



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Michael J. Tildesley, Thomas A. House, Mark C. Bruhn, Ross J. Curry, Maggie O'Neild, Justine L. E. Allpress, Gary Smith and Matt J. Keeling

Article Title: Impact of spatial clustering on disease transmission and optimal control

Year of publication: 2010

Link to published article: <http://dx.doi.org/10.1073/pnas.090904710>

Publisher statement: Copyright ©2011 by the National Academy of Sciences

# The impact of spatial clustering on disease transmission and optimal control

Michael J. Tildesley<sup>1,2</sup>, Thomas A. House<sup>3</sup>, Mark C. Bruhn<sup>4</sup>,  
Ross J. Curry<sup>4</sup>, Maggie O’Neil<sup>4</sup>, Justine L.E. Allpress<sup>4</sup>, Gary  
Smith<sup>5</sup> and Matt J. Keeling<sup>3</sup>

1. Centre for Immunity, Infection and Evolution, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, United Kingdom.
2. US National Institute of Health, Fogarty International Center, Bethesda, MD 20892 USA
3. Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK
4. RTI International, Research Triangle Park, NC, USA
5. Section of Epidemiology and Public Health, School of Veterinary Medicine, University of Pennsylvania, Kennett Square, PA 19348 USA

**Spatial heterogeneities and spatial separation of hosts are often seen as key factors when developing accurate predictive models of the spread of pathogens. The question we address in this paper is how coarse the resolution of the spatial data can be for a model to be a useful tool for informing control policies. We examine this problem using the specific case of foot-and-mouth disease spreading between farms using the formulation developed during the 2001 epidemic in the UK. We show that if our model is carefully parameterised to match epidemic behaviour, then using aggregate county-scale data from the USA is sufficient to closely determine optimal control measures (specifically ring culling). This result also holds when the approach is extended to theoretical distributions of farms where the spatial clustering can be manipulated to extremes. We have therefore shown that while spatial structure can be critically important in allowing us to predict the emergent population-scale behaviour from a knowledge of the individual-level dynamics, for this specific applied question such structure is mostly subsumed in the parameterisation allowing us to make policy predictions in the absence of high-quality spatial information. We believe that this approach will be of considerable benefit across a range of disciplines where data is only available at intermediate spatial scales.**

## **Introduction**

The spatial distribution of organisms is viewed as critically important for determining population dynamics. Numerous examples from the epidemiological and ecological literature have shown that spatial structure has a profound impact on how population-level dynamics emerge from individual-level behaviour (1-4). For infectious diseases in particular, where transmission generally occurs over relatively short distances, spatial structure (and in particular the spatial distribution of sessile hosts) plays three roles: hosts that are far from sources of infection are at very little risk; local transmission and depletion of susceptible hosts can dramatically reduce

the speed of epidemic growth; and local control measures can be applied using spatial proximity as a method of targeting at risk hosts. These three elements are present for any spatial distribution of hosts, but are generally amplified by clustering. The impact of spatial structure on the spread of infectious disease has been examined for humans (5), wildlife (6-7) and livestock (8-9), but the ability to make useful quantitative predictions relies on the availability of good quality spatial and epidemic data. In recent years considerable research has focused on the spread of livestock infections due to the extreme vulnerability of the livestock industry, the potential economic costs, the variety of strategies that can be used as control measures and the costs associated with such measures.

The UK 2001 epidemic of foot-and-mouth disease (FMD) provides a prime example of what can be achieved when comprehensive spatial models, detailed host data and detailed case data are brought together. This approach provided important insights and guidance during the 2001 epidemic (8-9) and has been used retrospectively to investigate a range of alternative control strategies (10-12). Whilst the location of livestock holdings in the UK is known, the same is not true for many other countries. In particular, in the USA, although the United States Department of Agriculture holds an agricultural census every five years, most of the census data that resides in the public domain is aggregated to county-scale preserving the anonymity of farmers but losing valuable spatial information. From the 2007 census, we find that the United States' livestock industry is dominated by cattle, with 936,669 premises recorded as having cattle, 83,134 as having sheep, 75,442 as having pigs (or hogs) and 91,462 as having goats.

In the event of an outbreak of FMD (or other livestock disease) in the USA, it is likely that predictions from mathematical models would be an integral part of policy making and would help advise regarding optimal control strategies to limit the size and duration of the outbreak.

