



Report Cover Page

CEBRA Project		
1402C and 1502C		
Title		
National-level farm demographic data for preparedness of highly-infectious livestock disease epidemics. Review of data sources in New Zealand, approach to modelling populations and the effect of population uncertainty on disease modelling.		
Author(s) / Address (es)		
Mary van Andel (MPI), Tracey Hollings (CEBRA), Andrew Robinson (CEBRA), Chris Jewell (Previously Massey University now Lancaster University), Mark Burgman (CEBRA), Daan Vink (MPI), Katrin Sattler (MPI), Masako Wada (Massey EpiCentre), Tim Carpenter (Massey EpiCentre), Richard Bradhurst (ANU), Graeme Garner (DAWR)		
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)		
Final Project Report		
Summary		
<p>Isolation and strict biosecurity measures implemented by Australia and New Zealand have prevented the incursion of many organisms of biosecurity concern. The agricultural industry is a key part of the economy for both countries, and preventing the arrival of diseases of concern, including foot-and-mouth (FMD) disease, is crucial to maintaining access to international markets, reputation, and protecting the economy and industry. Early detection of, and a rapid, effective response to such diseases have a large impact on limiting the economic damage caused by epidemics.</p> <p>An acknowledged weakness of biosecurity preparedness and response to agricultural diseases in both countries is the lack of a single source of accurate, up-to-date farm livestock demographics information. This report reviews the use of animal counts for investigation of, preparedness for and response to exotic animal disease outbreaks, and analyses the available New Zealand datasets in depth. Gaps and weaknesses in the current data landscape are documented. The project objectives then focus on developing methodologies to estimate national-level farm demographic data and assess the use of modelled and inaccurate data in disease simulation models. There are nine key deliverables outlined in the report which were carried out over the two year duration of the project.</p>		
CEBRA Use only	Received By:	Date:
	CEBRA / SAC Approval:	Date:
	DAWR Endorsement: () Yes () No	Date:

3 **National-level farm demographic data for preparedness of**
4 **highly-infectious livestock disease epidemics**

5 Review of data sources in New Zealand, approach to modelling populations
6 and the effect of uncertainty on disease modelling.

7
8 **Final Project Report**

9 30 June 2016

10
11 **CEBRA Project No. 1402C & 1502C**

12 Investigators:

13 Mary van Andel (MPI);

14 Tracey Hollings (CEBRA);

15 Andrew Robinson (CEBRA);

16 Chris Jewell (Lancaster University, formerly Massey University);

17 Mark Burgman (CEBRA);

18 Daan Vink (MPI);

19 Katrin Sattler (MPI);

20 Masako Wada (Massey EpiCentre);

21 Richard Bradhurst (ANU);

22 Graeme Garner (DAWR);

23 Tim Carpenter (Massey EpiCentre)

24

25

26 **Acknowledgements**

27 This report is a product of the Centre of Excellence for Biosecurity Risk Analysis (CEBRA). In
28 preparing this report, the authors acknowledge the financial and other support provided by the
29 Department of Agriculture and Water Resources (DAWR), Ministry for Primary Industries (MPI) and
30 the University of Melbourne. Thanks are also due toASUREQuality for the use of AgriBase data for
31 this project; to Hannes Calitz of the Institute of Fundamental Sciences at Massey University for
32 assistance with grid computing; and to Carolyn Gates and Joanna McKenzie of the Institute of
33 Veterinary, Animal and Biomedical Sciences at Massey University who provided input on early drafts
34 of several chapters of this work. Special thanks are due to Paul Bingham, manager of the
35 Surveillance and Incursion Investigation Team (Animals and Marine) in the Operations branch of
36 MPI.

37

38

39

40

41

42 **Table of contents**

43 [Acknowledgements](#)

44 [Table of contents](#)

45 [List of Tables](#)

46 [List of Figures](#)

47 [1 Executive Summary](#)

48 [2 Introduction](#)

49 [3 Deliverable 1: A review of the uses of farm-level animal count](#)
50 [information and the means by which the value of these data can be described](#)

51 [4 Deliverable 2: Review of Remotely Sensed Data Techniques](#)

52 [5 Deliverable 3: Overview of New Zealand national level data sources](#)
53 [available at the farm-level](#)

54 [6 Deliverable 4: Machine Learning Models](#)

55 [7 Deliverable 5: Report on Bayesian Zero Inflated Poisson Models to](#)
56 [predict animal counts – Predicting the spatial distribution of domestic animal](#)
57 [populations from incomplete demographic data](#)

58 [8 Deliverable 6: Model validation against external data sources](#)

59 [9 Deliverable 7: Use of model outputs in the Australian Animal Disease](#)
60 [Simulation model \(AADIS\)](#)

61 [10 Deliverable 8: Comparison of two national livestock databases for](#)
62 [informing decision making for a foot and mouth disease outbreak in New](#)
63 [Zealand](#)

