

Accepted Manuscript

Title: Predicting farm-level animal populations using environmental and socioeconomic variables

Authors: Mary van Andel, Christopher Jewell, Joanna McKenzie, Tracey Hollings, Andrew Robinson, Mark Burgman, Paul Bingham, Tim Carpenter



PII: S0167-5877(16)30481-0
DOI: <http://dx.doi.org/doi:10.1016/j.prevetmed.2017.07.005>
Reference: PREVET 4279

To appear in: *PREVET*

Received date: 22-10-2016
Revised date: 4-7-2017
Accepted date: 5-7-2017

Please cite this article as: van Andel, Mary, Jewell, Christopher, McKenzie, Joanna, Hollings, Tracey, Robinson, Andrew, Burgman, Mark, Bingham, Paul, Carpenter, Tim, Predicting farm-level animal populations using environmental and socioeconomic variables. *Preventive Veterinary Medicine* <http://dx.doi.org/10.1016/j.prevetmed.2017.07.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Title:

Predicting farm-level animal populations using environmental and socioeconomic variables

List of Authors and affiliations:

Mary van Andel – Ministry for Primary Industries, Operations Branch, Diagnostic and Surveillance Services Directorate. Wallaceville, New Zealand

Christopher Jewell – CHICAS, Faculty of Health and Medicine, Lancaster University, Lancaster, United Kingdom

Joanna McKenzie – International Projects Group, Institute of Vet, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

Tracey Hollings – Centre of Excellence for Biosecurity Risk Analysis, University Melbourne, Melbourne, Australia

Andrew Robinson - Centre of Excellence for Biosecurity Risk Analysis, University Melbourne, Melbourne, Australia

Mark Burgman – Centre for Environmental Policy, Imperial College, London, United Kingdom

Paul Bingham - Ministry for Primary Industries, Operations Branch, Diagnostic and Surveillance Services Directorate. Wallaceville, New Zealand

Tim Carpenter – Epicentre, Institute of Vet, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

Abstract:

Accurate information on the geographic distribution of domestic animal populations helps biosecurity authorities to efficiently prepare for and rapidly eradicate exotic diseases, such as Foot and Mouth Disease (FMD). Developing and maintaining sufficiently high-quality data resources is expensive and time consuming. Statistical modelling of population density and distribution has only begun to be applied to farm animal populations, although it is commonly used in wildlife ecology. We developed zero-inflated Poisson regression models in a Bayesian framework using environmental and socioeconomic variables to predict the counts of livestock units (LSUs) and of cattle on spatially referenced farm polygons in a commercially available New Zealand farm database, Agribase. Farm-level counts of cattle and of LSUs varied considerably by region, because of the heterogeneous farming landscape in New Zealand. The amount of high quality pasture per farm was significantly associated with the presence of both cattle and LSUs. Internal model validation (predictive performance) showed that the models were able to predict the count of the animal population on groups of farms that were located in randomly selected 3km zones with a high level of accuracy. Predicting cattle or LSU counts on individual farms was less accurate. Predicted counts were statistically significantly more variable for farms that were contract grazing dry stock, such as replacement dairy heifers and dairy cattle not currently producing milk, compared with other farm types. This analysis presents a way to predict numbers of LSUs and cattle for farms using environmental and socio-economic data. The technique has the potential to be extrapolated to predicting other pastoral based livestock species.

Keywords:

Biosecurity, Markov Chain Monte Carlo simulation, Zero-Inflated Poisson Regression, species distribution modelling, spatial epidemiology

Introduction

An exotic disease outbreak can be economically devastating for the primary production industries and related sectors in nations that have substantial primary sector export economies (Blake et al., 2003; Carpenter et al., 2011). The preparation for and rapid control of outbreaks of exotic animal disease are high priorities for the Competent Authorities of these countries.

Information on the size and distribution (i.e., the demography) of the farm animal populations at risk, and their proximity to sources of infection, are required to adequately prepare for and respond to these outbreaks (Woolhouse, 2003). Important aspects of planning for exotic disease outbreak management include: defining a strategy to reduce disease spread, vaccine bank resourcing, and identification of the locations of burial or pyre burning sites for disposal of animal carcasses. Each of these activities relies on knowledge about the farm animal demography of the area in question (Tomassen, de Koeijer et al. 2002). Impacts of insufficient planning for disease outbreak management are numerous and may extend many years after the end of an outbreak as economic losses, ground water contamination and air pollution from carcass disposal sites (Gwyther et al., 2011; Joung et al., 2013).

Demography fundamentally influences how disease will spread (Buhnerkempe et al., 2014; Porphyre et al., 2013a). Estimating the risk of spread of infection of contagious disease from known infected sites is important for prioritising interventions during an outbreak with finite disease control resources (Bessell et al., 2010; Carpenter et al., 2004; Haydon et al., 2003; Jewell and Brown, 2014; Keeling et al., 2001; Thrusfield et al., 2005). Farm animal demographic information facilitates a more timely identification of sites that are at high risk of becoming infected and/or at risk of spreading infection and guides decisions on control strategies early in an outbreak when epidemic information is scarce (Bessell et al., 2010; Hugh-Jones, 1972; Mansley, 2004; Sarandopoulos, 2015; Tomassen et al., 2002). This prioritisation is critical because delays in identification of infected premises (IP) can have a major effect on epidemic impacts (Carpenter et al., 2011).

Disease spread simulation models, which use animal demographic information, are useful for planning and preparing for exotic animal disease outbreaks. Models of disease transmission in farm animal populations typically assume that infection rate is a function of farm-level characteristics and the distance between pairs and contact structure of farms. Thus the distribution of farms and populations on the farms at risk are critical

components of the effective reproduction number, R_e (Dohoo et al., 2009, Porphyre et al., 2013). In the model, estimates of R_e indicate whether an outbreak is under control and can inform policy adjustments which are implemented in the field (Ferguson et al., 2001; Paine et al., 2010). If the population data are inaccurate the model predictions of disease spread and impact will be inaccurate in ways which are not readily quantifiable without some model based structure to capture their uncertainty.

Competent Authorities may maintain national animal registries of farm and livestock population distribution (Paiba et al., 2007). The maintenance of accurate data within central animal registries is an ongoing challenge; particular areas of difficulty include: recording of births and deaths, real time animal movements, and mismatches and inconsistencies between different sources of data in the repository (Jewell et al., 2015; Savill et al., 2007). Other Competent Authorities have addressed a lack of animal population data by simulating livestock populations within a country (Burdett et al., 2014) or globally (Robinson et al., 2014). Although simulated livestock population data are available on a global scale at 1 km² resolution from the Food and Agriculture Organization's Gridded Livestock of the World Project (Robinson et al., 2014), this is not sufficiently accurate at the farm level to plan disease control operations, which require knowledge of the numbers and species present on each IP (Mansley, 2004). An alternative approach, which is required in some countries for which details of farm locations and their animal populations are not available, is to model the location of farm sites and farm types based on environmental variables (Burdett et al., 2015).

New Zealand livestock farming is predominantly pasture based and the main stock categories farmed are dairy cattle, beef cattle, sheep and deer. These populations are represented in multiple repositories of animal demographic information (Jewell et al., 2015), which are maintained by the government and private industry bodies. One of these, AgriBase, is owned byASURE Quality Limited, a New Zealand state-owned enterprise that provides operational services to the Ministry for Primary Industries (MPI). MPI maintains a second properties database FarmsOnline (FOL) for biosecurity purposes but livestock counts are not available in this data source. Both AgriBase and FOL hold spatially referenced information on properties used for all farming activities including those used for recreational farming or "lifestyle" farming as it is known in New Zealand. However, national databases are expensive to maintain, and require full engagement of the landowners to keep current

both in terms of land parcels and livestock kept. The result is that there may be a high level of unquantified uncertainty around what we would otherwise like to regard as census data (Jewell et al., 2015). In this study, we investigate an approach to making use of sample-based data with known accuracy to predict counts of counts of livestock units (LSU's) and cattle present in production systems where animals live mainly outdoors and eat a predominantly pasture based diet on unsampled farms in New Zealand using environmental and socio-economic data. An LSU is a reference unit with allows various species of livestock to be aggregated based on nutritional requirement of each animal type; we have used the European Union definition of an LSU for this study (Anonymous, 2013).

The balance of the paper is constructed as follows. In the next section we describe the data sets, data handling, and modelling strategy, and the means by which we assessed and compared the models. We then provide a detailed description of livestock prediction in New Zealand, with Discussion, and our Conclusion. Supplementary material to the materials and methods section is provided.

Materials and Methods

Databases and study variables:

The predictive model was built using farm boundary and animal population data from AgriBase (Sanson and Pearson, 1997) which contained 135 002 farms. AgriBase contains spatially referenced farm boundary data and species-specific counts of animal populations for each farm. Some farm boundaries in the dataset overlapped one another as they are farmed under cross-lease agreements. These 6586 (4.5% of the total available farm records) overlapping farms were deleted from the dataset to avoid bias in predicting animal counts. In this case both of the overlapping farms were deleted. We deleted a further 36815 properties (27.02% of the total available farm records) that had no information on production type of any kind (agriculture or horticulture based) because we assumed that these represent missing data rather than the true absence of the animal species of interest. The remaining 91601 farms (68.50%) were used for model building and validation.

The two outcome variables of interest were farm level counts of LSUs and of cattle; both were drawn from the AgriBase dataset (Sanson and Pearson, 1997). LSUs were calculated for sheep, goats, cattle, pigs, deer and

horses following the European Union definition of individual species requirements in relation to a dairy cow (Anonymous, 2013) shown in Table 1. Poultry data were not included in the model.