The likelihood of a particular farm being infected with FMD is based upon many factors, including the type and size of the holding in question, but the proximity to other infectious farms has been consistently demonstrated to be the main contributing factor (13-15). It therefore seems vital to have complete information regarding the location and size of all farms in the region of an outbreak.

Given that information is only available at the county-scale in the USA, we consider the impact of making the naive assumption that farms are distributed randomly within each county. This assumption still allows us to implement spatial models, but loses any clustering that is present in the distribution of farms. In particular, we consider how a model using a random distribution of farms would be parameterised to match the temporal profile of an observed epidemic and show that this approach has strong applied benefits even in the absence of fine-scale positional data.

### The model

The model used throughout this paper is an adaptation of the model developed by Keeling *et al* (2001) (8) during the 2001 FMD epidemic, and discussed in detail elsewhere (16). The rate at which an infectious farm  $i$  infects a susceptible farm  $j$  is given by:

$$\text{Rate}_{ij} = \left( \sum_{s \in \text{species}} [N_{s,j}]^{p_s} S_s \right) \times \left( \sum_{s \in \text{species}} [N_{s,i}]^{q_s} T_s \right) \times K(d_{ij}) \quad (1)$$

$N_{s,i}$  is the number of livestock species  $s$  recorded as being on farm  $i$ ,  $S_s$  and  $T_s$  measure the species-specific susceptibility and transmissibility,  $d_{ij}$  is the distance between farms  $i$  and  $j$  and  $K$  is the distance-dependent transmission kernel, estimated from contact tracing (8).  $p_s$  and  $q_s$  are power law parameters accounting for a non-linear increase in susceptibility and transmissibility for species  $s$  as animal numbers on a farm increase (see Supplementary Material). This form of spatial transmission model has been shown to provide an accurate and robust description

of the UK 2001 outbreak of foot-and-mouth, capturing national, regional and individual-level patterns of infection (16-18). In keeping with observations from the UK 2001 epidemic, we assume that a farm acts as a single infectious unit, once infected it enters the latent period (which lasts 5 days) before becoming infectious where it remains until its livestock are culled. We note that, whilst this assumption may be valid for regions of the USA with similar farming practices as the UK, very different behaviour may occur in very large cattle farms (19). All parameters for the UK county models using the true farm locations take the same values as that obtained for the UK 2001 epidemic as discussed in detail elsewhere (16).

During the UK foot-and-mouth outbreaks in 2001 and 2007, in addition to routine culling of infected premises (IPs), all “premises where animals have been in direct contact with infected animals or have, in any way, become exposed to infection” were defined by veterinary judgement as dangerous contacts (DCs) and were pre-emptively culled in an effort to control disease. In our model, DCs are identified based on their risk of infection from the infecting source, biased toward identifying farms that have actually been infected (11), assuming that veterinary judgement benefits from a range of detailed local knowledge. Here we assume that once identified all livestock on an infected premises are culled within 24 hours and all associated pre-emptive culling is undertaken within 48 hours; this is somewhat optimistic but was one of the principle aims during the 2001 epidemic.

Retrospective analysis of the 2001 epidemic has determined that a policy of IP and DC culling alone would have resulted in a much larger epidemic than actually occurred, implying that other culling strategies, including culling of contiguous premises and farms within 3km of IPs, aided in disease control (12). With this in mind, we investigate the effectiveness of ring culling in addition to IP and DC culling. When an IP is reported, all farms within a particular radius of that IP will be targeted for culling. The radius of the ring is allowed to vary between simulations and we seek the radius which minimises the “Epidemic Impact”, defined as the total number of

farms with livestock culled (either as IPs, DCs or ring culled farms). The optimal ring size is clearly a trade-off between too little culling, in which case the epidemic is not controlled, and too much culling, in which case an excessive number of farms lose their livestock.

### **Methodological Approach**

Our goal is to test the accuracy of predictions made when detailed spatial data are not available, and the only recourse is to randomly scatter farms across the landscape. As a first step, and focusing on regions where the location of farms (and their livestock composition) is known, we perform multiple simulations to determine the range of epidemics that can be expected. Using the same spatial data, we then conduct further simulations to determine the ring cull radius ( $R_T$ ) that minimises the total number of farms losing livestock to either infection or control.