64 [11 Deliverable 9: Capacity development and information transfer](#)

65 [12 Recommendations](#)

66 [13 References](#)

67 [14 Appendices](#)

68

69

70 List of Tables

71 [Table 4.1: A list of representative peer reviewed articles applying automated and semi-automated](#)
72 [techniques to remotely sensed imagery to estimate the abundance of mammals](#)

73 [Table 5.1: Number of premises recorded on the Farms on Line \(FOL\) and AgriBase databases with](#)
74 [the reported number of animals per property for different animal species, on 16 September 2014](#)
75 [and 1 September 2014, respectively.](#)

76 [Table 5.2. Agreement between the AgriBase, Farms on line \(FOL\) and NAIT databases in terms of](#)
77 [recorded presence of animals of different species on properties in the databases, calculated as](#)
78 [percentage agreement and Cohen's kappa statistic \(K\)^b.](#)

79 [Table 6.1: Species equivalence values to an adult dairy cow that were used for analysis \(Eurostat,](#)
80 [2013; The Scottish Government, 2014\)](#)

81 [Table 6.2: Summary of the source and calculation of predictor variables used for modelling.](#)

82 [Table 6.3: The prediction root mean squared error \(RMSPE\) for LSU and cattle using the withheld](#)
83 [results for the two spatial stratifications \(regional and quarantine zone\) for both individual farms](#)
84 [within regions and to 3km quarantine zones. Table also shows the mean count of LSU and cattle per](#)
85 [farm and the standard deviation for all farms in brackets. KNN models could not be assessed in](#)
86 [quarantine zone stratification due to the categorical co-variate of 'region'. Results for Quarantine](#)
87 [zone predictions are in 1,000s of LSU and cattle. Full results for all models are shown in the](#)
88 [Supplementary Materials.](#)

89 [Table 6.4: Comparison of BRT and RF by the RMSPE for LSU across different size farms. The number](#)
90 [of farms in the regional stratification and quarantine stratification differ because not all farms were](#)
91 [selected as target data over the ten simulations in the quarantine zone stratification due to the bias](#)
92 [arising from farm size when selecting centroid farms.](#)

93 [Table 6.5: The LSU prediction root mean squared error \(RMSPE\) using the withheld results for the](#)
94 [regional spatial stratification for both individual farms within regions and to 3km quarantine zones.](#)
95 [Table also shows the mean number of LSU per farm in the region, the standard deviation for all](#)
96 [farms in brackets, and the total number of farms in each region. Results for Quarantine zone](#)
97 [predictions are in 1,000s of LSU.](#)

98 [Table 6.6: The cattle prediction root mean squared error \(RMSPE\) using the withheld results for the](#)
99 [regional spatial stratification for both individual farms within regions and to 3km quarantine zones.](#)
100 [Table also shows the mean number of cattle per farm in the region, the standard deviation for all](#)
101 [farms in brackets, and the total number of farms in each region. Results for Quarantine zone](#)
102 [predictions are in 1,000s of cattle.](#)

103 [**Table 7.1:** Species equivalence values used in the models constructed to predict counts of livestock](#)
104 [units \(LSUs\) on rural properties in New Zealand. These values are calculated following the European](#)
105 [Union definition of individual species requirements \(Eurostat, 2013\)](#)

106 [**Table 7.2:** Predictor variables included in Bayesian ZIP model predicting counts of cattle and counts](#)
107 [of livestock units \(LSU\) relative to dairy cattle on New Zealand farms. * Predictor variables included](#)
108 [in the final models.](#)

109 [**Table 7.3:** Mean and standard deviation of rank probability scores \(RPS\) for a model to predict](#)
110 [livestock units \(LSUs\) and cattle on New Zealand farms. RPS are reported for 2 models where](#)
111 [training data were withheld first at a 3km zone level and subsequently at a regional level. Results are](#)
112 [then examined at the model level and at the area level for each of these instances.](#)

113 [**Table 8.1:** Descriptive statistics for each mesh block in the Agricultural Production Survey \(APS\)](#)

114 [**Table 8.2:** Count of mesh blocks per region](#)

115 [**Table 8.3:** Counts of species present in each region according to the APS data extrapolated from the](#)
116 [2015 survey.](#)

117 [**Table 8.4:** Descriptive measures for numbers of AgriBase farms overlapping Statistics New Zealand](#)
118 [mesh blocks.](#)

119 [**Table 8.5:** Mean absolute error calculated for the difference between 1\) 2 observed data sources](#)
120 [namely Statistics New Zealand mesh block level agricultural production survey \(APS-mb\) data and](#)
121 [AgriBase data aggregated to the mesh block level for GST registered properties; 2\) Aggregated](#)
122 [results of the Random Forest \(RF\) model using AgriBase predictor variables compared with APS-mb](#)
123 [for the cattle model and the livestock units \(LSU\) model; 3\) Aggregated results of the Bayesian Zero](#)
124 [Inflated Poisson \(B-ZIP\) model using AgriBase predictor variables compared with APS-mb data for](#)
125 [the cattle model and the LSU model; 4\) Aggregated results of the RF model using FarmsOnLine \(FOL\)](#)
126 [predictor variables compared with APS-mb for the cattle model and LSU model; 5\) Aggregated](#)
127 [results of the B-ZIP model using FOL predictor variables compared with APS-mb for the cattle model](#)
128 [and LSU model.](#)

