agreement with the full simulations is described in appendix B, and further study using this amended model would improve estimates of persistence on large networks of farms.

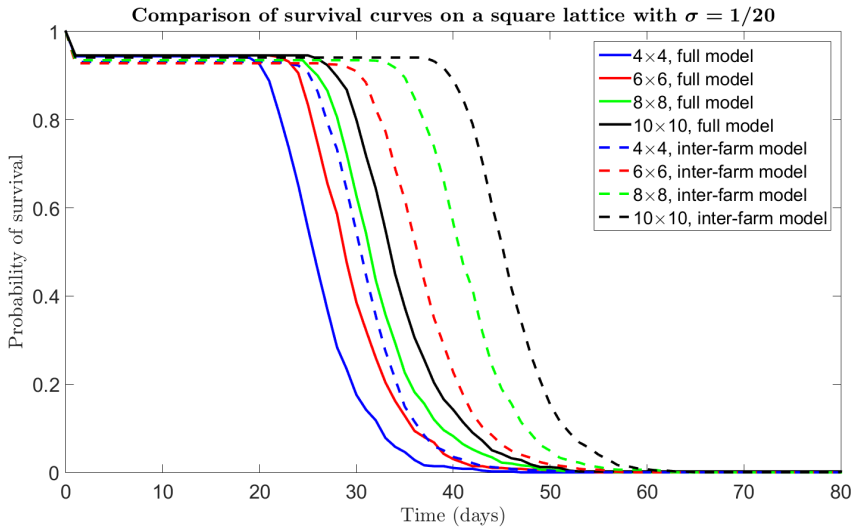

Figure 8: Comparison of the survival curves for different numbers of farms connected in a square lattice, with each farm of size $N = 100$, using the full compartmental model and the inter-farm model. Both models have similar behaviour, although the inter-farm model appears to lag behind the full network model.

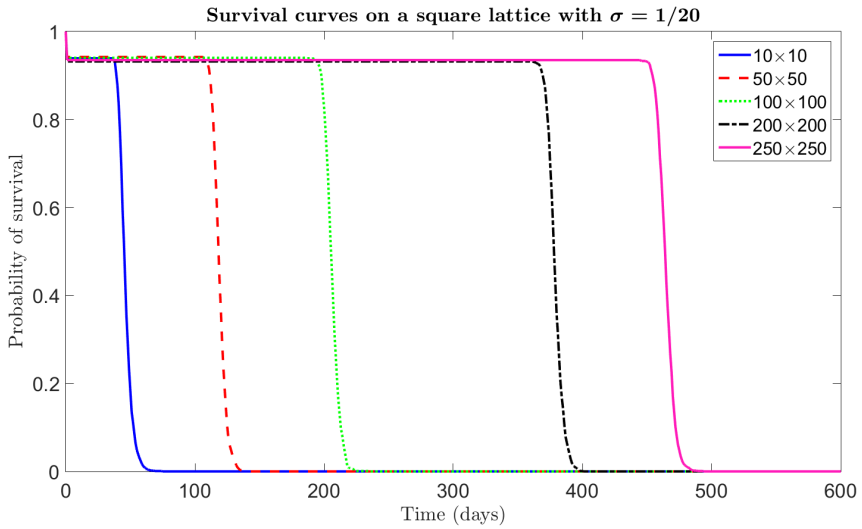

Figure 9: Survival curves for a number of lattices with large numbers of farms, each of size $N = 100$ with $\sigma = 1/20$ using the inter-farm model. Even for the largest of these lattices the probability of survival eventually reaches zero.

From figure 9 we can see that for lattices up to size 250×250 global extinction is observed in all simulations. Recall that from figures 3 and 4 we observed that under homogeneous mixing a population of this size resulted in a non-negligible probability of survival. To gain further insight into the behaviour of the epidemic on a lattice of farms, we refer to the snapshots of the time evolution of the system shown in figure 10. The epidemic begins in the top left corner of the lattice and spreads across it in a doughnut-shaped wave. The wave eventually reaches the centre of the graph, and is then unable to spread back out from this centre since it is surrounded by patches in which the infection has exhausted its supply of susceptibles.