In the second step, we distribute the same farms randomly within the given region to simulate the effects of not having the precise spatial locations but knowing the heterogeneities in the number of livestock. At this stage, we could simulate epidemics on the random spatial data set using the original UK parameters. However, it is naive to assume that the UK 2001 parameters could be used for a future epidemic in a different farm demography. We therefore choose instead to mimick what would happen during a real epidemic and estimate the parameters that allow us to accurately predict the epidemic behaviour – in this case matching the attack rates obtained using the randomly distributed farms to the attack rates obtained using the true spatial location of farms. This parameterisation is achieved by fitting a 2 parameter description of the transmission kernel (determining kernel width and height,  $K_w$  and  $K_h$  respectively), and uses the same approach as outlined in Tildesley *et al* (2008) (16) – to provide a best fit, parameters are found that minimise the average difference on a daily basis between ‘simulated epidemics’ from the random location data to the ‘observed epidemic’ simulated on the true spatial data for the cumulative number of farms reported and culled as well as the cumulative number of

cattle and sheep on such farms (see Supplementary Material). Using this random distribution of farms, but with a refined parameterisation, we determine the optimal ring cull radius for the random-location reparameterized model ( $R_{RR}$ ). Comparison between  $R_T$  and  $R_{RR}$ , and the Epidemic Impact under both culling regimes provides important insights into the necessity of detailed spatial data for informing policy. In all simulations, a single farm is randomly selected and seeded with infection. The onset of ring culling occurs 48 hours after the first case is reported, with subsequent ring culling occurring 48 hours after the reporting of the relevant IP. In essence our approach is to use the first set of simulations (using the true spatial locations) as a surrogate for real epidemic data; this has the advantage that a variety of control options can be tested and compared using both true and randomised spatial data. We now examine how this approach can be applied to county-level data from the UK and USA, as well as hypothetical data with arbitrary clustering.

### **County-Level Data**

We begin by examining spatial location data for the county of Cumbria in the UK. Cumbria has a high density of large cattle farms, and as a consequence was one of the worst affected areas in the UK 2001 outbreak. Comparing the true and randomised locations (figure 1a) we see significant local clustering as captured by the average density of farms around each farm. We note that the farm-centred density for randomised spatial locations is not constant (as theoretically expected) due to the finite scale of the region. For counties in the United States, although the true locations of farms are generally unknown, farms were geo-located for a handful of counties representing a diverse mixture of livestock operations within the USA. This spatial and heterogeneous farm information was generated using local and national agricultural, regulatory, farm subsidy, business, and tax assessor databases containing addresses or coordinates of farms. The locations of the farms were then georeferenced, visually checked, and modified if necessary



with aerial imagery, Google Streetview, Google & Mapquest maps, and other ancillary data. In a few cases where available data sources were not able to locate the number of farms that were indicated by the agricultural census, probable farm locations were based on land-parcels that were predominantly pasture, had farm buildings, and had access to a drinking water source. Figure 1b shows example data from Lancaster County in Pennsylvania, where we observe far stronger local clustering than in Cumbria despite a comparable density of farms. This pattern of local clustering of farms is consistent across other counties in the UK and USA (figure 1c), although the strength of clustering and the overall density of farms differs considerably. We show later that these farm-centred density plots can be captured as the sum of exponentials.

For the counties in the USA considered here, precise parameter values are unknown. Should UK parameters be applied to the USA, the lower overall farm density means that epidemics do not generally take off. Therefore, in order to provide epidemics of a sufficient size and duration for comparison with the random location model, we scale the UK values of  $K_h$  and  $K_w$  for the clustered distribution of farms ('true data') according to the relative densities of these counties in comparison to the UK, thus preserving the overall number of contacts between farms. It is not intended that the Epidemic Impacts given for the US counties should be in any indicative of the actual Epidemic Impacts in the event of an epidemic – rather it is the level of agreement between the results for the 'true data' and 'random data' which is of importance for targeting of control in the event of future epidemics.

Epidemics simulated on the true Cumbrian data predict an average Epidemic Impact of 2505 (31% of the population). In contrast, if the randomised data is used (with the same parameters) the average Epidemic Impact drops to 1765 as the lack of spatial clustering means that it is more difficult for the epidemic to spread – in the clustered scenario the minimum distance between farms is less and so the epidemic spreads more easily. However, we can re-parameterise such that the random model gives comparable Epidemic Impacts to the fully clustered model. Table