129 [**Table 8.6:** Mean rank probability scores \(RPS\) for a Bayesian ZIP model built with AgriBase data and](#)
130 [tested against mesh block level Agricultural Production Survey data and farm-level AgriBase data.](#)

131 [**Table 8.7:** Recorded farm type in FOL for each farm matched in the APS data.](#)

132 [**Table 8.8:** Mean absolute errors \(MAE\) for the Random Forest and Bayesian ZIP models as well rank](#)
133 [probability scores for the Bayesian ZIP models for the livestock unit and cattle prediction models](#)
134 [when compared against APS data collected at the farm-level.](#)

135 [**Table 8.9:** Mean absolute errors \(MAE\) for the Random Forest and Bayesian ZIP models as well rank](#)
136 [probability scores for the Bayesian ZIP models for the livestock unit and cattle prediction models](#)

137 [when compared against data collected at the farm-level by surveying veterinarians regarding their](#)
138 [knowledge of animal populations in their practice areas.](#)

139 [**Table 9.1:** Relationship between the farm types used in the New Zealand Standard Model in](#)
140 [InterSpread Plus and how these relate to farm types used to parameterise the Australian Animal](#)
141 [Disease Spread Model for New Zealand in this experiment.](#)

142 [**Table 9.2:** Counts of herds present in the herd file constructed for use in the Australian Animal](#)
143 [Disease Spread Model showing the number of herds of each type stratified by the farm type they are](#)
144 [present on.](#)

145 [**Table 9.3:** Percentages of new farm infections attributable to different spread pathways](#)

146 [**Table 9.4:** Relative infectivity and susceptibility weights for each of the species included in the](#)
147 [Australian Animal Disease spread model \(AADIS\).](#)

148 [**Table 9.5:** Tabulated results of four different disease spread models built in 2 different platforms \(the](#)
149 [Australian Animal Disease Spread Model – AADIS and InterSpread Plus – ISP\). The aim of this table is](#)
150 [to show the data from figure 9.3 in tabular form and to provide a reference point where the chosen](#)
151 [value of 0.55 for infectivity and susceptibility can be compared against results from the same silent](#)
152 [spread period modelled in Interpread Plus using the New Zealand Standard Model. Note that for the](#)
153 [“real” AADIS herd dataset, 3 additional datapoints were calculated \(0.55, 0.6, 0.65\) – these are](#)
154 [marked with a *.](#)

155 [**Table 9.6:** The results of t tests for comparing the median number of infected farms between a](#)
156 [model using a point estimate of herd size with a model using a beta pert distribution for farm size](#)
157 [after 50 days of spread.](#)

158 [**Table 10.1:** Summary statistics of the seven New Zealand districts selected for foot-and-mouth](#)
159 [disease simulation by AgriBase \(AB\) and FarmsOnLine \(FOL\) livestock databases](#)

160 [**Table 10.2:** Number of iterations \(and percentage\) where emergency vaccination was not required](#)
161 [to eradicate a simulated foot and mouth disease epidemic in the seven New Zealand districts](#)

162 [**Table 14.1:** MAE values for LSUs and individual species for the 13 selection variations described](#)
163 [above. The option \(option3\) that minimised the most MAE values for the individual species and for](#)
164 [LSUs was chosen as the most appropriate selection criteria. The final row represents mean error \(not](#)
165 [absolute error\) which was also taken into account.](#)

166 [**Table 14.2:** Input fields from the Agricultural Production \(APS\) Census data used in an analysis to](#)
167 [match the APS and FarmsOnLine datasets for the purposes of model validation.](#)

168 [**Table 14.3:** Input fields from FarmsOnLine \(FOL\) the Agricultural Production \(APS\) Census data used](#)
169 [in an analysis to match the APS and FOL datasets for the purposes of model validation.](#)

170 [Table 14.4: Workflow linking Agricultural Production \(APS\) Census data with target livestock present](#)
171 [to FOL farms.](#)

172 [Table 14.5: Fields for the matched records for analysis matching FarmsOnLine \(FOL\) with Agricultural](#)
173 [Production \(APS\) Census data](#)

174 [Table 14.6: Information on regional variance in matching success between the FarmsOnLine and](#)
175 [Agricultural Production Census information.](#)

176 [Table 14.7: Summary of survey results](#)

177 [Table 14.8: Counts of animal movements by year](#)

178 [Table 14.9: Counts of production types by year which are unspecified.](#)