Publicly available land cover, environmental and socio-economic data held by Land Information New Zealand (LINZ), Statistics New Zealand, Land Environments of New Zealand (LENZ) and Land Resources Information New Zealand (LRINZ) were used to generate candidate predictor variables (Table 2). The New Zealand Transverse Mercator (NZTM) projection was used for geographic information projection in this analysis. Predictor variables were loaded into a geographic information system (GIS) relational database (PostgreSQL 8.4 with PostGIS 2.2.0 extensions). This provided a convenient platform for interfacing with statistical software (Rv3.0.1, Foundation for Statistical Computing, Vienna, Austria), enabling large amounts of geospatial data to be processed and extracted for each farm. Predictor variables included in the models, where they were sourced and how they were derived are presented in Table 2.

A Spearman correlation matrix was used to identify significant pairwise collinearity amongst the predictor variables, defined arbitrarily as a correlation of $|r| > 0.7$. Our choice to use the Spearman's correlation coefficient rather than a variable selection process within the modelling framework allowed the analysts to make variable selections based on knowledge of the biological system under examination. Collinearity was identified between the following pairs of variables: mean *annual temperature* and mean *minimum temperature*, mean *solar exposure* and minimum *solar exposure*, and *slope* and *aspect*. Collinearity was addressed by creating bivariate transformations to include in the model and by retaining one of each of the collinear pairs (see supplementary material equations 1 and 2 for further detail of the bivariate transformations).

An interaction term was included to represent the relationship between *deprivation index* and *distance from urban centres* to capture the population of lifestyle farms. Statistics New Zealand notes that this population of lifestyle farms in rural areas with high urban influence has the highest median and mean income (and household expenditure) of any demographic in New Zealand (Pink, 2011). Land held by this population demographic, which commutes to city centres for work, is thus unlikely to be used in the same way as land further from cities, where agriculture is the principal source of income. The log of the area in hectares of each

farm covered by *high quality pasture*, *low quality pasture*, *native forest* and *endemic forest* and the complement of the farm area were used in the final model. Each of these was offset by 0.01 to make the log transformation possible for farms with no coverage of some of the categories. The log transformation of these variables was used in the modelling to promote for numerical stability.

Statistical model framework

A zero-inflated Poisson (ZIP) regression model was used to generate the predicted counts of LSUs and of cattle for farm properties. A ZIP model was chosen because many rural properties are not engaged in livestock production; AgriBase records 37% of properties as having no animals present (Jewell et al., 2015).

Consequently, the ZIP model fit the data significantly better than did a comparable Poisson model.

The model was trained to predict farm level populations using a Bayesian approach that calculated a distribution of the predicted number of cattle or LSUs within a farm polygon on the land surface in New Zealand. The outcome variable for an individual farm (either number of cattle or LSUs) was modelled as a ZIP random variable, conditional on the values of the predictor variables for that farm (Mullahy, 1986). A two-stage process was thus used, in which (i) a Bernoulli random variable first represented the probability of animals being present on a farm; (ii) then, assuming that animals are present, the number of animals was assumed to follow a Poisson distribution with mean related via a log link function to the predictor variables. Both the Bernoulli and the Poisson stages involved modelling using the suite of predictor variables. Further detail of the model framework, model equations and model code are available in the supplementary materials.

The vectors of coefficients are calculated for the same set of predictor variables for the both the Poisson and Bernoulli components. We have allowed the prediction of the excess zeros and the prediction of the counts on each farm to be conditional on the same predictor variables, however as Table 4 shows, different predictors are identified as important in the Poisson regression step when compared with the Bernoulli step, thus x_i (equation 5 and 6 in the supplementary material) appears in the linear predictors for both the Poisson and the Bernoulli steps of the model. The importance of each covariate in the final model was explored using reversible jump MCMC which is fully described in the supplementary material (Carlin and Chib, 1995; Green, 1995).

The MCMC was coded in R using a single-site adaptive Metropolis-Hastings step for each coefficient (Chib and Greenberg, 1995; Jewell, 2016a). A total of 50 000 iterations of each model was performed and the first 20 000 runs were discarded as burn-in leaving 30 000 iterations which were thinned systematically every 30 iterations to 1000 observations to be used in all further analyses.

Predictive performance:

To assess how well the model was able to predict the data which it was built from, a frequentist model residual analogue was calculated (rank probability score, referred to hereafter and in the supplementary material as RPS) and five “residual”, i.e., RPS checking analyses were completed. The statistical detail of the calculation of the RPS and the analyses performed on this residual analogue are explained in detail in the supplementary material.

The first two methods of residual (RPS) checking were: i) an examination of spatial autocorrelation of RPS in two large regions (figure 1 shows the regions and figure 2 the variograms), and ii) the differences in model performance amongst farm types (figure 3). In these analyses the farm is the unit of interest. The third and fourth methods were performed at iii) small (3km radial zone) and iv) larger (New Zealand administrative regions) spatial scales and calculated area level RPS values (table 3 and figure 4). In these analyses the area (either the region or the 3km zone) is the primary unit of interest. The final method v) used Receiver Operator Characteristic (ROC) curves (figure 5) to examine the performance of the Bernoulli component of the ZIP model (which predicts the presence or absence of livestock or cattle on the farm of interest).

We chose 3km radial zones as a unit of interest to examine model performance because the outcomes (farm level counts of cattle or LSUs) of these models are primarily required for biosecurity purposes. In the event of a suspected or actual outbreak of FMD, the population in the 3km radial zone around an infected farm is at high risk of local spread (Thrusfield et al., 2005). This means that accurate data on susceptible animal populations on farms within 3km of a farm with a suspected outbreak of FMD should be available as early as possible in the investigation and risk assessment process. In the New Zealand context, a 3km zone is an appropriate choice based on local observations and farm density.

Results

Predictor variables

Predicted coefficients of covariates for the Bernoulli and Poisson components of the ZIP models and their associated odds ratios and risk ratios are presented in Table 4. The most important covariate for predicting the presence or absence (Bernoulli component of the ZIP model) of cattle and of LSUs was the log of the area (hectares) of high quality pasture followed by region in which the farm was located. Region was the most important predictor variable for counts of LSUs and cattle (the Poisson part of the models). The use of the zero-inflated distribution allows the excess appearance of zeros to be linked to the covariates. As an example, consider the Bernoulli component of the cattle model. As shown in the model covariate table presented in Table 4, the size of the area of high quality pasture present on the farm of interest is predictive of the presence of cattle (OR 1.91, CI between 1.91 and 1.92), however some regions have additional increases in the likelihood of cattle being present (multiplicative at the log scale with the covariate representing the size of high quality pasture). The difference between the covariates for the regional fixed effects for Northland and Nelson illustrate this point well. If the Manawatu- Wanganui region (a region with a large amount of farm land and a sizeable cattle population) is taken to be the referent category, then the Nelson region is half as likely (OR of 0.52 with CI between 0.5 and 0.54 to have cattle present on a particular farm of interest, and farms in the Northland region are 1.77 times more likely with CI between 1.75 and 1.83 to have cattle present on the property. No evidence to remove any of the covariates from either the Poisson or Bernoulli parts of the model was found when the covariates were explored using reversible jump MCMC. We have included the OR for the interaction terms in table 4 even though the OR for an interaction term cannot be interpreted without taking into consideration the OR for both of the terms involved in the interaction. Table 4 is a summary of the joint posterior distribution over the coefficients. It displays marginal information for each coefficient and ignores correlations between the respective marginal posterior distributions, and the fact that for some realisations of the posterior the value of certain coefficients may be 0.

Predictive performance

The primary measure of the performance of our models is how well they predicted counts of LSUs or cattle at various spatial scales as measured by the RPS. Results for internal validation at each of the five scales are presented below.

- i) The spatial variogram of farm-level RPSs showed slight spatial autocorrelation in the LSU model over short distances in both the North Island and South Island test regions (Figure 2). This correlation was apparent up to about 30km in the North Island and about 50km in the South Island for the LSU model. In contrast, in the cattle model in the North Island, correlation of the RPS persisted up to 60km and in the South Island it reached a sill at around 50km. Autocorrelation in RPS was consistently larger in the South Island when compared with the North Island for the LSU model indicating a higher level of spatial correlation of model error.
- ii) Significant differences in mean RPS values were found for groups of farms stratified by the farm type recorded in AgriBase (dairy farms, pastoral livestock farms, pig breeder and finisher farms, grazing and dairy heifer rearing properties and other farm types) when tested with the Kruskal-Wallis test ($p < 0.001$ with 6 degrees of freedom) for both the LSU and the cattle models. In both models, the farm type 'dry stock' which included contract grazing and dairy heifer rearing properties were predicted with the least accuracy (figure 3).
- iii) Prediction at the 3km zone level was most accurate with mean RPS of 0.093 for the LSU model and 0.162 for the cattle model summarised over the 10 model fits (table 3).
- iv) When region was removed from the model to allow internal validation between regions, the model that predicted cattle numbers performed better overall than the model that predicted LSUs (mean RPS of 0.317 and standard error of 0.062 for the LSU model compared with a mean of 0.247 and standard error 0.033 for the cattle model) (table 3). When regional RPSs from each model run were aggregated and plotted a large amount of variability in the RPSs was observed at the region level (figure 4).
- v) The area under the curve (AUC) was estimated for the models and thresholds that maximised both sensitivity and specificity. As an example, for the model that predicted LSUs, when the cutoff for representing the posterior prediction of having no animals on each property was 0.25, 90% of farms with no animals were correctly identified (specificity). The sensitivity of this estimate was poor (40%) with many false negative results (i.e. farms identified by the model as having no animals, which in fact had animals present). The largest area under the curve (AUC) for the LSU model was found at a point mass of zeros of 0.15, representing correct identification of

68% of the farms without animals. In the cattle model the area under the curve was maximised at a cutoff of 0.35 where 68% of the area was captured.

Discussion

To the best of our knowledge no work like this has been published in the peer-reviewed literature for New Zealand. The small scale of our spatial unit of interest (the farm), our focus on operational uses of the data for controlling disease outbreaks (as opposed to use in disease spread models only) and the large amount of existing farm-level animal data for New Zealand set this study apart from similar studies performed in other countries. Our technique differs from the FLAPs approach (Burdett et al., 2015) as it does not model farm locations but rather models the animal counts for known farm polygons. When compared with the approach taken by the Gridded Livestock of the World (GLW) project (Robinson et al., 2014) the main difference is that our model provides counts at the farm level (rather than a raster cell level count or animal density), allowing farm level (rather than raster cell level) attributes to play a role in predicting animal counts.