Region	Epidemic Impact (T)	Epidemic Impact (R)	Epidemic Impact (RR)
Cumbria	2505 (2121 – 2976)	1765 (101 – 2287)	2429 (1432 – 3054)
Devon	519 (27 – 1998)	190 (31 – 561)	545 (34 – 981)
Clwyd	679 (475 – 1131)	388 (52 – 798)	641 (148 – 1355)
Aberdeenshire	80 (20 – 263)	25 (16 – 41)	76 (21 – 121)
Lancaster, PA	1284 (954 – 1634)	75 (20 – 216)	1197 (576 – 1545)
Cuming, NE	454 (443 – 461)	441 (423 – 453)	453 (437 – 463)
Wright/Humboldt, IA	134 (78 – 171)	121 (60 – 152)	133 (75 – 165)
Franklin, TX	244 (20 – 318)	118 (15 – 182)	220 (16 – 304)

Table 1: Mean epidemic impact for epidemics seeded within a given county for the true clustered location data (T) and for an equivalent data set with random farm locations (R) within each county. Epidemic impact is also given for random farm locations but with the model re-parameterised to provide a best fit to epidemics simulated on the true location data (RR). Values in brackets give the 95% prediction intervals; all results are from 10000 stochastic simulations.

1 gives a full list of Epidemic Impacts for all nine counties examined in figure 1 (note Wright and Humboldt counties border one another so the combined epidemic impact is given here); in all cases re-parameterisation for the random locations generates epidemic impacts that are in close agreement with predictions from the full spatial model – this is unsurprising as the results of the full spatial model are used for parameterisation.

Whilst obtaining a good fit to the epidemic profile and hence being able to predict the likely extent of an epidemic are potentially informative, by far the most useful application of mathematical models is to inform the optimal policy for controlling an epidemic. As such accurate mathematical models can be used to experimentally test a variety of control strategies to assess

which is optimal – although it is the role of policy makers to decide which quantity needs to be optimised. Here we consider ring culling (in addition to IP and DC culling) and determine what size of ring minimises the average Epidemic Impact. The simplest way to determine this value is through multiple simulations. We first adopt this approach using the true locations of farms in Cumbria and Lancaster counties (figure 2a and b, blue lines) which predict clear minima at the optimal radii ( $R_T$ ) of approximately 3.6km (2.24 miles) in both cases. A similar approach can be used to optimise other forms of control, such as vaccination under logistical constraints (11) or localised (contiguous) culling (12). However, the key question is whether we can perform the same meaningful calculation in the absence of detailed spatial location data. We therefore perform the same numerical experiment using the reparameterised model and the random farm locations (red line, figures 2a and b), and observe that although using random spatial locations leads us to over-estimate the effect of control measures, the optimal ring cull radius ( $R_{RR}$ ) is close to the result from fully clustered spatial simulations. In fact this level of agreement is so close that using the ring cull radius  $R_{RR}$  in the fully clustered simulations only has a marginal effect on the average Epidemic Impact. For Lancaster county using a ring cull radius of  $R_{RR} = 3.8\text{km}$  leads to an average epidemic impact of 678 farms, an increase of just 3 farms from the true optimal value at  $R_T = 3.6\text{km}$ . Table 2 shows how this process holds across all nine counties discussed so far.

Whilst the optimal ring cull radius is predicted to vary between counties, the increase in Epidemic Impact from the use of predictions made in the absence of spatial information is limited to no more than 2 or 3 farms in each case. This implies that, for the counties considered in this analysis, precise knowledge of farm location is not required to advise regarding optimal control strategies. It should be stressed that the reparameterisation of the model is a crucial step in this process – should this be omitted from the procedure, models are unable to predict the true optimal radii with any accuracy. This effect is particularly noticeable in Lancaster county

Region	$R_T$	$R_{RR}$	Epidemic Impact Difference
Cumbria	3.6 (3.5-3.8)	3.8 (3.6-4.0)	3 (0-11)
Devon	2.8 (2.7-3.0)	2.8 (2.7-3.1)	0 (0-3)
Clwyd	3.6 (3.5-3.7)	3.2 (3.1-3.4)	3 (1-7)
Aberdeenshire	2.4 (2.3-2.7)	2.0 (1.9-2.2)	2 (1-3)
Lancaster, PA	3.6 (3.5-3.7)	3.8 (3.7-4.0)	2 (0-6)
Cuming, NE	0 (0-0)	0 (0-0)	0 (0-0)
Wright/Humboldt, IA	0 (0-0)	0 (0-0)	0 (0-0)
Franklin, TX	5.5 (5.4-5.6)	6.0 (5.8-6.1)	1 (0-3)

Table 2: The optimal ring cull radius in kilometres which minimises the epidemic impact for both the true clustered location data ( $R_T$ ) and the reparameterised random location data ( $R_{RR}$ ). Also shown is the increase in epidemic impact if the optimal ring cull radius for the reparameterised random data were implemented on the true location data (as opposed to the optimal ring cull radius for the true data). 95% confidence intervals for optimal ring cull radii and epidemic impact difference are given in brackets. All results are for 10000 stochastic simulations.