179 [Table 14.10: Break down of those NAIT number types that had “None” or blank in 2015 as their](#)
180 [production types](#)

181 [Table 14.11: Counts of movements to and from production types of interest by year.](#)

182 [Table 14.12: Values used in AADIS for movements onto dairy, beef and deer farms. Values are](#)
183 [frequencies of animal movements onto farms per day.](#)

184 [Table 14.13: Values used in AADIS for movements off dairy, beef and deer farms. Values are](#)
185 [frequencies of animal movements off farms per day.](#)

186 [Table 14.14: Movement frequencies per day from farms to each of the PICA types and production](#)
187 [types by season](#)

188 [Table 14.15: Movement frequencies per day to farms from farms and sale yards by each of the](#)
189 [production types by season](#)

190 [Table 14.16: AADIS parameters from NAIT data. Distances shown in km.](#)

191 [Table 14.17: AADIS distance parameters in km for each type of movement that involves beef, cattle](#)
192 [and deer farms for each season.](#)

193 [Table 14.18: Counts of animal movements stratified by season when more than 150 animals and](#)
194 [more than 250 animals were moved at one event.](#)

195 [Table 14.19: Descriptive statistics of movement distances for consignments of more than 250](#)
196 [animals in each season](#)

197 [Table 14.20: Mean animal counts per premises by species, enterprise type, and district, based on](#)
198 [AgriBase data](#)

199 [Table 14.21: New Zealand Standard Model \(NZSM\): foot-and-mouth disease \(FMD\) simulation](#)
200 [parameters](#)

201 [Table 14.22: Estimated economic parameters for estimation of the economic impacts of a foot-and-](#)
202 [mouth disease \(FMD\) epidemic](#)

203 [**Table 14.23:** Estimated slaughtering capacity by area for subsequent culling of vaccinated animals](#)
204 [for control of a foot-and-mouth disease outbreak by vaccinate-to-die](#)
205
206

207 List of Figures

208 [Figure 4.1:: Using Particle Tracker \(Sbalzarini and Koumoutsakos, 2005\) in ImageJ designed for video](#)
209 [imaging in cell biology to detect sheep; a\) original image of farming area in NZ at 0.4m resolution; b\)](#)
210 [particle tracker program for automated detection of 'particles' to detect sheep; c\) a representative](#)
211 [area of peri-urban small scale farming in NZ at 0.4m resolution; d\) This relatively less homogenous](#)
212 [landscape has less success with particle tracker plugin for detecting sheep. Images obtained from](#)
213 <http://data.linz.govt.nz>

214 [Figure 5.1: Schematic diagram of the relationships between existing and possible national level](#)
215 [sources of animal counts, how data sets can currently be joined and challenges and strengths](#)
216 [associated with the data sources.](#)

217 [Figure 5.2: The spatial distribution of properties in New Zealand recorded in the \(a\) AgriBase and \(b\)](#)
218 [Farms on Line databases reporting the presence of cattle, sheep, goats, deer and pigs, on 1](#)
219 [September 2014 and 16 September 2014, respectively, calculated using spatial kernel density](#)
220 [estimation with a 37 km bandwidth. The scale bars represent the density of farm centroids per km²](#)
221 [for the different regions.](#)

222 [Figure 5.3: The density of farm centroids per km² recorded in the AgriBase database as a proportion](#)
223 [of that in the Farms on Line databases for \(a\) cattle, \(b\) pigs, \(c\) deer, and \(d\) small ruminants \(sheep](#)
224 [and goats\). Note that the scale for small ruminants is different to the other three species.](#)

225 [Figure 6.1: Map of Livestock Units \(LSU\) per hectare across New Zealand from Agribase™. White](#)
226 [areas represent masked areas or areas of missing data. Areas of more than 10 LSU/cattle per hectare](#)
227 [are represented in dark green. LSU represented per hectare is used for display purposes only.](#)

228 [Figure 6.2: Scatterplot of the observed against the predicted a\) farm-level LSU and b\) total](#)
229 [quarantine zone LSU for the regional spatial stratification, plotted by region. Total quarantine zone](#)
230 [LSU was calculated by summing the observed and predicted values for all farms within the 3km](#)
231 [quarantine zone. Cattle results are not shown as plots they are a subset of LSU.](#)

232 [Figure 6.3: Scatterplot of the observed against the predicted a\) farm-level LSU and b\) total](#)
233 [quarantine level LSU for the quarantine zone spatial stratification, by region. Total quarantine level](#)
234 [LSU was calculated by summing the observed and predicted values for all farms within the 3km](#)
235 [quarantine zone. Cattle results are not shown as they are a subset of LSU.](#)

236 [Figure 6.4: Relative influence of predictor variables from Random Forest models for LSU and cattle.](#)
237 [Plots show the increase in MSE of predictions if the variable of interest were randomly permuted.](#)

238 [Figure 7.1: Map showing the 16 regions and the two main islands of New Zealand. Grey shaded](#)
239 [areas represent the areas used to compute variograms in the North Island and South Island when](#)