A regular lattice structure does not allow FMD to persist, even in very large populations. Such a structure

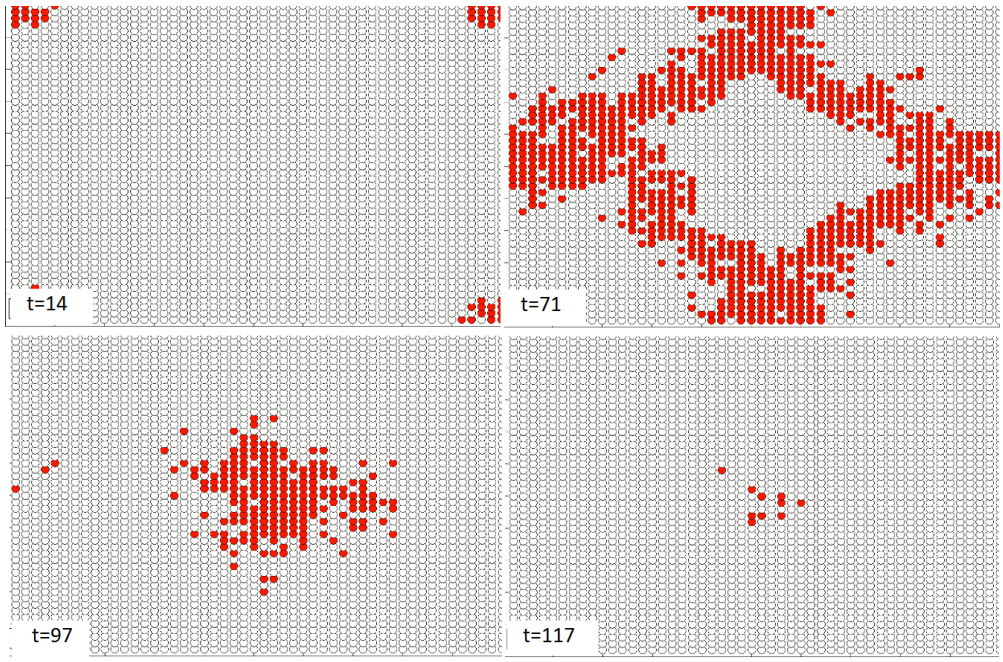

Figure 10: An epidemic on the 50×50 lattice with $\sigma = \frac{1}{20}$ at four different time points. Each lattice point represents a farm, with red patches corresponding to infected sites. The epidemic begins in the top-left corner, and spreads in a wave across the lattice.

is clearly an unrealistic assumption, and in reality we expect contacts between farms to be much less regular. We next consider a random network, constructed by placing farms at random points in a two dimensional box and then placing connections to a certain number of nearest-neighbour farms. A typical example is shown in figure 11 and a comparison of the full internal dynamics with this farm level model for a feasible number of farms is given in figure 12. Once again the farm level model overestimates survival probabilities, and so we should interpret its findings as an estimate of qualitative behaviour, rather than exact numerical predictions. To give an idea of the relative speeds of the two models, 1,000 repetitions of calculating the survival curve for 100 farms of size $N = 100$, using $\sigma = \frac{1}{20}$ took 6,077 seconds, whereas using the inter-farm model this only took 15 seconds. For 5,000 farms of the same size and the same value of σ, the inter-farm model took 1328 seconds, and for 10,000 farms 7352 seconds (all ran on a quad core 2.40Ghz laptop, using MATLAB and a maximum time for each survival curve of 10 years).