– in the absence of reparameterisation, the random-location model predicts very low epidemic impacts and the optimal strategy is to employ no ring culling. After reparameterisation, as discussed above, it is optimal to ring cull at  $\approx 3.8\text{km}$ .

### Spatial Clustering

Whilst the nine counties considered so far imply that our results are general, to fully test this approach we generate alternative farm distributions based around a given average farm-centred density distribution,  $\overline{D}(r)$ . We define  $\overline{D}(r)$  as the number of farms per unit area at a distance  $r$  from an index farm, averaged over all possible index farms in the population. For highly clustered distributions, we expect  $D(r)$  to decrease nonlinearly with  $r$  as observed in figure (1). In practice we find that  $\overline{D}(r)$  can be fit by a sum of exponentials:

$$\overline{D}(r) = S_{\text{inf}} + (S_0 - S_{\text{inf}}) \left( \sum_i A_i e^{-B_i r} \right) \quad (2)$$

where  $S_{\text{inf}}$  defines the long-distance asymptotic density,  $S_0$  defines the average local density around a farm, and we insist that  $\sum_i A_i = 1$ . We find that a sum of three exponentials (and hence 7 parameters) is sufficient to fully capture the observed density distributions from all nine counties (see Supplementary Material).

We now generate theoretical spatial distributions of farms to test the validity of our approach over a wider range of parameters. In particular, we distribute  $N = 1000$  farms in a  $50 \times 50$  km area, according to a given density distribution. For each simulation, one farm is randomly seeded with infection. To simplify our analysis we ignore heterogeneity in farm size and composition, and use a reduced formulation for the density distribution; in particular, transmission of infection between infectious and susceptible farms is now a highly simplified version of equation 1, depending only on their separation ( $\text{Rate}_{ij} = K(d_{ij})$ ), while the density distribution is

defined in terms of a single exponential decay:

$$D(r) = S_{\text{inf}} + (S_0 - S_{\text{inf}}) e^{-Br} \quad (3)$$

We fix the population size and area that we are going to study, which reduces our free parameters to the ratio  $S_0 : S_{\text{inf}}$ , and the exponent  $B$ . Figure 2c and 2d, show an example of applying our methodology to these theoretical spatial distributions ( $S_0 : S_{\text{inf}} = 10, B = 0.4, N = 1000$  such that  $S_{\text{inf}} = 0.4$ ). This distribution is far more clustered than any of the previous real-world examples, and yet reparameterising and determining the optimal ring cull radius remains a valid approach. The estimated optimal radius  $R_{RR}$  is an over-estimate of the true value, but culling in the highly clustered distribution with a radius of  $R_{RR}$  only increases the epidemic impact by 3 farms from the true optimal value of 152 farms. Here we find that our methodology benefits from a general principle that it is usually better to over-target control (ie bias control more toward high risk hosts than is strictly optimal) than under-target, hence it is better to ring cull using a radius that is slightly larger than optimal compared to one that is smaller than optimal.

We can expand these theoretical spatial distributions to a range of clustering, and repeat our basic analysis (figure 3); we allow our two fundamental parameters ( $S_0 : S_{\text{inf}}$  and  $B$ ) to vary over a grid of values. For each clustered distribution, figure 3a shows the percentage reduction in the epidemic impact from the full clustered model (compared to IP and DC culling alone) by introducing ring culling at the optimal radius ( $R_{RR}$ ) predicted from a random distribution of farms. This illustrates the substantial applied benefits that can be accrued from using well-parameterised mathematical models, even when some of the finer scale spatial information is missing. Localised ring culling is seen to have the greatest percentage impact as the ratio  $S_0 : S_{\text{inf}}$  decreases and as  $B$  increases; that is a decrease in epidemic impact of approximately 90% is possible unless the spatial distribution is strongly clustered over relatively large spatial scales. However, it is important to understand whether better data would allow

us to improve on these results; figure 3b shows the further reduction in epidemic impact to be gained by culling at the true optimal radius ( $R_T$ ) rather than the radius ( $R_{RR}$ ) predicted from a random distribution. We observe that over the vast majority of spatial patterns considered, there is only a relatively small average improvement to be gained from using exact spatial knowledge.