A model that included the log of the area (hectares) of high quality pasture and region in which the farm was located predicted presence or absence of cattle and of livestock in general (represented as LSUs) with reasonable accuracy. Region was the most important predictor variable for counts of LSUs and cattle.

The models performed better when used to predict the aggregated cattle or LSU population of 3km radial zones. Prediction at the level of single farms was poorer (as could be expected). Even though the purpose of the models is to predict presence or absence of animals and the counts of animals present on the farms rather than to examine the effects of different covariates, the relatively uninformative nature of many covariates in the models should be mentioned (that is, the odds ratios for many covariates in the predictive models are close to one, as shown in Table 4). We believe that these covariates still explain some of the variance (and improve prediction) without necessarily explaining an amount of variance beyond the traditional probability threshold value of 0.05. In addition, many of these covariates that do not explain a large amount of variance may have been expected to play a larger role if the modelling had not provided evidence to the contrary. At the farm level we were able to identify that while counts on dairy farms were predicted more accurately (low RPS shown in figure 3), predicted counts were less accurate for farms classified as contract grazing and dairy heifer rearing. These results could be due to real differences in practices such as constantly changing stock

numbers on dairy heifer rearing and contract grazing properties as animals are bought and sold and animals arrive for and leave after grazing contracts. These factors mean that contract grazing properties do not have the same predictable animal counts as properties of a similar size which have more stable population of resident animals. Alternatively, the lower accuracy in prediction could be due to misclassification bias in the dataset used to build the model. Only an exercise that collects actual farm level data will clarify the reason for low predictability of LSUs and cattle numbers on contract grazing and dairy heifer rearing farms. The high level of regional heterogeneity in the farm-level animal populations in New Zealand makes some regions appropriate training data sets for particular regions but not for others (figure 4). In particular, the Waikato (an area of high dairy cattle farming density) is well predicted by the models and the Tasman Area is very poorly predicted.

The livestock unit (LSU) model is important to test the hypothesis that in an animal production system where animals live mainly outdoors and eat a predominantly pasture based diet, remotely sensed geospatial layers provide sufficient information to extrapolate counts of LSUs at the farm level. As biosecurity applications include stocking of vaccine banks and predicting the spread of disease based on species composition at the farm level, such a model or the outputs of the LSU model must be adjusted to predict single (we present the cattle model as an example) or specified combinations of species populations on each farm (Burdett et al., 2015; Robinson et al., 2014). A useful characteristic of our model is the ability to predict LSU and cattle counts in the 3km radial zones with a high degree of accuracy reflected by the lower RPS for these zones compared with the predictions for individual farms (Table 3). As accurate data on susceptible animal populations on farms within 3km of a farm with a suspected outbreak of FMD are needed by the responders to a disease event as early as possible in the New Zealand context our model may provide some idea of the magnitude of the risk presented by an IP before farm level information can be collected.

The advantage of using Bayesian inference to estimate the coefficients of the ZIP model is that a predictive distribution of the number of animals on a farm may be estimated explicitly, without resorting either to asymptotic assumptions or complex bootstrapping procedures (West, 2014). This provides a precise measure of uncertainty surrounding predicted animal numbers for each farm. Samples from the predictive distribution may then be used as input for Monte Carlo disease simulation models, such as the Davis Animal Disease Spread model (DADs) and the Australian Animal Disease Spread model (AADIS), so that uncertainty in the

population at risk of infection is reflected in the simulated outputs (Bates et al., 2003; Bradhurst et al., 2015).

This is an area of future exploration. Having information on farm-level populations is important because in a disease outbreak, not all farms have the same risk of disease (Gates et al., 2014). In particular (in the context of this study), the importance of herd size is well established as a risk factor for disease spread based on detailed data collected from outbreaks in naïve populations (Hugh-Jones, 1972; Keeling et al., 2001; Kitching et al., 2006; Muroga et al., 2013; Tildesley et al., 2008; Tildesley and Ryan, 2012).

Advances in modelling techniques and increased computational ability have resulted in a trend toward using individual animals as the unit of interest in disease spread modelling. The inclusion of within-herd spread in detailed process disease spread models is being employed to represent the effects of farm and herd level heterogeneity (Bates et al., 2003; Bradhurst et al., 2015). These models are being used to evaluate the impacts of management decisions at a microeconomic level on individual farms (Halasa et al., 2013; Porphyre et al., 2013a) and variation in start conditions (farm level populations). The effect of herd size on disease spread has the potential to alter the model outputs that inform policy decisions. The adoption of such modelling techniques requires farm level animal population data and a clear understanding of the inadequacies in the data used and the impacts of any inadequacies on model outcomes. The accuracy of these data are seldom understood or examined (Honhold and Taylor, 2006; Jewell et al., 2015). Recent studies have explored how different models may be used to represent similar outbreak situations and have compared model outputs within and between countries (Dubé et al., 2007; Roche et al., 2015; Sanson et al., 2013, 2011) but little attention has been paid in the published literature to the effect of the population information on the model outcomes and resulting policy advice (Wada et al., 2016). Similarly, although demographic data were identified as lacking during the 2001 FMD outbreak in the UK, the extent of the impact of the inaccuracies on either operations or decisions made based on disease spread modelling is hard to establish (Honhold and Taylor, 2006; Keeling et al., 2001; Kitching et al., 2006; Thrusfield et al., 2005).

Spatial correlation of the RPSs was present at distances less than 18km for the LSU model and distances less than 70km for the cattle model after inclusion of region and other environmental variables. This spatial autocorrelation of RPS could be reduced in future models by identifying missing predictor variables that capture the spatial autocorrelation or by the inclusion of a statistical effect which would extend the ZIP model

to a geostatistical model (Diggle and Ribeiro, 1998). Given the large number of farms present in our dataset the addition of this statistical effect would pose a significant computational challenge.

The interpretation of the ROC curves depends on the purpose for which the modelled data are being used.

Unless we are able to measure the cost of error and uncertainty around knowing and not knowing if an animal is present on a particular property, it is difficult to choose a preferred cutoff at which to interpret the performance of the Bernoulli component of the model and to appreciate the impact of uncertainty associated with that cutoff. In our case we used the ROC curves as a predictive performance exercise to examine the performance of the Bernoulli component when predicting the absence of animals on a farm. Our interpretation of the results of the ROC curves is that our proposed model does provide valuable information about the demographics of animals present on a farm but it is not able to predict with acceptable accuracy the presence or absence of an animal population. This finding agrees with other studies that used spatial environmental variables to predict the distribution of animal species - distinguishing between areas with low numbers of animals and areas where animals were absent is seldom achievable (Porphyre et al., 2013b).

Aside from the statistical variability that our model is able to estimate around each population estimate (the error), ongoing fluctuations can be expected to occur due to seasonal changes in production and changing long term trends in species composition in different areas. An appreciation of these population trends is important for animal disease outbreak response, as having incorrect data representing a population early in a disease response may delay or influence decision making to the same degree as a lack of information could delay or influence response efforts. Further work is needed to examine how long term and seasonal demographic changes are best represented in the model. An approach to long term changes is to use the presented LSU model and combine this with information present in the New Zealand agricultural census. Strategies that could be explored to address seasonal changes include the incorporation of regional level animal movement data into the model or the use of normalised difference vegetation index (NDVI), which has successfully been used in the past to predict herbivore migration associated with pasture availability in ecological models (Marshal et al., 2006; Mueller et al., 2008; Myneni et al., 1995; Reed et al., 1994; Tveraa et al., 2013). In addition to seasonal and production linked trends, herd size is affected by economic factors such as the milk price. These sources of uncertainty resulting from seasonality and economics suggest that maintenance of up to date national level registries is still an important function for the competent authority for biosecurity.

We are able to identify some sources of bias in our models. Aside from bias present in the data used to build the models, which may be affected by response bias by those populating the data and will be affected by the currency of the data held in the databases, our methods may have introduced bias. We have mentioned that 4.5% of farm records were deleted due to overlapping polygons and that another 27% were deleted due to missing data on production type. If either missing-ness or overlapping-ness are non-randomly distributed amongst the full dataset, bias will result from their removal. When properties were selected to have a 3km zone situated around them, this was done based on the property overlapping the centroid of the grid square. This biases our selection method toward larger farms raising the question of whether the accuracy of fit is affected by the selection of farms in a 3km zone surrounding larger farms when compared with those farms surrounding small farms. Finally when distances to features were measured, the calculation was made from the centroid of the farm to the closest feature rather than from the farm boundary to the closest feature. It is possible that this might bias the estimate of the distance from a farm to a feature for larger farms when compared with smaller farms predisposing our models to type 1 error.

Although our goal was to predict the animal numbers on farms for which data were not available, we have only been able to test the model using withheld samples of farms for which the data were available, that is, we have performed internal validation. Internal validation is inherently unsatisfying; it demonstrates that the sample of data that is available shows reasonable consistency from the point of view of fitting the model, but does not provide solid evidence of the extensibility of the model to novel situations. Therefore for the deployment of the model it is necessary for us to assume that the farm population for which data are missing is represented statistically by the farms for which we do have information. There are numerous reasons that would undermine this assumption, for example, that the missing farms have a greater proportion of peri-urban properties, that non-inclusion is not homogeneously distributed across all production types or that overlapping properties are not randomly distributed in space. However, we proceed under the assumption that some information is better than no information. Further work is underway to develop an independent dataset that can be used for a more complete external validation of our models, but that is beyond the scope of this article.