Despite the similarities with the lattice model in that there are only nearest neighbour interactions between farms, the survival curves in figure 13 indicate that this additional complexity can lead to persistence given a large enough network of farms. Although, as we have mentioned, the farm level model overestimates survival probabilities compared to direct simulation, it seems reasonable to infer that persistence really is possible for networks with $5,000$ or more farms.

Figure 14 illustrates the effect on persistence behaviour of varying the σ parameter. This relationship is more complicated than that between persistence and the number of farms. For low levels of contact we observe no persistence, whilst for large values of σ there is strong persistence as the disease can move with ease around the network. The behaviour is harder to interpret for intermediate values of σ. In this case it is possible that there is a trade-off between the ability to spread quickly across the network and the need to preserve susceptible farms to allow for future infection.

From figure 15, we see that introducing a more realistic spatial structure reduces the critical community size when compared to a single farm with homogeneous mixing. For $\sigma = 1/1000$, we observe no persistence past 1,000 days. As σ increases, the spatial structure plays a larger role in the dynamics and a much lower

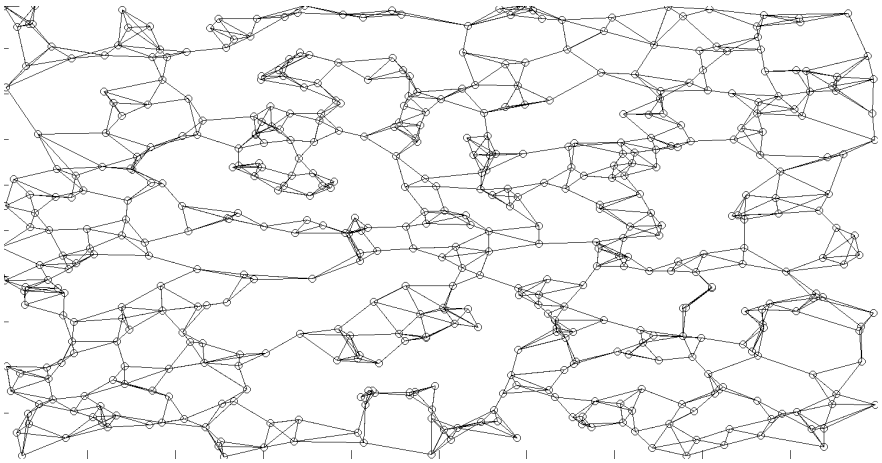

Figure 11: An example of the random network model, with 500 farms placed at random and connected to their five nearest neighbours.

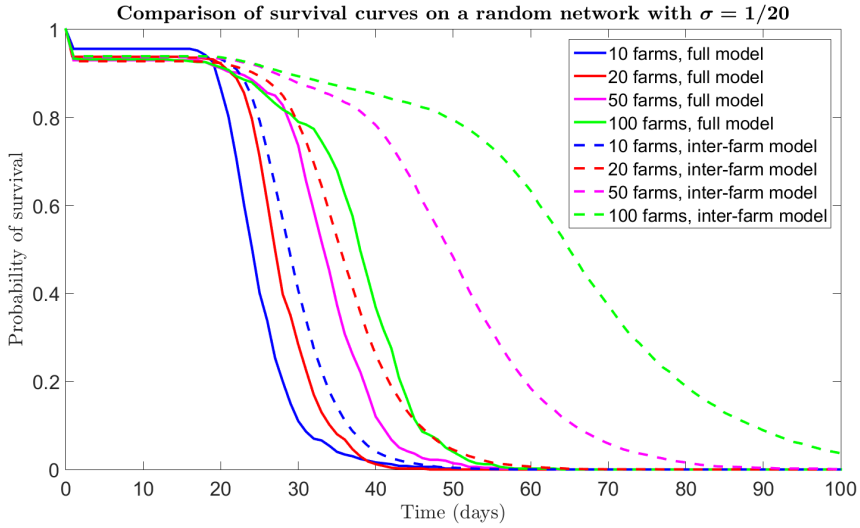

Figure 12: Comparison of survival curves for the full internal dynamics and farm level model for the random network with different numbers of farms, each of size $N = 100$. Connections are made between each farm and its five nearest neighbours.

total population size is needed to reach the same probability of persistence as in the homogeneous mixing case in figure 4. It is clear that persistence is highly dependent on the value of σ, however for all but the case of $\sigma = 1/1000$ we observe some persistence when the network has at least 5,000 farms of size 100 (total population of 500,000 or greater).