For models such as these to be used prospectively in the event of an epidemic of FMD, the approach described above could not be adopted in the same way. The results so far all use the entire epidemic on the clustered location data to reparameterise a random location model, prior to the development of preferred ring cull strategies. In order for such a model to be used during an outbreak, as soon as an epidemic is reported, the early pattern of cases must be used to fit a model allowing an early investigation of intervention strategies. We therefore repeat the process described above, but this time use only the first two weeks of the epidemic from the true spatial model to fit the kernel parameters for the random location model. In the absence of ring culling, epidemics on the true spatial distribution of farms last for typically 100-120 days, so the first 14 days of the epidemic represents the early growth phase. The fitting process is identical to that described above: kernel parameters are determined that generate a best fit to the first two weeks of data; we find the optimal cull radius as suggested by these parameters and the random location model; and finally we determine the effects of culling at this radius in the true model which contains the clustered locations of farms. The results are summarised in figures 3c and d. If ring culling at  $R_{RR}$  were implemented within the model using the true clustered locations of farms, then the percentage reduction in the epidemic compared with IP and DC culling alone is still around 90% for most of parameter space with a reduction to around 65-70% if farms are clustered over large spatial scales (figure 3c), mirroring that seen when we fit to the entire epidemic. From figure 3d we see that there is a further reduction in epidemic impact when culling at the true optimal radius  $R_T$  but this improvement, though slightly higher than when we fit to

the entire epidemic, is still relatively small. This result highlights that, during the early stages of an outbreak, epidemic data could be used to fit a random location model which could then be used to predict preferred strategies for targeted interventions.

## **Discussion**

Spatial clustering of individuals plays a highly significant role in ecological dynamics (1-2; 20-21) and in the spread of infectious disease (22-24). Therefore, when precise demographic data is unavailable, alternate methods need to be adopted for mathematical modellers to predict the best control policy to combat an epidemic and therefore provide useful policy advice. In the context of livestock diseases, in the event of an animal movement ban, the location of individual livestock is fixed, with infection between farms only occurring via movement of people, machinery and other forms of local transmission. This epidemiological situation has been investigated in detail for the UK, Denmark, Australia and New Zealand (25-26) where detailed spatial location data is recorded and available, but not for the USA where only data aggregated at the county scale is released.

This paper investigates the effect of assuming that farms are randomly located within a county (or given area), ignoring the spatial clustering that is known to exist; it is also assumed throughout that numbers of cases are also only available at this aggregated scale. The first observation is that using the same basic model parameters (i.e. identical  $K_w$  and  $K_h$ ), but ignoring spatial clustering, leads to far smaller epidemics that underestimate the true level of optimal control. However, when the models using random spatial locations are re-parameterised to match epidemics (either from real or surrogate sources), they can still be used to provide meaningful predictions for control. In this paper we have focused on optimal control by localised ring culling, although alternative measures such as ring vaccination or stringent quarantine (27), could be investigated in the same way. The optimal ring size predicted from random spatial



locations (but a reparameterised model) is in general a slight over-estimate of the true optimal radius; using this over-estimate still leads to a substantial improvement in the predicted epidemic impact and is only slightly worse than using the true value optimised to account for spatial structure in the distribution of farms.

We therefore find that recognising that these epidemics take place in a spatial landscape is vital (measures such as ring culling cannot be applied in a non-spatial environment) but once a model has been parameterised to match the epidemic data, much of the spatial structure is subsumed into the re-parameterisation. As such, aggregate data and aggregate case reports can be used to derive predictions for control of epidemics with spatially localised transmission. Although we have focused on the spread of foot-and-mouth disease between livestock on farms, many other examples of localised transmission of infection occur which could be considered with a similar methodology, including plant and crop diseases (28), wildlife diseases (29), human infections (30-31) and other livestock diseases (32).