240 [the spatial autocorrelation of model fit \(shown in this case by the rank probability score\) is](#)
241 [presented.](#)

242 [Figure 7.2: Box plots of the Rank Probability Scores \(RPS\) from 50 resampled fits of our model where](#)
243 [region was removed from A, the LSU model and B the cattle model. Data were divided into a training](#)
244 [\(6 regions from the North Island and 5 from the South Island\) and target dataset \(the remaining 5](#)
245 [regions\) by a stratified random sample of for each of the 50 resampled fits. In each instance RPS was](#)
246 [calculated at the region level. All 50 fits were then aggregated and the RPS for each region averaged](#)
247 [over all 50 fits to produce the boxplots which show the mean RPS for each region.](#)

248 [Figure 7.3: Boxplot of rank probability scores \(RPS\) for each of the 7 farm classes in AgriBase for](#)
249 [models predicting livestock units \(A\) and cattle \(B\) on rural properties in New Zealand. This figure](#)
250 [shows that in both models predictions for grazing and dry cow properties are the least accurately](#)
251 [represented. Box widths are proportional to the square root of the number of observations present](#)
252 [in each group.](#)

253 [Figure 7.4: Variograms for the LSU model \(A\) and the cattle model \(B\) showing the spatial covariance](#)
254 [in the farm-level Rank Probability Score \(RPS\) in an area chosen in the North Island \(NI\) and the](#)
255 [South Island \(SI\). The dotted lines represent the 5% and 95% of the prediction interval estimated by](#)
256 [calculating a variogram for 100 random permutations of the real RPS values for each region.](#)

257 [Figure 7.5: Receiver operating characteristic \(ROC\) curve for a model predicting the absence of cattle](#)
258 [\(dotted line\) and the absence of livestock units \(LSUs shown with a solid line\) on rural properties in](#)
259 [New Zealand.](#)

260 [Figure 8.1: Rank probability score \(RPS\) at the mesh block level. Livestock unit model predictions at](#)
261 [the mesh block level are compared with Statistics New Zealand Agricultural Production Survey data.](#)

262 [Figure 8.2: Rank probability score \(RPS\) at the mesh block level. Cattle model predictions at the](#)
263 [mesh block level are compared with Statistics New Zealand Agricultural Production Survey data.](#)

264 [Figure 8.3: Regional distribution of Agricultural Production Survey \(APS\) analysis sample and](#)
265 [achieved matches per territorial authorities as per 2013. 492 entries of the APS analysis sample](#)
266 [\(N = 41182\) could not be unambiguously attributed to an area unit due to incorrect/ insufficient](#)
267 [location information in the APS dataset and are consequently not represented in the figures.](#)

268 [Territorial authority boundaries were sourced from Statistics New Zealand, licensed under CC-BY.](#)

269 [Figure 8.4: Mean rank probability score \(RPS\) for each farm in the APS validation data stratified by](#)
270 [region.](#)

271 [Figure 9.1: Farms in AADIS with multiple herds present on them. Farms had between 1 and 5 herds](#)
272 [on them \(inclusive\) with the majority of farms 43365 of the total 76469 \(56.7%\) having only 1 herd](#)

273 [present on the farm, 35.9% having 2 herds present, 6.7% having 3 herds present, 0.6% with 4 herds](#)
274 [and only 0.02% of the population \(13 farms\) having 5 herds on a single farm.](#)

275 [Figure 9.2: Variation in size amongst herds within the 10 herd types. From 1 to 10: Large sheep,](#)
276 [small sheep, large pigs, small pigs, large deer, small deer, large dairy, small dairy, large beef, and](#)
277 [small beef. Note that the x and y axes scales differ to allow better visualisation of the smaller classes](#)
278 [although bin width is set to 10 throughout.](#)

279 [Figure 9.3: Plot shows the log number of infectious premises \(IPs\) for different Infectivity and](#)
280 [susceptibility values \(x axis\) for 3 models built in the Australian Animal Disease Simulation Model](#)
281 [\(AADIS\). The three models have the same parameterisation for all settings apart from the herd file](#)
282 [used.](#)

283 [Figure 9.4: Proportions of herd infections by each spread pathway for a model run from a set seed in](#)
284 [the Waikato that used a point estimate of farm size \(blue\) compared with a model that used a beta](#)
285 [pert distribution for farm size \(pink\). From left to right spread pathways are airborne spread, direct](#)
286 [spread, indirect spread, local spread between farms, sale yard spread and within farm between herd](#)
287 [spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.](#)

288 [Figure 9.5: Proportions of herd infections by each spread pathway for a model run from a set seed in](#)
289 [the Taranaki that used a point estimate of farm size \(blue\) compared with a model that used a beta](#)
290 [pert distribution for farm size \(pink\). From left to right spread pathways are airborne spread, direct](#)
291 [spread, indirect spread, local spread between farms, sale yard spread and within farm between herd](#)
292 [spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.](#)