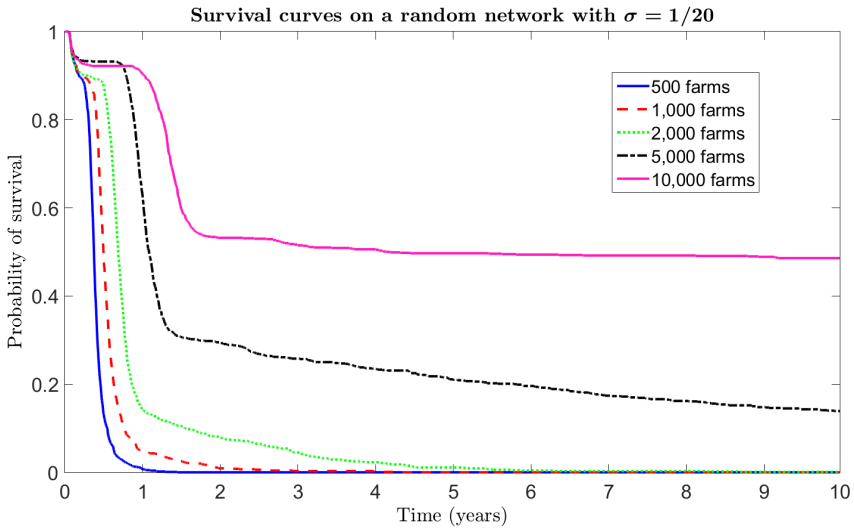

Figure 13: Survival curves for the random network model with $\sigma = 1/20$ and varying numbers of farms, each of size $N = 100$, using the inter-farm model. Each farm is connected to its five nearest neighbours.

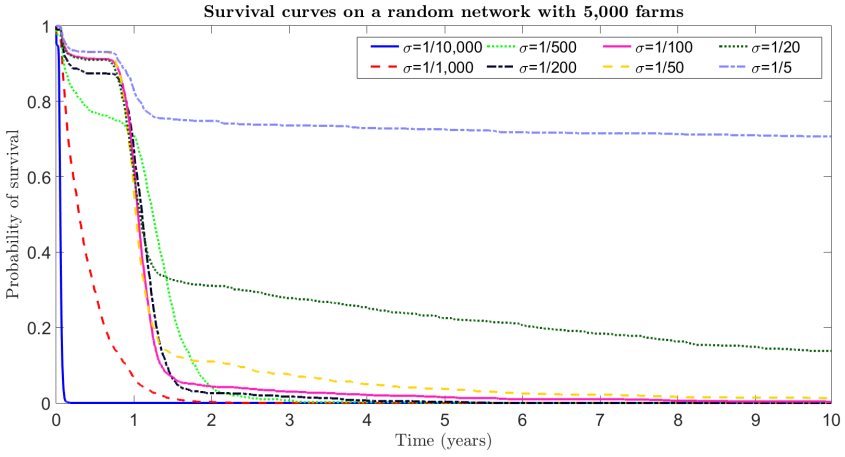

Figure 14: Survival curves for the random network model with 5,000 farms of size $N = 100$ and varying connection parameter σ using the inter-farm model. Connections are made between each farm and its five nearest neighbours.