In the USA, where precise knowledge of farm location is unknown but aggregate farm statistics are available, this work provides an insight into the role that mathematical modelling can still play in informing disease control policy. However, several further points are worth considering. Firstly, this methodology relies on the availability of epidemic data, and can therefore only be used once an epidemic is in progress; it cannot be used to inform policy before an epidemic arises as case report data is vital if the model is to be re-parameterised to capture the effects of spatial structure. For detailed predictive models to be available before an epidemic occurs would either require a vast change to the way information on farms is handled and distributed, or would require ground-truthed synthetic maps based on topological and geographical features. As an extension to this work, alternate maps could be developed for regions with known farm locations such as the UK, using these known geographical features to eliminate regions such as mountains and lakes which are not suitable for farming. This would affect

the spatial clustering of farms and potentially the subsequent disease dynamics. Using the same approach as adopted in this paper, this work would provide strong support for the development of ground-truthed synthetic maps such as those for the five counties in the USA investigated in this paper. However, even when such spatial information is available, the different farming and management practices in different regions of the country make it difficult to utilise a universal model of transmission. Our approach can only be used reactively; therefore it is vitally important that case report data is available in real-time and parameterisation is performed using state-of-the-art methodology, such that timely policy advice can be given during the early stages of the epidemic when it will have the greatest effect. However, given the simplicity and parsimony of our approach it is likely that it can be used robustly in many situations, providing rapid policy advice that could minimise potential losses.

### **Acknowledgements**

This work was funded by the MIDAS (NIGMS, NIH), the Wellcome Trust, the SFC/DEFRA VTRI programme, the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. Support for MJT was also provided through a Wellcome Trust strategic award for the Centre for Immunology, Infection and Evolution, University of Edinburgh.

### **References**

1. Hassell, M.P & May, R.M. (1974) Aggregation of Predators and Insect Parasites and its Effect on Stability. *J. An. Ecol.* **43** 567-594.
2. Pacala, S.W., Hassell, M.P. & May, R.M. (1990) Host Parasitoid Associations in Patchy Environments. *Nature* **344** 150-153.

3. Kareiva, P. (1990) Population-Dynamics in Spatially Complex Environments *Phil. Trans. Roy. Soc. Lond. B* **330** 175-190.
4. Tilman, D. (1994) Competition and Biodiversity in Spatially Structured Habitats *Ecology* **75**, 2-16.
5. Ferguson, N.M.; Cummings, D.A.T.; Cauchemez, S, *et al* (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia *Nature* **437** 209-214.
6. Swinton, J., Harwood J., Grenfell B.T. & Gilligan C.A. (1998) Persistence thresholds for phocine distemper virus infection in harbour seal *Phoca vitulina* metapopulations. *J. An. Ecol.* **67** 54-68.
7. Gudelj, I. & White, K.A.J. (2004) Spatial heterogeneity, social structure and disease dynamics of animal populations. *Theoretical Population Biology* **66** (2), 139-149.
8. Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M.E., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. & Grenfell, B.T (2001) Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813-817.
9. Ferguson, N.M., Donnelly, C.A. & Anderson, R.M. (2001a) Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542-548.
10. Keeling, M.J., Woolhouse, M.E.J., May, R.M., Davies, G. & Grenfell, B.T. (2003) Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**, 136-142.
11. Tildesley, M.J., Savill, N.J., Shaw, D.J., Deardon, R., Brooks, S.P., Woolhouse, M.E., Grenfell, B.T. & Keeling, M.J. (2006) Optimal reactive vaccination strategies for a foot-and-mouth outbreak in Great Britain *Nature* **440**, 83-86.
12. Tildesley, M.J., Bessell, P.R., Keeling, M.J. & Woolhouse, M.E.J. (2009) The role of pre-emptive culling in the control of Foot-and-Mouth Disease. *Proc. Roy. Soc. B*, **276**, 3239-3248.
13. Hugh-Jones, M.E. (1972) Epidemiological studies on 1967-1968 Foot and Mouth Epidemic

- attack rates and cattle density. *Research in Veterinary Science* **13:5**, 411-417
14. Ferguson, N.M., Donnelly, C.A. & Anderson, R.M. (2001b) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155-1160.
  15. Taylor, N.M., Honhold, N., Paterson, A.D. & Mansley, L.M. (2004) Risk of foot-and-mouth disease associated with proximity in space and time to infected premises and the implications for control policy during the 2001 epidemic in Cumbria. *Veterinary Record* **154**, 617-626.
  16. Tildesley, M.J., Deardon, R., Savill, N.J., Bessell, P.R., Brooks, S.P., Woolhouse, M.E.J, Grenfell, B.T. & Matt J Keeling, M.J. (2008) Accuracy of models for the 2001 Foot-and-Mouth Epidemic. *Proc. Roy. Soc. B* **275** (1641), 1459-1468.
  17. Diggle, P.J. (2006) Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Statistical Methods in Medical Research* **25**, 325-336.
  18. Deardon, R., Brooks, S.P., Grenfell, B.T., Keeling, M.J., Tildesley, M.J., Savill, N.J., Shaw, D.J. & and Mark E. J. Woolhouse, M.E.J (2009) Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, in press.
  19. Bates, T.W., Thurmond, M.C. & Carpenter, T.E. (2001) Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. *American Journal of Veterinary Research* **62**, 1121-1129.
  20. Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M., Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* **7**, 601-613.
  21. Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters* **8**, 224-239.
  22. Keeling, M.J. (1999) The effects of local spatial structure on epidemiological invasions. *Proc. Roy. Soc. Lond. B* **266** 859-867.