Conclusions

We have constructed a model that predicts the presence and absence as well as the numbers of livestock units and cattle at the farm level in New Zealand. While wildlife ecology uses similar models for conservation and research purposes, the use of modelled or inaccurate point estimates of animal populations for biosecurity operations (rather than when preparing for a disease outbreak) may introduce systematic error which could increase the number of properties affected by, and the eventual cost of a fast moving animal disease.

Predictions were accurate at a wide spatial scale (3km zones or administrative regions) and our model allows us to explore uncertainty around a point estimate in animal numbers at the farm level by incorporating farm population posterior distribution estimates in disease simulators. Future work will further refine the modelling approach to improve its internal validity and provide external validation of the model using data collected from the field to fully assess the precision of the estimates of livestock numbers and of cattle on farms and to test the extensibility of the derived covariates to predict farm level counts on an unrelated spatial farms database (FarmsOnLine). While probabilities of presence and distributions of animal counts are useful in exotic disease preparedness and disease spread simulation modelling, when faced with the already considerable uncertainty inherent early in a disease outbreak response the decision maker requires the most accurate and correct information to make high quality decisions. Every reasonable effort must be made to strengthen data linkages between existing animal population data sources by aligning data collection efforts undertaken by government and industry and by the use of a single, national farm identifier.

Acknowledgements

We thank the Ministry for Primary Industries in New Zealand for support of this project and the Centre of Excellence for Biosecurity Risk Analysis (CEBRA) at University Melbourne for funding. Thanks are due toASUREQuality and Robert Sanson for the use of AgriBase data for this project; to Hannes Calitz of the Institute of Fundamental Sciences at Massey University for assistance with grid computing; to Graeme Garner of the Australian Federal Department of Agriculture and Water Resources, Daan Vink of the Ministry for Primary Industries and Carolyn Gates of the EpiCentre at Massey University who provided input on early on drafts of this work.

Table 1: Species equivalence values used to calculate livestock units (LSUs) on rural properties in New Zealand, following the European Union definition of individual species feed requirements (Anonymous, 2013).

	Livestock Units (LSUs) relative to a grazing dairy cow
Dairy cattle	1
Beef cattle	0.8
Deer	0.8
Sheep & Goats	0.1
Pigs	0.4
Horses	0.8

Table 2: Predictor variables included in a Bayesian Zero Inflated Poisson model predicting counts of cattle and counts of livestock units on New Zealand farms.

* Predictor variables included in the final models. Δ Predictor variables significant in the LSU model. ¥ Predictor variables significant in the cattle model

Predictor Variable	Description (units)	Reference
Distance to sealed roads * $\Delta\text{¥}$	Distance from the centroid of the farm to the nearest major sealed road (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to urban centres * $\Delta\text{¥}$	Distance from the centroid of the farm to the nearest urban centre (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to rivers * $\Delta\text{¥}$	Distance from the centroid of the farm to the nearest river (km)	Land and Information NZ (LINZ). https://data.linz.govt.nz/
Social deprivation index * $\Delta\text{¥}$	Value at the farm centroid (categorical variable between 1 and 10 with 1 being the greatest degree of deprivation)	Statistics NZ census data. Decile measure of socio-economic status
Interaction term*	Interaction term between deprivation index and distance to urban centres to account for the effect of lifestyle farming	
Area of high quality pasture (log of) * $\Delta\text{¥}$	Area of grassland per farm classified as high producing by the national land cover database (ha)	Land Cover Database (LCDB) Version 4. A satellite derived GIS vector polygon layer of land use. https://iris.scinfo.org.nz
Area of low quality pasture (log of) * $\Delta\text{¥}$	Area of grassland per farm classified as poorly producing grassland by the national land cover database (ha)	Land Cover Database (LCDB) Version 4. A satellite derived GIS vector polygon layer of land use. https://iris.scinfo.org.nz
Mean annual temperature	The average annual temperature across the farm calculated by averaging each raster cell of a modelled temperature surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Mean minimum temperature	The average minimum temperature across the farm calculated by averaging each raster cell of a modelled temperature surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Temperature difference*	Mean annual temperature minus the mean minimum temperature for each farm	
Mean solar radiation	The annual average solar radiation across the farm calculated by averaging each raster cell of a modelled solar radiation surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com

Mean winter solar radiation	The average minimum daily solar radiation across the farm calculated by averaging each raster cell of a modelled solar radiation surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Solar difference*Δ¥	The difference between the mean solar radiation and the mean winter solar radiation for each farm	
Mean water deficit*Δ¥	The average of the annual water deficit across each farm calculated by averaging each raster cell of a modelled water deficit surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Mean aspect	Mean of farm aspect or the direction in degrees which a slope faces (degrees)	
Mean slope	Mean of farm slope (slope is a unitless ratio)	
Slope x Aspect*	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Sine of Slope*Δ¥	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Cosine of Slope*Δ¥	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Standard deviation of slope*Δ¥	Standard deviation of slope for the farm	
Mean and standard deviation of elevation*Δ¥	Mean and standard deviation of elevation for the farm calculated by averaging each raster cell of a surface generated by a digital elevation model contained within the farm polygon (m)	Digital elevation model from Land and Information NZ (LINZ). https://data.linz.govt.nz/
Regional Council*Δ¥	One of the 17 New Zealand Regional councils that the farm centroid falls within (categorical variable)	https://koordinates.com/layer/4240-nz-regional-councils-2012-yearly-pattern/
Size of farm	Farm area (ha)	Agribase™ for the farm boundaries and post gis https://koordinates.com/layer/168-nz-native-polygons-topo-150k/ & https://koordinates.com/layer/131-nz-exotic-polygons-topo-150k/
Forest cover (log of)*Δ¥	The sum of endemic and exotic forest cover for each farm (ha)	
Remaining land cover (log of)* Δ¥	Remaining area of the farm once high quality pasture, low quality pasture and forest cover had been taken from the total (ha)	

Table 3: Mean and standard deviation of rank probability scores (RPS) for models predicting livestock units (LSUs) and cattle on farms within 3 km zones and administrative regions in New Zealand.

Target data predicted	Number of fits	Level of RPS calculation	LSU model		Cattle model	
			mean RPS	Standard deviation	mean RPS	Standard deviation
200 3km zones around randomly selected farms	10	farm	0.221	0.006	0.214	0.014
	10	area (3km zone)	0.093	0.009	0.162	0.0122
3 North Island and 2 South Island regions	50	farm	0.317	0.062	0.247	0.033
	50	area (administrative region)	0.261	0.227	0.265	0.219

Table 4: Summary of the mean and standard deviation of the joint posterior estimates of covariates for species distribution models for New Zealand farms using counts of livestock units (LSU) as the outcome variable in the first instance and counts of cattle in the second instance. Risk ratios are reported for the Poisson component and odds ratios for the Bernoulli component for each model along with 95% credible intervals.

Predictor variable	LSU model			Cattle model		
	mean	std dev	OR (5%-95%CI)	mean	std dev	OR (5%-95%CI)
Bernoulli component						
intercept	3.353	0.007	28.58 (28.48 - 28.68)	3.074	0.012	21.62 (21.53 - 21.79)
Other land cover in hectares (log)	0.013	0.000	1.01 (1.01 - 1.01)	0.008	0.000	1.01 (1.01 - 1.01)
High quality pasture in ha (log)	0.628	0.002	1.87 (1.87 - 1.88)	0.649	0.003	1.91 (1.91 - 1.92)
Low quality pasture in ha (log)	0.009	0.000	1.01 (1.01 - 1.01)	0.006	0.000	1.01 (1.01 - 1.01)
Forest cover in ha (log)	0.014	0.000	1.01 (1.01 - 1.01)	0.009	0.000	1.01 (1.01 - 1.01)
Distance to a road from farm centroid	0.050	0.000	1.05 (1.05 - 1.05)	0.072	0.001	1.07 (1.07 - 1.08)
Distance to an urban centre from farm centroid	0.006	0.000	1.01 (1.01 - 1.01)	0.003	0.000	1 (1 - 1)
Temperature difference	0.006	0.000	1.01 (1.01 - 1.01)	0.004	0.000	1 (1 - 1)
Solar difference	0.011	0.000	1.01 (1.01 - 1.01)	0.000	0.001	1 (1 - 1)
Distance to a river from farm centroid	0.007	0.000	1.01 (1.01 - 1.01)	0.007	0.000	1.01 (1.01 - 1.01)
Mean slope	-0.029	0.000	0.97 (0.97 - 0.97)	-0.047	0.001	0.95 (0.95 - 0.96)
Standard deviation of slope	-0.015	0.000	0.99 (0.98 - 0.99)	-0.019	0.000	0.98 (0.98 - 0.98)
Mean digital elevation model	-0.001	0.000	1 (1 - 1)	-0.002	0.000	1 (1 - 1)
Standard deviation of digital elevation model	0.002	0.000	1 (1 - 1)	0.002	0.000	1 (1 - 1)
Cosine of slope	0.051	0.000	1.05 (1.05 - 1.05)	0.049	0.001	1.05 (1.05 - 1.05)
Sine of slope	0.008	0.001	1.01 (1.01 - 1.01)	0.006	0.001	1.01 (1 - 1.01)
Product of slope and aspect	0.000	0.000	1 (1 - 1)	0.000	0.000	1 (1 - 1)
Deprivation index	0.022	0.000	1.02 (1.02 - 1.02)	0.027	0.001	1.03 (1.03 - 1.03)