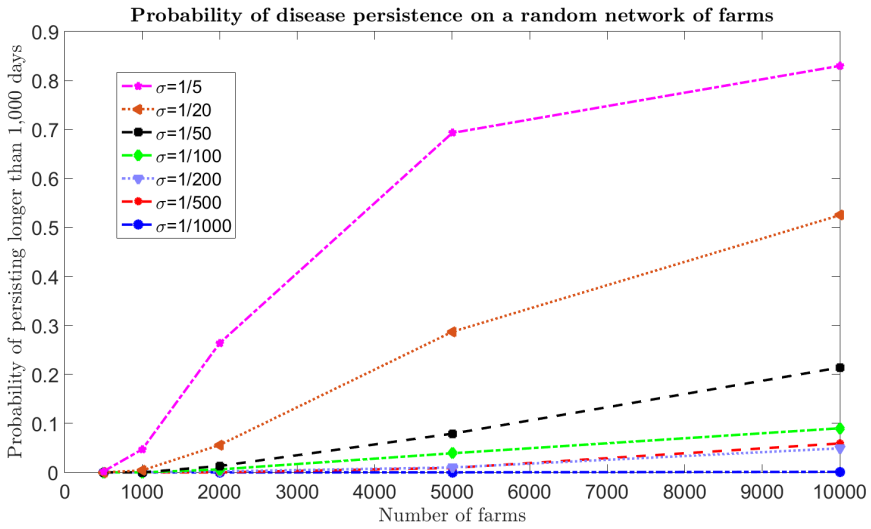

Figure 15: Probability of survival past 1,000 days as a function of the number of farms in the random network (each farm is of size $N = 100$) and for a number of different values of the σ parameter using the inter-farm model.

4 Conclusions and further work

For isolated populations of cattle we have demonstrated that there is a clear relationship between population size and persistence behaviour of FMD. The probability that a single import of infection will lead to FMD becoming endemic in the population increases with population size, and to estimate the critical community size we should choose a minimum persistence probability and then take the corresponding population size as our estimate. Under the assumption of homogeneous mixing an import of disease can lead to long-term persistence with probability 0.1 in populations of size 1.5 million and probability 0.5 in populations of size 15 million. It therefore seems reasonable to suggest that an estimate for the critical community size for FMD would fall between these bounds. In particular, the probability of persistence when the disease is confined to a single realistically sized farm is close to zero, and so we can conclude that FMD can only become endemic when it is allowed to spread between different herds of cattle.

For homogeneous ensembles of interacting farms, we have shown that the level of inter-farm contact has important implications for persistence. The results visualised in figure 5 indicate that FMD can not persist for very low levels of contact, but persists with increasing probability once the contact rate increases past a certain threshold. This supports the assertion that the structure of the population and level of interaction between herds is, along with population size, an important factor in the persistence behaviour of FMD. Solving the master equations for the system showed that the average number of infected individuals per farm oscillates towards its equilibrium value, and it would be interesting in further work to study the causes of these oscillations and see if they can be related to demographic turnover and waves of infection.

Simulations of epidemics on networks demonstrate that persistence behaviour is dependent on both population size and spatial structure. On a regular lattice of farms, persistence was never observed in our simulations, whereas on random arrangements of farms the persistence shows dependence on both contact rate and number of farms. For the random network, we find that for very low values of contact between farms persistence is never observed. For higher levels of between-farm interaction, the probability of persistence in a population of a given size increases, with larger numbers of farms resulting in higher persistence probabilities. In particular, FMD is able to persist in smaller populations when they are divided into farms, as opposed to when such populations are allowed to mix freely. High levels of contact between farms correspond more closely to the homogeneous freely mixing state, and it therefore seems reasonable to suggest that there may be a limit on how much increasing the contact level can increase persistence probabilities. On the other hand, the network structure still places limits on the range of interactions which can take place, even for very high contact levels, and so most network models will not specifically tend towards the homogeneous model.