23. Eames, K.T.D. & Keeling, M.J. (2003) Contact Tracing and Disease Control. *Proc. Roy. Soc. Lond. B* **270** 2565-2571
24. Brown, D.H. & Bolker, B.M. (2004) The effects of disease dispersal and host clustering on the epidemic threshold in plants. *Bulletin of Mathematical Biology* **66**, 341-371.
25. Tildesley, M.J. & Keeling, M.J. (2008) Modelling foot-and-mouth disease: A comparison between the UK and Denmark. *PREVET* **85**, 107-124.
26. Garner, M.G. & Beckett, S.D. (2008) Modelling the spread of foot-and-mouth disease in Australia. *Australian Veterinary Journal* **83**, 758 - 766.
27. Akey, B.L. (2003) Low-Pathogenicity H7N2 Avian Influenza Outbreak in Virginia during 2002. *Avian Diseases* **42** Special Issue, Proceedings of the Fifth International Symposium on Avian Influenza, 1099-1103.
28. Marcus, R., Fishman, S., Talpaz, H., Salomon, R. & Bar-Joseph, M. (1984) On the spatial distribution of citrus tristeza virus disease. *Phytoparasitica* **12** (1), 45-52.
29. Dobson, A. & Foufopoulos, J. (2001) Emerging infectious pathogens of wildlife. *Phil. Trans. Roy. Soc B*, **356**, 1001-1012.
30. Lipsitch M. *et al.* (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966-1970.
31. Wearing, H.J. & Rohani, P. (2006) Ecological and immunological determinants of dengue epidemics. *PNAS* **103**, 11802-11807.
32. Savill, N.J., St Rose, S.G., Keeling, M.J. & Woolhouse, M.E.J (2006) Silent spread of H5N1 in vaccinated poultry. *Nature* **442**, 757.

## Figure Legends

**Figure 1** Graphs showing the average density of farms against radius around each farm as the radius varies for the true data (blue line) and random data (red line) in (a) Cumbria, UK and (b) Lancaster County, USA. Insets show farm locations for each respective county for the true data (left plot) and the random data (right plot). The color scale on the insets shows the number of cattle on each farm. (c) Average density of farms against radius around each farm for the true data for Devon, Aberdeenshire, Clwyd in the UK and Cuming, Wright, Humboldt and Franklin in the USA.

**Figure 2** Epidemic impact against ring cull radius for epidemics in (a) Cumbria and (b) Lancaster County. In both figures, the blue line shows the mean epidemic impact for simulations using the true location data whilst the red line shows the mean epidemic impact for the reparameterised random data. Both lines are calculated as locally smooth splines fit to 10000 simulation results. (c) Farm network and (d) mean epidemic impact against ring cull radius for the random data (red line) and the true data (blue line). For (c) and (d),  $S_0 = 4$ ,  $N = 1000$ ,  $B = 0.4$ . In (a), (b) and (d), the black dots show the minima of each line.

**Figure 3** Using the full clustered data, these graphs showing the impact of ring culling at the true ( $R_T$ ) and approximated ( $R_{RR}$ ) optimal radius, as the distribution of farms controlled by the parameter  $B$  and the ratio  $S_0 : S_{\text{inf}}$  vary. In (a) and (b) the random location model is fitted to the entire epidemic derived from a simulation using the spatially clustered location data, whilst in (c) and (d) only the first 14 days of the epidemic are used to fit the random location model. In (a) and (c) the colour scale gives the percentage of farms that would be saved by additional ring culling at radius  $R_{RR}$  compared with IP and DC culling alone. In (b) and (d) the colour scale gives the additional saving in the epidemic impact when ring culling at the true optimal radius

$(R_T)$  compared with at  $R_{RR}$ .