Deprivation index * Distance to an urban centre	-0.001	0.000	1 (1 - 1)	-0.002	0.000	1 (1 - 1)
Mean water deficit	-0.002	0.000	1 (1 - 1)	0.019	0.002	1.02 (1.02 - 1.02)
Mean solar hours	-0.009	0.000	0.99 (0.99 - 0.99)	-0.024	0.002	0.98 (0.97 - 0.98)
Mean temperature	-0.001	0.000	1 (1 - 1)	-0.001	0.000	1 (1 - 1)
Fixed effects for region (Manwatu-Wanganui as the referent category)						
Auckland	-0.159	0.004	0.85 (0.85 - 0.86)	0.145	0.012	1.16 (1.14 - 1.18)
Bay of Plenty	0.204	0.004	1.23 (1.22 - 1.23)	0.333	0.008	1.4 (1.38 - 1.41)
Canterbury	-0.047	0.004	0.95 (0.95 - 0.96)	-0.055	0.016	0.95 (0.92 - 0.96)
Gisborne	0.548	0.004	1.73 (1.72 - 1.74)	0.764	0.008	2.15 (2.13 - 2.16)
Hawkes Bay	0.106	0.003	1.11 (1.11 - 1.12)	0.130	0.004	1.14 (1.13 - 1.14)
Marlborough	-0.359	0.005	0.7 (0.69 - 0.7)	-0.361	0.016	0.7 (0.68 - 0.71)
Nelson	-0.718	0.016	0.49 (0.47 - 0.5)	-0.660	0.025	0.52 (0.5 - 0.54)
Northland	0.169	0.004	1.18 (1.18 - 1.19)	0.573	0.016	1.77 (1.75 - 1.83)
Otago	-0.344	0.004	0.71 (0.71 - 0.71)	-0.586	0.021	0.56 (0.53 - 0.57)
Southland	-0.288	0.003	0.75 (0.75 - 0.75)	-0.547	0.023	0.58 (0.55 - 0.59)
Taranaki	0.179	0.002	1.2 (1.19 - 1.2)	0.340	0.005	1.4 (1.4 - 1.41)
Tasman	-0.459	0.005	0.63 (0.63 - 0.64)	-0.310	0.015	0.73 (0.71 - 0.75)
Waikato	0.257	0.003	1.29 (1.29 - 1.3)	0.447	0.007	1.56 (1.56 - 1.58)
Wellington	-0.066	0.003	0.94 (0.93 - 0.94)	-0.157	0.004	0.85 (0.85 - 0.86)
West Coast	-0.341	0.004	0.71 (0.71 - 0.71)	-0.179	0.010	0.84 (0.82 - 0.85)

Poisson component	LSU model			Cattle model		
	mean	std dev	RR (5%-95%CI)	mean	std dev	RR (5%-95%CI)
intercept	-2.438	0.046	0.09 (0.08 - 0.09)	-1.394	0.036	0.25 (0.23 - 0.26)
Other land cover in hectares (log)	0.022	0.002	1.02 (1.02 - 1.02)	0.004	0.001	1 (1 - 1.01)
High quality pasture in ha (log)	-0.040	0.003	0.96 (0.96 - 0.97)	-0.086	0.003	0.92 (0.91 - 0.92)
Low quality pasture in ha (log)	-0.015	0.003	0.98 (0.98 - 0.99)	-0.018	0.002	0.98 (0.98 - 0.99)
Forest cover in ha (log)	0.009	0.002	1.01 (1.01 - 1.01)	-0.002	0.001	1 (1 - 1)
Distance to a road from farm centroid	-0.049	0.027	0.95 (0.91 - 0.99)	-0.199	0.027	0.82 (0.78 - 0.86)
Distance to an urban centre from farm centroid	-0.011	0.002	0.99 (0.99 - 0.99)	-0.025	0.002	0.98 (0.97 - 0.98)
Temperature difference	0.001	0.002	1 (1 - 1)	0.000	0.001	1 (1 - 1)
Solar difference	0.020	0.008	1.02 (1.01 - 1.03)	0.038	0.006	1.04 (1.03 - 1.05)
Distance to a river from farm centroid	-0.005	0.003	1 (0.99 - 1)	-0.011	0.002	0.99 (0.99 - 0.99)
Mean slope	0.107	0.006	1.11 (1.1 - 1.12)	0.123	0.006	1.13 (1.12 - 1.14)
Standard deviation of slope	-0.160	0.008	0.85 (0.84 - 0.86)	-0.145	0.006	0.86 (0.86 - 0.87)
Mean digital elevation model	0.002	0.000	1 (1 - 1)	0.003	0.000	1 (1 - 1)
Standard deviation of digital elevation model	-0.008	0.001	0.99 (0.99 - 0.99)	-0.013	0.001	0.99 (0.99 - 0.99)
Cosine of slope	0.032	0.016	1.03 (1.01 - 1.06)	0.007	0.012	1.01 (0.99 - 1.03)
Sine of slope	0.001	0.017	1 (0.97 - 1.03)	-0.036	0.014	0.96 (0.94 - 0.99)
Product of slope and aspect	0.000	0.000	1 (1 - 1)	0.000	0.000	1 (1 - 1)

Deprivation index	-0.006	0.006	0.99 (0.98 - 1)	-0.044	0.005	0.96 (0.95 - 0.96)
Deprivation index * Distance to an urban centre	0.002	0.000	1 (1 - 1)	0.001	0.000	1 (1 - 1)
Mean water deficit	-0.002	0.008	1 (0.99 - 1.01)	-0.011	0.007	0.99 (0.98 - 1)
Mean solar hours	0.045	0.006	1.05 (1.04 - 1.06)	0.037	0.005	1.04 (1.03 - 1.05)
Mean temperature	0.001	0.000	1 (1 - 1)	0.003	0.000	1 (1 - 1)
fixed effects for region (Manawatu-Wanganui as the referent category)						
Auckland	0.057	0.088	1.06 (0.92 - 1.23)	0.213	0.068	1.24 (1.11 - 1.39)
Bay of Plenty	0.954	0.085	2.6 (2.27 - 2.99)	0.453	0.068	1.57 (1.41 - 1.77)
Canterbury	1.085	0.078	2.96 (2.6 - 3.37)	1.469	0.061	4.35 (3.93 - 4.81)
Gisborne	0.300	0.118	1.35 (1.11 - 1.63)	0.386	0.096	1.47 (1.26 - 1.73)
Hawkes Bay	0.683	0.070	1.98 (1.76 - 2.23)	0.576	0.057	1.78 (1.62 - 1.96)
Marlborough	1.156	0.096	3.18 (2.71 - 3.72)	1.283	0.084	3.61 (3.14 - 4.14)
Nelson	0.627	0.215	1.87 (1.32 - 2.66)	0.870	0.182	2.39 (1.77 - 3.22)
Northland	0.026	0.116	1.03 (0.85 - 1.24)	-0.089	0.091	0.92 (0.79 - 1.07)
Otago	1.661	0.108	5.27 (4.41 - 6.29)	2.141	0.084	8.5 (7.41 - 9.81)
Southland	1.409	0.127	4.09 (3.31 - 5.05)	2.240	0.093	9.4 (8.07 - 10.95)
Taranaki	0.214	0.068	1.24 (1.11 - 1.39)	-0.226	0.054	0.8 (0.73 - 0.87)
Tasman	1.373	0.099	3.95 (3.35 - 4.63)	1.224	0.079	3.4 (2.99 - 3.88)
Waikato	0.209	0.073	1.23 (1.09 - 1.39)	-0.047	0.057	0.95 (0.87 - 1.05)
Wellington	0.625	0.065	1.87 (1.68 - 2.08)	0.664	0.049	1.94 (1.79 - 2.1)
West Coast	1.691	0.098	5.42 (4.6 - 6.37)	1.643	0.077	5.17 (4.56 - 5.87)

Figure 1: Map showing the 16 regions and the two main islands of New Zealand. Grey shaded areas represent the areas used to compute variograms in the North Island and South Island to evaluate spatial autocorrelation of model fit shown in this case by the rank probability score.

Figure 2: Variograms for the LSU model (A) and the cattle model (B) showing the spatial covariance in the farm-level Rank Probability Score (RPS) in areas presented in figure 1 in the North Island (NI) and the South Island (SI). The dotted lines represent the 5% and 95% of the prediction interval estimated by calculating a variogram for 100 random permutations of the real RPS values for each region. The distance at which the errors stop being correlated is read off where the solid lines cross into the 95% prediction intervals (dotted lines) for each island and for each model.

Figure 3: Boxplot of RPS for each of the 7 farm types in AgriBase for models predicting livestock units (light grey) and 5 farm classes in AgriBase for models predicting cattle (dark grey) on rural properties in New Zealand. This figure shows that in both models predictions for dry stock (contract grazing and dairy heifer rearing properties) were least accurate. Box widths are proportional to the square root of the number of observations present in each group.

Figure 4: Box plots of the RPS from 50 resampled fits of the model where region was removed from A, the LSU model and B the cattle model. The upper bound of each box is the 75th percentile, lower bound is the 25th percentile, the median is indicated within the box and the dotted whiskers terminate at the 95th and 5th percentiles. Outliers are shown as hollow circles.

Figure 5: Receiver operating characteristic (ROC) curve for a model predicting the absence of cattle (dotted line) and the absence of LSUs (shown with a solid line) on rural properties in New Zealand. The x-axis of the ROC curve shows $1 - (\text{Pred Animals} | \text{Data Animals})$, traditionally Sensitivity. The y-axis represented $(\text{Pred no animals} | \text{Data no animals})$, traditionally $1 - \text{Specificity}$.

References

Anonymous, 2013. Eurostat: Statistics Explained [WWW Document]. URL

<http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:LSU>

Bates, T.W., Thurmond, M.C., Carpenter, T.E., 2003. Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. *Am. J. Vet. Res.* 64, 195–204.

doi:10.2460/ajvr.2003.64.195

Bates, T.W., Thurmond, M.C., Carpenter, T.E., 2003. Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. *Am. J. Vet. Res.* 64, 195–204.

Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E.J., 2010. Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain. *Int. J. Infect. Dis.* 14,

e210–e215. doi:<http://dx.doi.org/10.1016/j.ijid.2009.05.003>

Bivand, R., Rudel, C., Pebesma, E., Hufthammer, K.O., 2014a. rgeos: Interface to Geometry Engine - Open

Source (GEOS).

Bivand, R., Rundel, C., Pebesma, E., Hufthammer, K.O., 2014b. rgdal: Bindings for the Geospatial Data Abstraction Library.