293 [Figure 9.6: Proportions of herd infections by each spread pathway for a model run from a set seed in](#)
294 [Canterbury that used a point estimate of farm size \(blue\) compared with a model that used a beta](#)
295 [pert distribution for farm size \(pink\). From left to right spread pathways are airborne spread, direct](#)
296 [spread, indirect spread, local spread between farms, sale yard spread and within farm between herd](#)
297 [spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.](#)

298 [Figure 10.1: Flowchart illustrating the method of assigning AgriBase farm types to FarmsOnLine](#)
299 [records, for use of New Zealand Standard Model for foot-and-mouth disease epidemic simulation.](#)

300 [Figure 10.2: Estimated counts of livestock premises centroids by distance from an arbitrary livestock](#)
301 [premises centroid for all districts \(n = 69\) in the 16 New Zealand regions based on AgriBase \(AB\) and](#)
302 [FarmsOnLine \(FOL\) livestock databases](#)

303 [Figure 10.3: Estimated counts of centroids of livestock premises within 5 km of an arbitrary premises](#)
304 [centroid for 54 New Zealand districts, based on AgriBase \(triangle\) and FarmsOnLine \(square\)](#)
305 [livestock database, and the mean of the two \(circle\), in the descending order by the mean estimates.](#)

306 [The seven districts selected for foot and mouth disease \(FMD\) simulation are shown in red.](#)

307 [Figure 10.4: The locations of the seven New Zealand districts selected for simulation of foot and](#)
308 [mouth disease \(FMD\) epidemics](#)

309 [Figure 10.5: The cumulative number of premises detected with foot and mouth disease \(FMD\), by](#)
310 [number of days elapsed since first detection for three control policies and seven districts, based on](#)
311 [disease simulation using AgriBase \(AB\) and FarmsOnLine \(FOL\) data. The policies are stamping-out](#)
312 [only \(SO\), SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and](#)
313 [dry stock grazing premises \(SOV_{targeted}\) or all premises \(SOV\). The median \(solid line\) and the 95th](#)
314 [percentile \(dot-dash line\), with 5th and 95th percentile range \(shade\) are shown.](#)

315 [Figure 10.6: The simulated epidemiological outcomes of foot and mouth disease \(FMD\) epidemics](#)
316 [lasting for >14 days for three control policies and seven New Zealand districts, using AgriBase \(AB\)](#)
317 [and FarmsOnLine \(FOL\) data \(total iterations: n = 22,466 and 24,284 for AB and FOL\). The policies are](#)
318 [stamping-out only \(SO\), SO with emergency vaccination within 3 km vaccination zone targeting](#)
319 [pastoral, dairy and dry stock grazing premises \(SOV_{targeted}\) or all premises \(SOV\). Box represents the](#)
320 [first and the third quartiles, and whisker the 5th and 95th percentiles, and outliers \(<5th and >95th](#)
321 [percentiles\) were removed.](#)

322 [Figure 10.7: The estimated short-term costs \[A\] and total costs \[B\] of simulated foot and mouth](#)
323 [disease \(FMD\) epidemics in seven New Zealand districts, with seven control strategies, based on](#)
324 [AgriBase \(AB\) and FarmsOnLine \(FOL\) databases. The strategies are stamping-out only \(SO\), SO with](#)
325 [emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing](#)
326 [premises \(SOV_{targeted}\) or all premises \(SOV\), with vaccinate-to-die \(VTD\) or vaccinate-to-live \(VTL\)](#)
327 [policy. Box represents the first and the third quartiles, and whisker the 5th and 95th percentiles, and](#)
328 [outliers \(<5th and >95th percentiles\) were removed.](#)

329 [Figure 10.8: Cumulative density functions of the estimated marginal benefit cost ratio \(B/C\) of four](#)
330 [vaccination-based foot and mouth disease \(FMD\) control strategies relative to stamping-out alone](#)
331 [\(SO\), for simulated epidemics lasting for >14 days in seven New Zealand districts, based on AgriBase](#)
332 [\(AB\) and FarmsOnLine \(FOL\) databases. The strategies are SO with emergency vaccination within 3](#)
333 [km vaccination zone targeting pastoral, dairy and dry stock grazing premises \(SOV_{targeted}\) or all](#)
334 [premises \(SOV\), with vaccinate-to-die \(-VTD\) or vaccinate-to-live \(VTL\). The dotted line shows B/C =](#)
335 [1, above which the benefits exceed the additional costs \(i.e. i.e. addition of emergency vaccination is](#)
336 [more beneficial than SO\).](#)

337 [Figure 10.9: Sensitivity analyses for local spread probabilities 1.5 times that of the original values:](#)
338 [cumulative density functions of the estimated marginal benefit cost ratio \(B/C\) of four vaccination-](#)
339 [based foot and mouth disease \(FMD\) control strategies using vaccination relative to stamping-out](#)
340 [alone \(SO\), for simulated epidemics lasting for >14 days in seven New Zealand districts, based on](#)