Further work on this topic could incorporate data on the locations and sizes of cattle populations in the UK in order to develop more realistic network models. We have already mentioned that there are between nine and ten million cattle in the UK, and that the average farm is home to between 100 and 200 cows [3, 5]. This leaves us with somewhere between $50,000$ and $100,000$ farms in the UK, which, depending on the structure of connections between farms and rate of contact between them, could leave the country at high risk of endemic FMD. Work combining the methodology we have used here with more realistic network structures would help to clarify the magnitude of this risk. In particular, work towards quantifying levels of interaction between farms would be useful in motivating a choice of the contact parameter σ.

It would be interesting to compare results from a network based on cattle farms in the UK to one based on a country such as Turkey, where FMD is endemic and occurs seasonally [19]. The seasonal nature of FMD in many countries suggests that models incorporating seasonality would add greater realism to the results covered here. One limitation of the work presented here is that we have focused solely on interactions within a single host species. Global prevalence of FMD displays different spatial patterns for different species, and certain strains of FMD have adapted to specific mammal species [19]. This provides a strong motivation for extending our work to incorporate multiple host species and virus strains. A further extension would be to incorporate a latent period and multiple stages of infection to create a more realistic life cycle of the infection.

Acknowledgements

We would like to thank the Pirbright Institute, along with the EPSRC, MRC and the Mathematics for Real-World Systems CDT for providing funding.

References

[1] Livestock disease. *POST PN 392* (October 2006).

[2] Anderson, R. M., May, R. M., Joysey, K., Mollison, D., Conway, G. R., Cartwell, R., Thompson, H. V., and Dixon, B. The invasion, persistence and spread of infectious diseases within animal and plant communities [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 314*, 1167 (1986), 533–570.

[3] Brooks-Pollock, E., and Keeling, M. Herd size and bovine tuberculosis persistence in cattle farms in great britain. *Preventive veterinary medicine 92*, 4 (2009), 360–365.

[4] Charleston, B., Bankowski, B. M., Gubbins, S., Chase-Topping, M. E., Schley, D., Howey, R., Barnett, P. V., Gibson, D., Juleff, N. D., and Woolhouse, M. E. J. Relationship between clinical signs and transmission of an infectious disease and the implications for control. *Science 332*, 6030 (2011), 726–729.

[5] Eurostat, the statistical office of the European Union. Livestock and meat, 2015.

[6] Foreign Agricultural Service, United States Department of Agriculture. Livestock and poultry: World markets and trade, April 2015.

[7] Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics 115*, 4 (2001), 1716–1733.

[8] Grimmett, G., and Stirzaker, D. *Probability and Random Processes*. Oxford University Press, 2001.

[9] Gubbins, S. Personal communication (May 2015).

[10] Keeling, M. J. Models of foot-and-mouth disease. *Proceedings of the Royal Society of London B: Biological Sciences 272*, 1569 (2005), 1195–1202.

[11] Keeling, M. J., and Grenfell, B. T. Disease extinction and community size: Modelling the persistence of measles. *Science 275*, 5296 (1997), 65–67.

[12] Keeling, M. J., and Grenfell, B. T. Individual-based perspectives on R_0. *Journal of Theoretical Biology 203*, 1 (2000), 51–61.

[13] Keeling, M. J., and Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2007.

[14] Keeling, M. J., and Ross, J. V. On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface 5*, 19 (2008), 171–181.

[15] Kermack, W. O., and McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 115*, 772 (1927), 700–721.

[16] Kitching, P., Hammond, J., Jeggo, M., Charleston, B., Paton, D., Rodriguez, L., and Heckert, R. Global fmd control—is it an option? *Vaccine 25*, 30 (2007), 5660–5664.

[17] Kitching, R. Foot-and-mouth disease: current world situation. *Vaccine 17*, 13 (1999), 1772–1774.

[18] Mahy, B. W. Introduction and history of foot-and-mouth disease virus. In *Foot-and-Mouth Disease Virus*. Springer, 2005, pp. 1–8.