Blake, A., Sinclair, M.T., Sugiyarto, G., 2003. Quantifying the impact of foot and mouth disease on tourism and the UK economy. *Tour. Econ.* 9, 449–465.

Bradhurst, R.A., Roche, S.E., East, I.J., Kwan, P., Garner, M.G., 2015. A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. *Front. Environmental Sci.* 3, 1–20.
doi:10.3389/fenvs.2015.00017

Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* 78, 1–3.
doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Buhnerkempe, M.G., Tildesley, M.J., Lindström, T., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Werkman, M., Keeling, M.J., Wennergren, U., Webb, C.T., Lindstrom, T., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Werkman, M., Wennergren, W., 2014. The Impact of Movements and Animal Density on Continental Scale Cattle Disease Outbreaks in the United States. *PLoS One* 9, e91724.
doi:10.1371/journal.pone.0091724

Burdett, C.L., Kraus, B., Bjork, K.E., Miller, R.S., Oryang, D., Garza, S.J., 2014. Predicting the location and populations of individual livestock farms in the United States, in: 99th ESA Annual Meeting, Sacramento CA. p. <http://eco.confex.com/eco/2014/webprogram/Paper488>.

Burdett, C.L., Kraus, B.R., Garza, S.J., Miller, R.S., 2015. Simulating the Distribution of Individual Livestock Farms and Their Populations in the United States : An Example Using Domestic Swine (*Sus scrofa domesticus*) Farms. *PLoS One* 10, 1–21. doi:10.5061/dryad.46pb3

Carlin, B.P., Chib, S., 1995. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *J. R. Stat. Soc. Ser. B* 57, 473–484. doi:10.2307/2346151

Carpenter, T.E., O'Brien, J.M., Hagerman, A.D., McCarl, B. a, 2011. Epidemic and Economic Impacts of Delayed Detection of Foot-And-Mouth Disease: A Case Study of a Simulated Outbreak in California. *J. Vet. Diagnostic Investig.* 23, 26–33. doi:10.1177/104063871102300104

- Carpenter, T.E., Thurmond, M.C., Bates, T.W., 2004. A simulation model of intraherd transmission of foot and mouth disease with reference to disease spread before and after clinical diagnosis. *J. Vet. Diagn. Invest.* 16, 11–16. doi:10.1177/104063870401600103
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. *Am. Stat.* 49, 327–335. doi:10.2307/2684568
- Diggle, P., Ribeiro, P.J., 1998. Model-based Geostatistics. *J. R. Stat. Soc. Ser. C (Applied Stat.* 47, 299–350. doi:10.1007/978-0-387-48536-2
- Dohoo, I., Martin, W., Stryhn, H., 2009. *Veterinary Epidemiologic Research*. VER Inc., Charlottetown, Prince Edward Island, Canada.
- Dubé, C., Stevenson, M.A., Garner, M.G., Sanson, R.L., Corso, B.A., Harvey, N., Griffin, J., Wilesmith, J.W., Estrada, C., 2007. A comparison of predictions made by three simulation models of foot-and-mouth disease. *N. Z. Vet. J.* 55, 280–288. doi:10.1080/00480169.2007.36782 ER
- Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* 413, 542–548.
- Gates, M.C., Humphry, R.W., Gunn, G.J., Woolhouse, M.E.J., 2014. Not all cows are epidemiologically equal: quantifying the risks of bovine viral diarrhoea virus (BVDV) transmission through cattle movements. *Vet. Res.* 45, 110. doi:10.1186/s13567-014-0110-y
- Green, P.J., 1995. Reversible jump Markov chain monte carlo computation and Bayesian model determination. *Biometrika* 82, 711–732. doi:10.1093/biomet/82.4.711
- Gwyther, C.L., Williams, A.P., Golyshin, P.N., Edwards-Jones, G., Jones, D.L., 2011. The environmental and biosecurity characteristics of livestock carcass disposal methods: A review. *Waste Manag.* 31, 767–778. doi:http://dx.doi.org/10.1016/j.wasman.2010.12.005
- Halasa, T., Willeberg, P., Christiansen, L.E., Boklund, A., Alkhamis, M., Perez, a, Enøe, C., 2013. Decisions on control of foot-and-mouth disease informed using model predictions. *Prev. Vet. Med.* 112, 194–202. doi:http://dx.doi.org/10.1016/j.prevetmed.2013.09.003
- Haydon, D.T., Chase-Topping, M., Shaw, D.J., Matthews, L., Friar, J.K., Wilesmith, J., Woolhouse, M.E.J., 2003.

- The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. Biol. Sci.* 270, 121–127. doi:10.1098/rspb.2002.2191
- Honhold, N., Taylor, N.M., 2006. Data quality assessment: comparison of recorded and contemporary data for farm premises and stock numbers in Cumbria, 2001, *Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a meeting held at Exeter, UK, 29-31 March 2006.*
- Hugh-Jones, M.E., 1972. Epidemiological studies on the 1967-1968 Foot and Mouth Epidemic: Attack Rates and Cattle Density. *Res. Vet. Sci.* 13, 411–417.
- Jewell, C., Brown, R., 2014. Bayesian data assimilation provides rapid decision support for vector borne diseases Supporting Information Preparation of NAIT movement network 1–6.
- Jewell, C.P., 2016a. Bayesian ZIP model with covariates on the zero probability v1. doi:http://fhm-chicas-code.lancs.ac.uk/jewell/zipBayes
- Jewell, C.P., 2016b. Variogram v1. doi:http://fhm-chicas-code.lancs.ac.uk/jewell/GPUVariogram.git
- Jewell, C.P., Andel, M. Van, Vink, W.D., Mcfadden, A.M.J., 2015. Compatibility between livestock databases used for quantitative biosecurity response in New Zealand. *N. Z. Vet. J.* 169. doi:10.1080/00480169.2015.1117955
- Joung, H.K., Han, S.H., Park, S.-J., Jheong, W.-H., Ahn, T.S., Lee, J.-B., Jeong, Y.-S., Jang, K.L., Lee, G.-C., Rhee, O.-J., Park, J.-W., Paik, S.Y., 2013. Nationwide Surveillance for Pathogenic Microorganisms in Groundwater near Carcass Burials Constructed in South Korea in 2010. *Int. J. Environ. Res. Public Health* 10, 7126–7143. doi:10.3390/ijerph10127126
- Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., 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* (80-.). 294, 813–817. doi:10.1126/science.1065973
- Kitching, R.P., Thrusfield, M., Taylor, N.M., 2006. Use and abuse of mathematical models : an illustration from the 2001 foot and mouth. *Rev. Sci. Tech.* 25, 293–311.
- Lambert, D., 1992. Zero-Inflated Poisson Regression, With an Application To Defects in Manufacturing.

Technometrics 34, 1–14.

Mansley, L.M., 2004. The challenge of FMD control in the 2001 UK FMD epidemic, in: Madec, F. (Ed.), *Animal Production in Europe : The Way Forward in a Changing World “in-Between”* congress of the International Society for Animal Hygiene; Saint-Malo, France . pp. 345–350.

Marshal, J.P., Bleich, V.C., Krausman, P.R., Reed, M.L., Andrew, N.G., 2006. Factors affecting habitat use and distribution of desert mule deer in an arid environment. *Wildl. Soc. Bull.* 34, 609–619. doi:10.2193/0091-7648(2006)34[609:FAHUAD]2.0.CO;2

Mueller, T., Olson, K.A., Fuller, T.K., Schaller, G.B., Murray, M.G., Leimgruber, P., 2008. In search of forage: predicting dynamic habitats of Mongolian gazelles using satellite-based estimates of vegetation productivity. *J. Appl. Ecol.* 45, 649–658. doi:10.1111/j.1365-2664.2007.01371.x

Mullahy, J., 1986. Specification and testing of some modified count data models. *J. Econom.* 33, 341–365. doi:http://dx.doi.org/10.1016/0304-4076(86)90002-3

Muroga, N., Kobayashi, S., Nishida, T., Hayama, Y., Kawano, T., Yamamoto, T., 2013. Risk factors for the transmission of foot-and-mouth disease during the 2010 outbreak in Japan : a case – control study. *BMC Vet. Res.* 9, 1–9.

Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. THE INTERPRETATION OF SPECTRAL VEGETATION INDEXES. *IEEE Trans. Geosci. Remote Sens.* 33, 481–486. doi:10.1109/36.377948

Paiba, G.A., Roberts, S.R., Houston, C.W., Williams, E.C., Smith, L.H., Gibbens, J.C., Holdship, S., Lysons, R., 2007. UK surveillance: Provision of quality assured information from combined datasets. *Prev. Vet. Med. Lead Pap. Elev. Symp. Int. Soc. Vet. Epidemiol. Econ. (ISVEE), Cairns, Aust.* 81, 117–134. doi:http://dx.doi.org/10.1016/j.prevetmed.2007.04.006

Paine, S., Mercer, G.N., Kelly, P.M., Bandaranayake, D., Baker, M.G., Huang, Q.S., Mackereth, G., 2010. Transmissibility of 2009 pandemic influenza A (H1N1) in New Zealand : effective reproduction number and influence of age , ethnicity and importations 1–9.

Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J., O’Brien, J., 2014. *sp: Classes and Methods for Spatial Data.*

Pink, B., 2011. New Zealand: An Urban/Rural Profile

http://www.stats.govt.nz/browse_for_stats/Maps_and_geography/Geographic-areas/urban-rural-profile.aspx. Statistics New Zealand, Wellington, New Zealand.

Porphyre, T., Auty, H.K., Tildesley, M.J., Gunn, G.J., Woolhouse, M.E.J., 2013a. Vaccination against Foot-And-Mouth Disease: Do Initial Conditions Affect Its Benefit? *PLoS One* 8, e77616.

doi:10.1371/journal.pone.0077616

Porphyre, T., McKenzie, J., Byrom, A.E., Nugent, G., Shepherd, J., Yockney, I., 2013b. Spatial prediction of brushtail possum (*Trichosurus vulpecula*) distribution using a combination of remotely sensed and field-observed environmental data. *Wildl. Res.* 40, 578. doi:10.1071/WR13028

Reed, B.C., Brown, J.F., van der Zee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. MEASURING PHENOLOGICAL VARIABILITY FROM SATELLITE IMAGERY. *J. Veg. Sci.* 5, 703–714. doi:10.2307/3235884

Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D’Aietti, L., Hay, S.I., Gilbert, M., 2014. Mapping the Global Distribution of Livestock. *PLoS One* 9, e96084.