341 [AgriBase \(AB\) and FarmsOnLine \(FOL\) databases. The strategies are SO with emergency vaccination](#)
342 [within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises \(SOVcattle\) or](#)
343 [all premises \(SOV\), vaccinate-to-die \(-VTD\) or vaccinate-to-live \(VTL\). The dotted line shows B/C = 1,](#)
344 [above which the benefits exceed the additional costs \(i.e. i.e. addition of emergency vaccination is](#)
345 [more beneficial than SO\).](#)

346 [Figure 14.1: Correlation identified amongst predictor variables extracted from GIS and land cover](#)
347 [data](#)

348 [Figure 14.2: remaining correlation once synthesized variables have been added to the dataset](#)
349 [representing variables included in the model building process.](#)

350 [Figure 14.3: The red bars represent the size of error calculated at the mesh block level for strategy 2](#)
351 [\(red\) and strategy 3 \(blue\). While removing the lifestyle properties does decrease the MAE \(Table](#)
352 [14.1\) for some species, and affects the mean error positively, when plotted the variation in errors](#)
353 [remains the same \(the large amount of overlapping red and blue which shows as purple\).](#)

354 [Figure 14.4: Regional distribution of APS analysis sample and achieved matches per area unit](#)
355 [\(aggregations of meshblocks\) as per 2013. 492 entries of the APS analysis sample \(N = 41182\) could](#)
356 [not be unambiguously attributed to an area unit due to incorrect/ insufficient location information](#)
357 [in the APS dataset and are consequently not represented in the figures. Area unit boundaries were](#)
358 [sourced from Statistics New Zealand licensed under CC-BY.](#)

359 [Figure 14.5: Regional distribution of APS analysis sample and achieved matches per territorial](#)
360 [authorities as per 2013. 492 entries of the APS analysis sample \(N = 41182\) could not be](#)
361 [unambiguously attributed to an area unit due to incorrect/ insufficient location information in the](#)
362 [APS dataset and are consequently not represented in the figures. Territorial authority boundaries](#)
363 [were sourced from Statistics New Zealand licensed under CC-BY.](#)

364 [Figure 14.6: Regions covered by the validation survey. 1: Peri-urban area / rural with strong urban](#)
365 [influence; 2: Semi-rural area / rural with moderate urban influence; 3: Rural area / rural with low](#)
366 [urban influence](#)

367 [Figure 14.7: Peri-urban survey region \(South Auckland\)](#)

368 [Figure 14.8: Semi-rural survey region \(Hamilton area\)](#)[Figure 14.9: Rural survey region \(Central](#)
369 [Hawkes Bay\)](#)

370 [Figure 14.10: Counts of observations of movements to production types](#)

371 [Figure 14.11: Movements from production types](#)

372 [Figure 14.12: Movements off dairy production properties. Distance is shown on the x axis and](#)
373 [frequency on the y axis. Aside from the peak at 0 which are associated with registration, the](#)
374 [distribution adequately approximates a beta pert.](#)

375 [Figure 14.13: Distance travelled during gypsy day movements \(in this case defined as a consignment](#)
376 [of more than 100 animals moved together\) by month and region in km. Note the scale of the y axis is](#)
377 [1000 km. This graph shows all ~5000 movements present in the data.](#)

378 [Figure 14.14: Distance travelled during gypsy day movements \(in this case defined as a consignment](#)
379 [of more than 100 animals moved together\) by month and region in km. Note the scale of the y axis is](#)
380 [300 km. These movements are those that go within region only.](#)

381 [Figure 14.15: Distance travelled during gypsy day movements \(in this case defined as a consignment](#)
382 [of more than 100 animals moved together\) by month and region in km. Note the scale of the y axis is](#)
383 [1200 km. These movements are those that leave their region of origin for another region.](#)

384 [Movements are plotted by the region that they originate in. For instance in June, many movements](#)
385 [leave the Waikato and some move as far as 1100km.](#)

386 [Figure 14.16: Distance travelled during gypsy day movements \(in this case defined as a consignment](#)
387 [of more than 100 animals moved together\) by month and region in km. Note the scale of the y axis is](#)
388 [1200 km. These movements are those that leave their region of origin for another region.](#)

389 [Movements are plotted by the region that they go to. For instance in June, many movements are](#)
390 [made to Southland and some move from as far away as 1000km.](#)

391

392

393 **Executive Summary**

394 Isolation and strict biosecurity measures implemented by Australia and New Zealand have prevented
395 the incursion of many organisms of biosecurity concern. The agricultural industry is a key part of the
396 economy for both countries, and preventing the arrival of diseases of concern, including foot-and-
397 mouth (FMD) disease, is crucial to maintaining access to international markets, reputation, and
398 protecting the economy and industry. Early detection of, and a rapid, effective response to such
399 diseases have a large impact on limiting the economic damage caused by epidemics.