[19] Sumption, K., Rweyemamu, M., and Wint, W. Incidence and distribution of foot-and-mouth disease in asia, africa and south america; combining expert opinion, official disease information and livestock populations to assist risk assessment. *Transboundary and Emerging Diseases 55* (2008), 5–13.

Appendix A Final size of local epidemics

In the inter-farm model of section 3.1 we assumed that once the infection reaches a farm, all cattle on the farm become infected and have passed into the recovered class by the end of the epidemic. To justify this assumption, we refer to the plot in figure A1. This shows the probability distribution for the number of susceptibles remaining at the end of an epidemic on a given farm, given that the infection is able to establish itself. It is clear from this plot that the remaining number of susceptibles following an epidemic on a farm is close to zero with high probability, justifying our assumption that following an epidemic, all members of a herd will be in the recovered class.

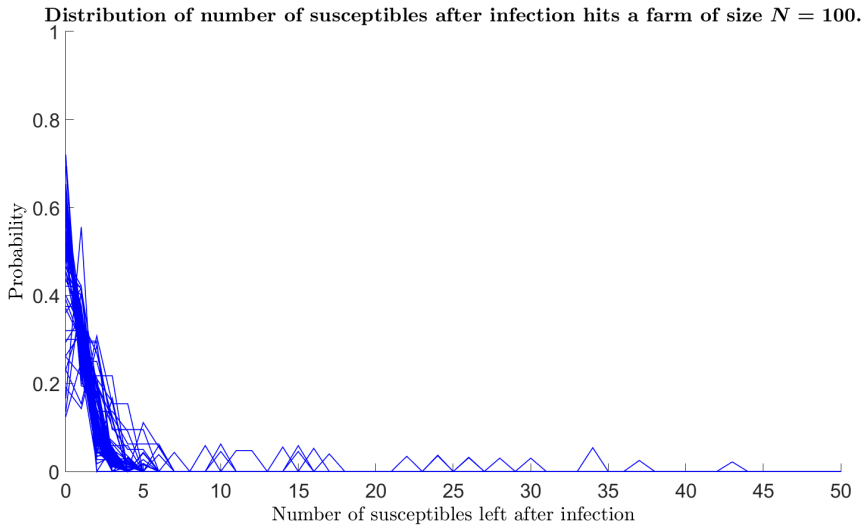

Figure A1: Distribution of the number of susceptibles left in a farm of size $N = 100$ after infection, given that the infection does not immediately die out. Distributions are plotted for each initial number of susceptibles, ranging from 1 to 99.

Appendix B Improvement to inter-farm model

In this appendix we describe an improvement to the inter-farm model introduced in section 3.1 which matches up better with the full network model without any significant effect on computational speed. The epidemic duration t_{end} is chosen in the same way as before, and again the formula from equation (9) is used to find the force of infection λ_{ij} on neighbouring farms. The epidemic duration and force of infection are used as before to calculate the probability that a neighbouring farm is infected, but we use a different procedure to find the time at which this infection event occurs. In the original model we solved the master equations to find I_t, the expected number of infected individuals present on the farm on day t of the epidemic, and then normalised this time evolution to get a probability distribution for infection time T. In fact, a better choice would be to make the probability of infection of farm j by farm i on day T of the epidemic on farm i proportional to

$$\exp\left(\frac{-\sigma_{ij}\beta S_j}{N}\sum_{t=0}^{T-1} I_t\right)\left(1-\exp\left(\frac{-\sigma_{ij}\beta S_j I_T}{N}\right)\right). \tag{10}$$

Here the first term in the product corresponds to the probability that infection has not taken place prior to day T, whilst the second term is the probability of infection on day T itself. A comparison with the full network model for the same parameter choices as in figure 12 can be seen in figure B1. This choice of infection time results in a better agreement between the two models. By neglecting to include a term corresponding to the fact that infection has not yet taken place, the model in section 3.1 overestimates infection probabilities, resulting in a longer epidemic.