Roche, S.E., Garner, M.G., Sanson, R.L., Cook, C., Birch, C., Backer, J.A., Dubé, C., Patyk, K.A., Stevenson, M.A., Yu, Z.D., Rawdon, T.G., Gauntlett, A.F., 2015. Evaluating vaccination strategies to control foot-and-mouth disease: a model comparison study. *Epidemiol. Infect.* 143, 1256–1275. doi:10.1017/S0950268814001927

Sanson, R., Pearson, A., 1997. Agribase - a national spatial farm database, in: Eighth IVSEE, Paris. p. 12.12.1-12.16.3.

Sanson, R., Rawdon, T., Owen, K., Yu, Z., 2013. Evaluating the benefits of vaccination for foot-and-mouth disease as an adjunct to a standard stamping out policy.

Sanson, R.L., Harvey, N., Garner, M.G., Stevenson, M.A., 2011. Foot and mouth disease model verification and “relative validation.” *Rev. Sci. Tech. L’Office Int. Des Epizoot.* 30, 527–540.

Sarandopoulos, J., 2015. Early Predictors of the Size and Duration of Foot-and-Mouth Disease Epidemics. The University of Melbourne.

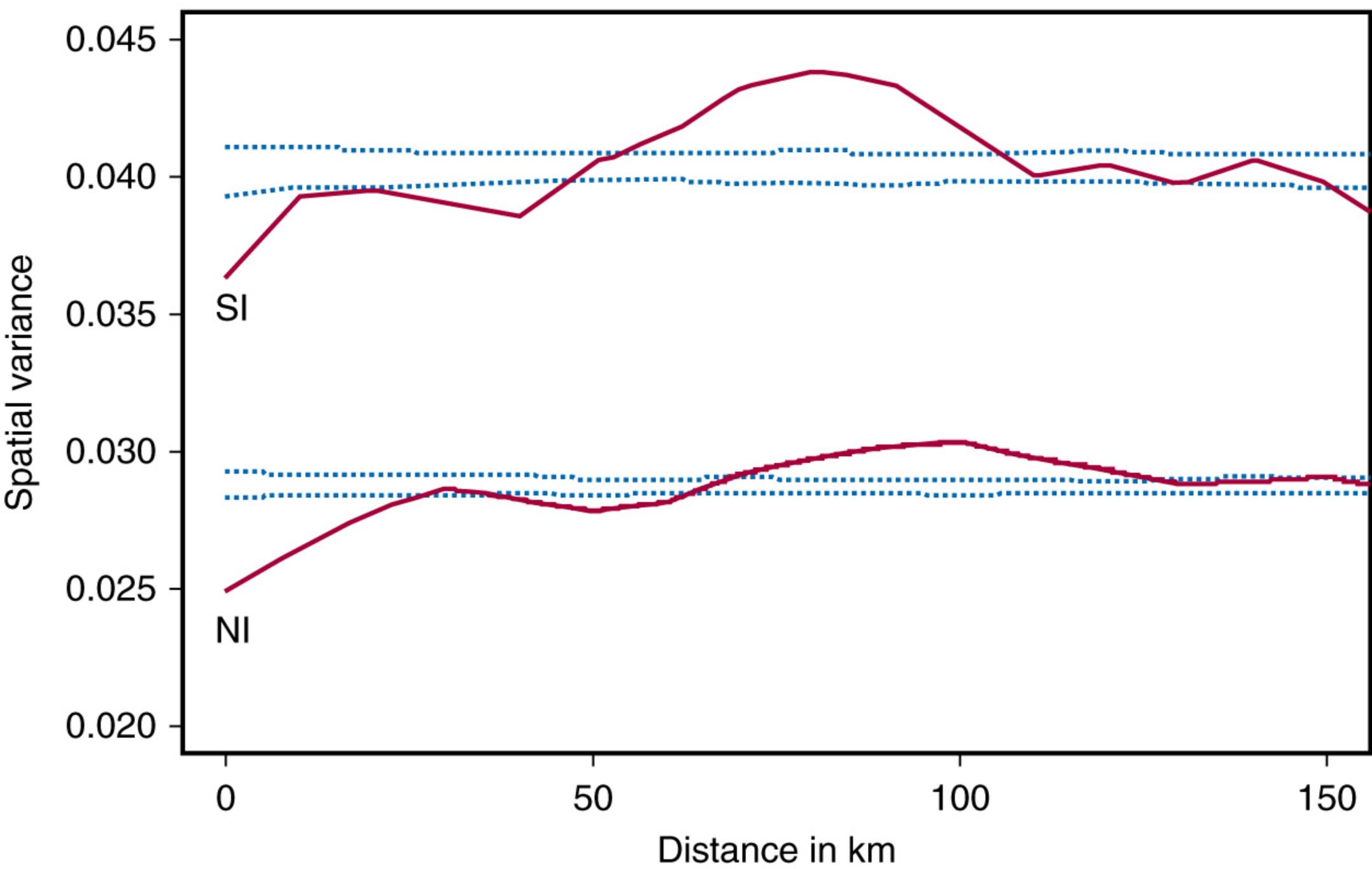
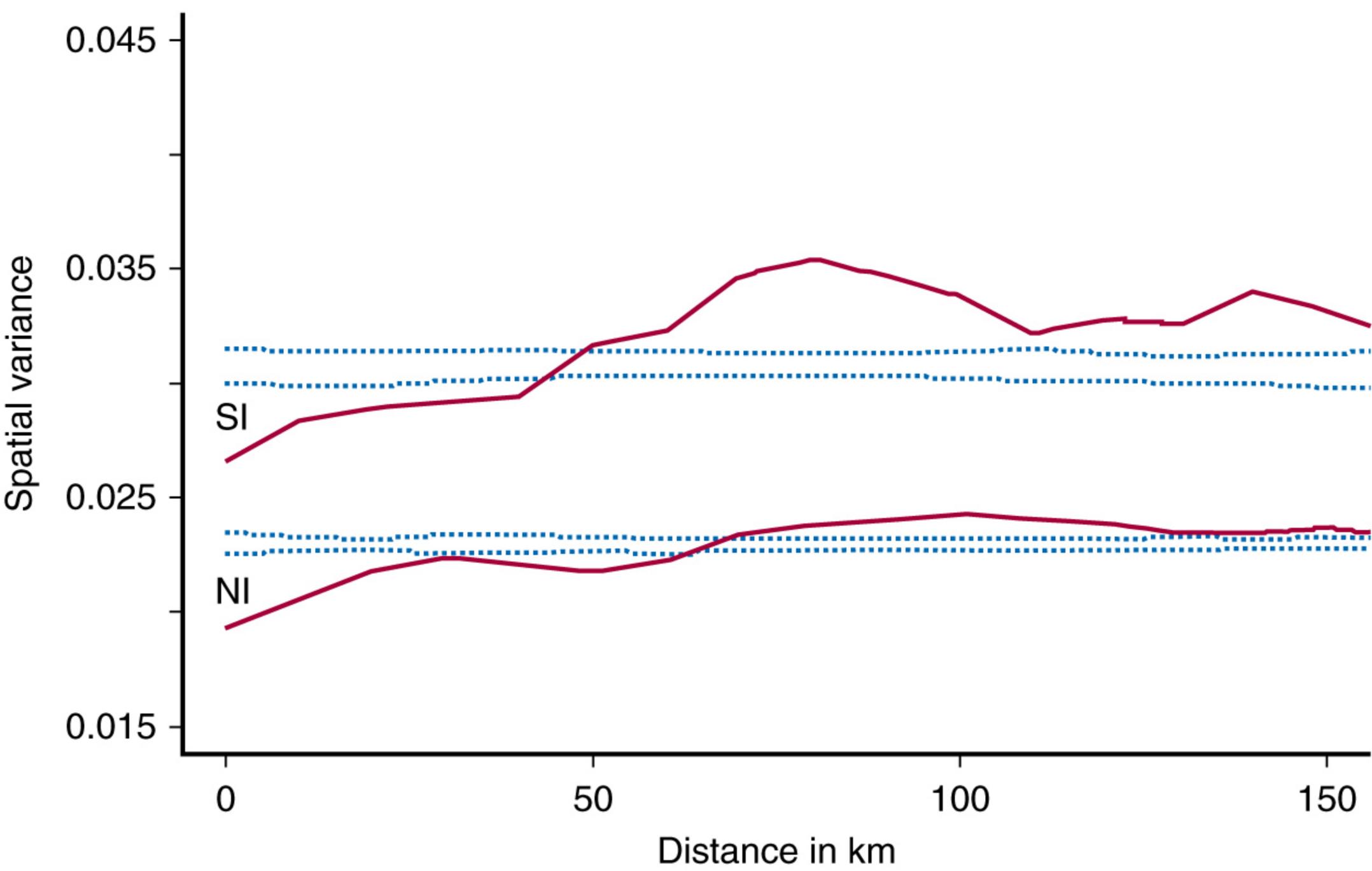
Savill, N.J., Shaw, D.J., Deardon, R., Tildesley, M.J., Keeling, M.J., Woolhouse, M.E.J., Brooks, S.P., Grenfell, B.T., 2007. Effect of data quality on estimates of farm infectiousness trends in the UK 2001 foot-and-mouth

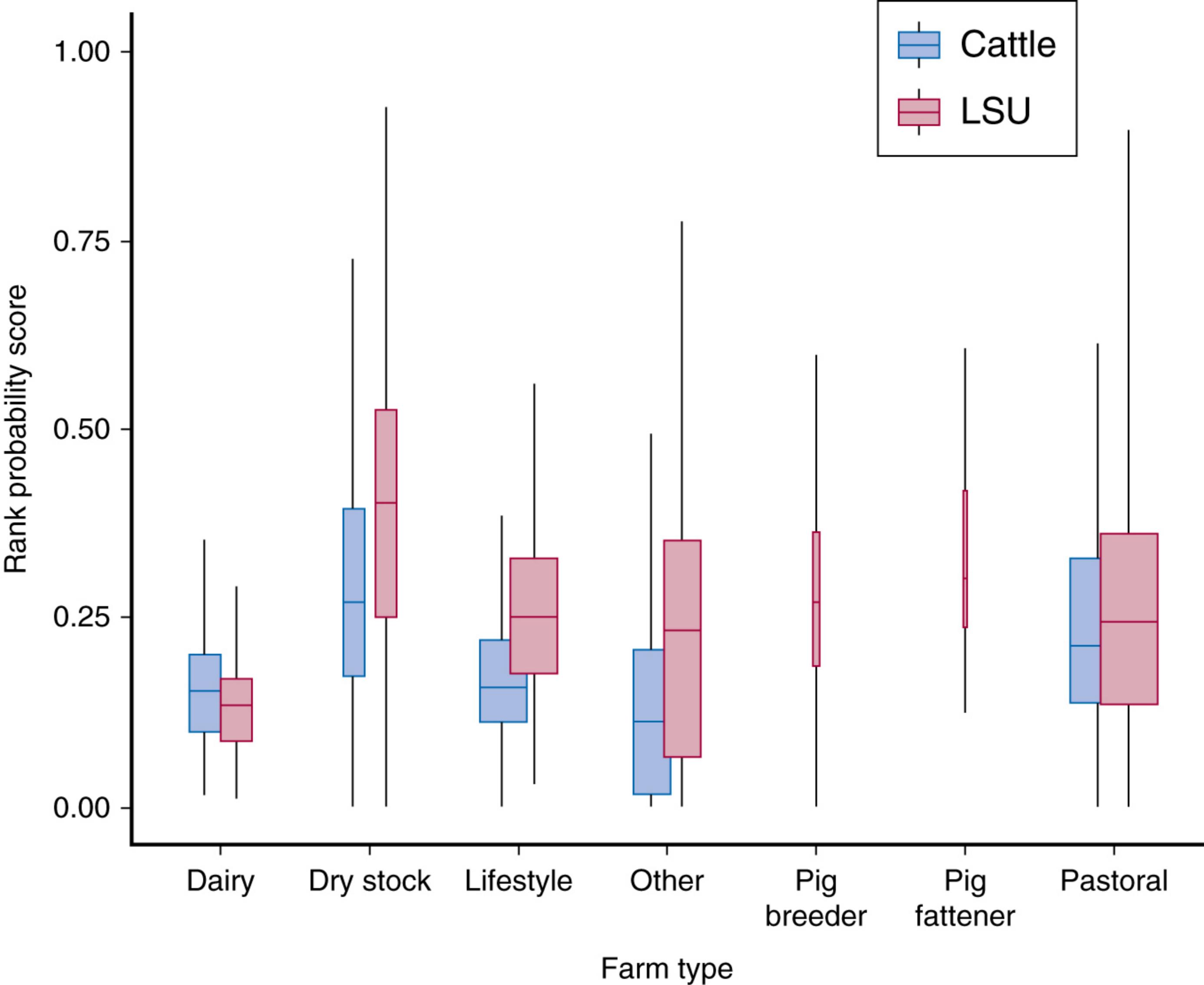
- disease epidemic. *J. R. Soc. Interface* 4, 235–241.
- Stage, A.R., 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. *For. Sci.* 22, 457–460.
- Stage, A.R., Salas, C., 2007. Interactions of elevation, aspect, and slope in models of forest species composition and productivity. *For. Sci.* 53, 486–492.
- Team, R.C., 2013. R: A language and environment for statistical computing. .
- Thrusfield, M., Mansley, L.M., Dunlop, P., Taylor, J., Pawson, A., Stringer, L., 2005. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 1: Characteristics and control. *Vet. Rec.* 156, 229–252.
- Tildesley, M.J., Deardon, R., Savill, N.J., Bessell, P.R., Brooks, S.P., Woolhouse, M.E.J., Grenfell, B.T., Keeling, M.J., 2008. Accuracy of models for the 2001 foot-and-mouth epidemic. *Proc. Biol. Sci.* 275, 1459–1468. doi:10.1098/rspb.2008.0006
- Tildesley, M.J., Ryan, S.J., 2012. Disease Prevention versus Data Privacy: Using Landcover Maps to Inform Spatial Epidemic Models. *PLoS Comput. Biol.* 8, 1–13. doi:10.1371/journal.pcbi.1002723
- Tomassen, F.H.M., de Koeijer, A., Mourits, M.C.M., Dekker, A., Bouma, A., Huirne, R.B.M., 2002. A decision-tree to optimise control measures during the early stage of a foot-and-mouth disease epidemic. *Prev. Vet. Med.* 54, 301–324. doi:http://dx.doi.org/10.1016/S0167-5877(02)00053-3
- Tveraa, T., Stien, A., Bardsen, B.-J., Fauchald, P., 2013. Population Densities, Vegetation Green-Up, and Plant Productivity: Impacts on Reproductive Success and Juvenile Body Mass in Reindeer. *PLoS One* 8. doi:10.1371/journal.pone.0056450
- Wada, M., Carpenter, T.E., van Andel, M., 2016. The importance of accurate spatial livestock data to inform decision making: a case study of foot-and-mouth disease control in New Zealand, in: *Geovet. Valdivia, Chile*.
- West, M., 2014. Bayesian Forecasting, in: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd. doi:10.1002/9781118445112.stat00219
- Woolhouse, M.E.J., 2003. Foot-and-mouth disease in the UK: What should we do next time? *Joural Appl.*

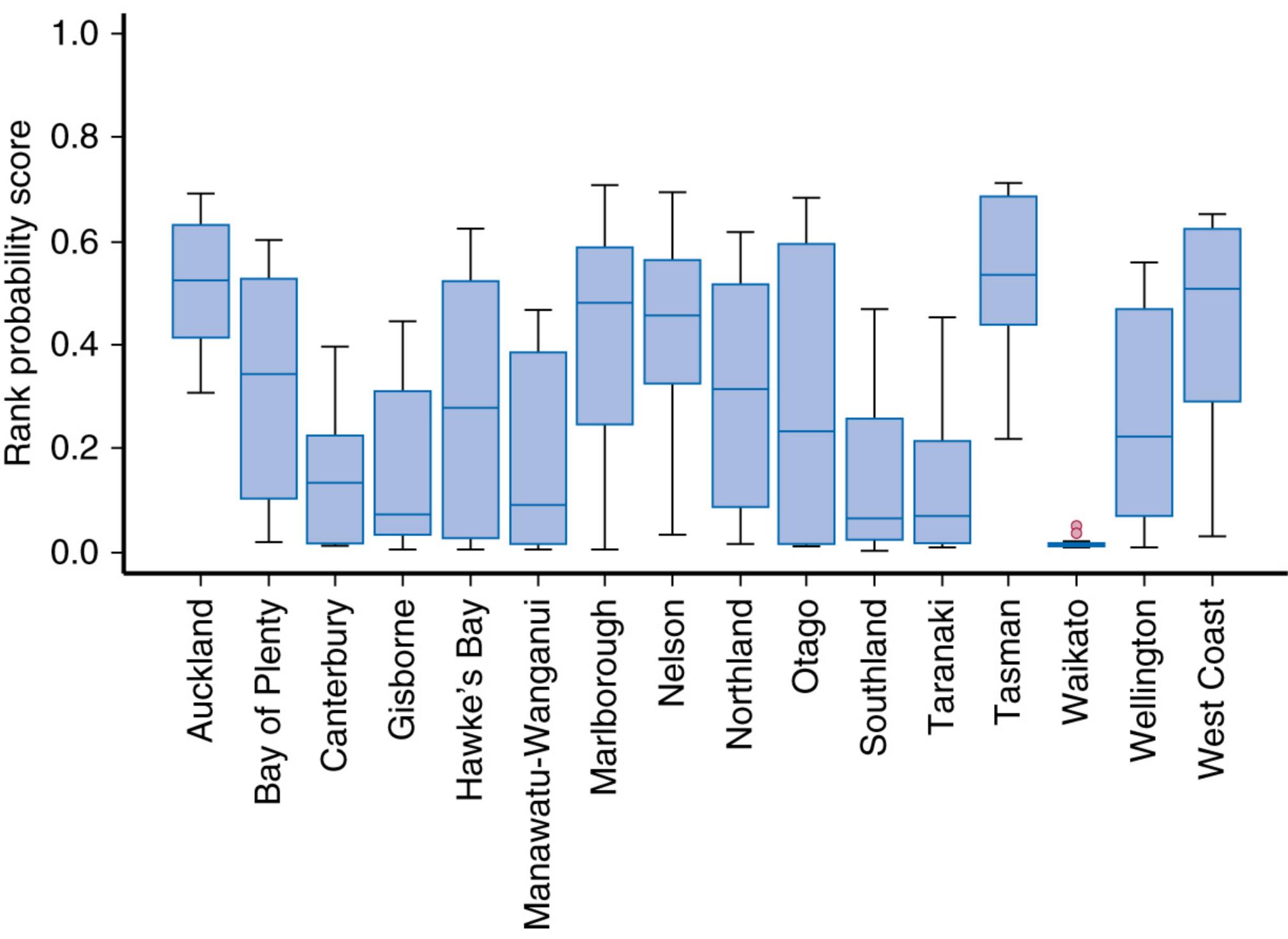
Microbiol. 94, 126–130.

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577. doi:ROC; Receiver-Operating Characteristic; SDT; Signal Detection Theory



A**B**



A**B**