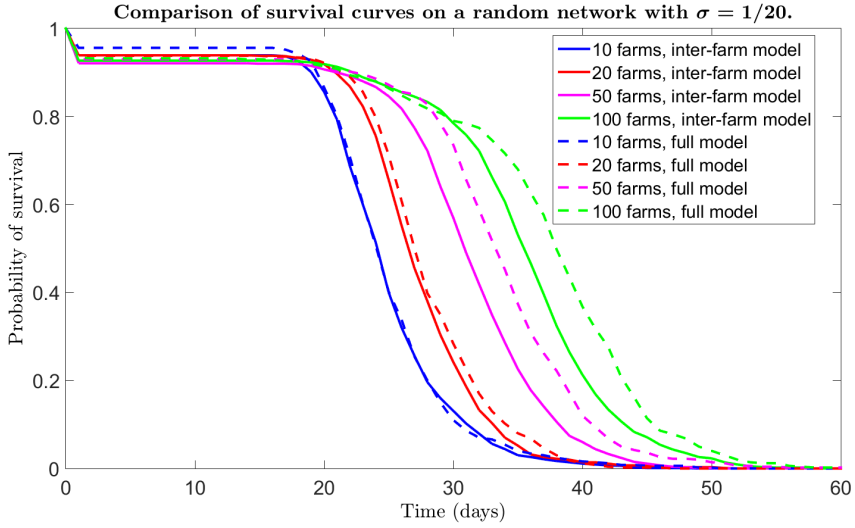

Figure B1: Comparison of survival curves for the full internal dynamics and the improved inter-farm model for the random network with different numbers of farms, each of size $N = 100$. Connections are made between each farm and its five nearest neighbours. The alterations to the inter-farm model has removed much of the lag between the survival curves from the two models.

Figure B2 shows the survival curves for a larger network of farms, where each farm is connected to its five nearest neighbours and with the same parameters used in figure 13. We see that there is still long term persistence, although larger networks of farms are required and there is a higher probability of early extinction than in the original inter-farm model.

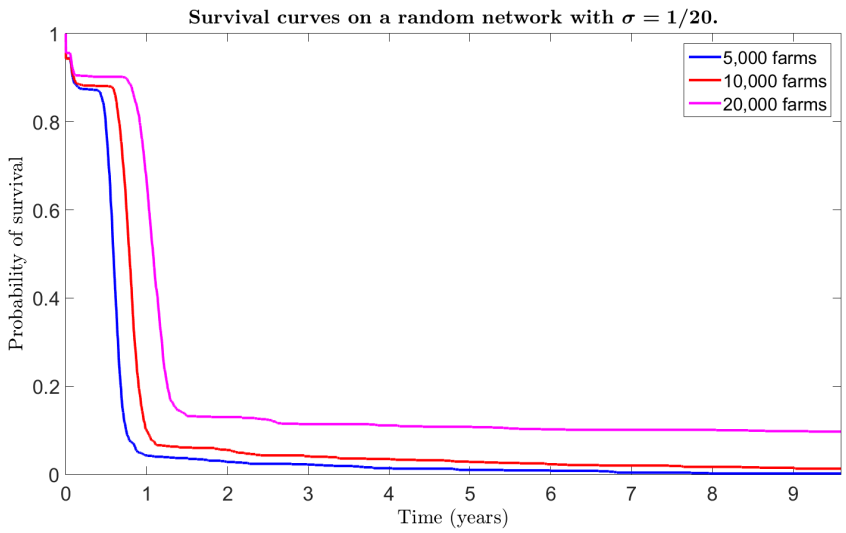

Figure B2: Survival curves on the random network using the improved inter-farm model with $\sigma = 1/20$ and varying numbers of farms, each farm of size $N = 100$. Each farm is connected to its five nearest neighbours.