400 An acknowledged weakness of biosecurity preparedness and response to agricultural diseases in
401 both countries is the lack of a single source of accurate, up-to-date farm livestock demographics
402 information.

403 It is inconvenient but undeniable that livestock producers do not always make decisions based solely
404 on economics. This is particularly the case when it comes to “lifestyle” farming, which is an
405 important sector included in our study. Parallel work is being undertaken by the Ministry for Primary
406 Industries to better understand this sector, their practices and drivers. The authors further note that
407 all farm types are dynamic systems and that using static representations of the populations in these
408 systems will always be problematic.

409 This report reviews the use of animal counts for investigation of, preparedness for and response to
410 exotic animal disease outbreaks, and analyses the available New Zealand datasets in depth. Gaps
411 and weaknesses in the current data landscape are documented. The project objectives then focus on
412 developing methodologies to estimate national-level farm demographic data and assess the use of
413 modelled and inaccurate data in disease simulation models. There are nine key deliverables outlined
414 in the report, as follows:

- 415 1) We first reviewed the uses of farm-level animal populations and possible ways to value this
416 data
- 417 2) We investigated the use of remote sensing technologies for maintenance of farm animal
418 population data for use by the competent authority. We found that currently, remote
419 sensing does not provide a robust solution.
- 420 3) We then assessed the available national farm-level livestock databases, Agribase and
421 FarmsOnline (FOL), to identify gaps in the available data and to understand the interactions
422 between the datasets.
- 423 4) We reviewed ways to model animal populations at the farm-level using three machine
424 learning techniques to predict counts of livestock units and cattle

- 425 5) This was developed concurrently with a Bayesian framework that uses the same remotely
426 sensed environmental information, predictor and response data to enable comparison of
427 different modelling techniques
- 428 6) We performed further data matching to validate the results of both the machine learning
429 and Bayesian models against external data sources to identify the models that performed
430 best for the NZ situation. The findings indicate random forest models provide the best
431 solution for modelling counts of Livestock Units (LSUs) and cattle and that these models are
432 able to predict with reliable results based on the AgriBase™ database. Our solution did not
433 allow the use of the models developed with AgriBase to be used to predict LSU or cattle
434 counts in the second livestock database, FOL, with any success due to irregularity in the FOL
435 assignment of land parcels to farms and to the low quality of the land use information in
436 FOL. This result supports the previously stated need for national level animal demographic
437 databases to be aligned so that they are fully interoperable to avoid duplication of effort and
438 waste of resources. An acknowledged shortcoming of this project is that there was no field
439 based collection of animal counts. Due to biases in centrally held databases and time
440 differences in collection periods this is a serious shortcoming.
- 441 7) Following on from these results we used a farm-level disease simulation model (InterSpread
442 Plus - ISP) and a within herd spread model (the Australian Animal Disease spread model –
443 AADIS) to examine the effect that inaccuracy in demographic information (at the level of the
444 farm population and at the level of the animals within herds) has on disease spread
445 simulation models that are used for biosecurity preparedness. We found that in the silent
446 spread phase, estimation of distributions for herd sizes (rather than estimation of point
447 estimates for herds) provided results that did not differ from those simulations that used
448 point estimates. We found that in AADIS herd level heterogeneity was important when herd
449 size was regarded as a significant risk factor for infectivity and susceptibility to disease. Our
450 experiment at the farm-level with ISP produced results confirming that biases that affect the
451 counts of farms present in a national database will affect decision making in locations where
452 farms are highly clustered. This work forms a foundation for further experiments but is not a
453 complete topic . Further work is ongoing.
- 454 8) Finally, we use the two existing New Zealand farms databases, AgriBase and FOL, to test
455 sensitivity of model-based recommendation for control of a foot-and-mouth disease (FMD)
456 outbreak in New Zealand, by variation in density of livestock premises and enterprise
457 categorisation. Simulations using both AgriBase and FarmsOnLine data resulted in somewhat
458 consistent policy recommendations for FMD control. The results may provide policymakers

459 with indication regarding areas where efforts should be focused in increasing the accuracy of
460 national livestock data for New Zealand.

461 9) This work enabled capacity development and information transfer of MPI biosecurity
462 personnel by developing expertise and knowledge around the existing data available, its
463 strengths and weaknesses in addition to the ability to generate input data for use in
464 epidemic modelling for preparedness work for infectious disease outbreaks where complete
465 demographic data are unavailable.

466

467 The authors point out that privacy and confidentiality issues are not addressed in this project. This is
468 not because we do not value these aspects of the discussion, but rather that they are outside of the
469 scope of this work . We have limited ourselves to the technical case for and discussion of baseline
470 demographic information for use in exotic disease epidemiology.

471 This project is been very valuable for MPI as it has provided a means to develop understanding of
472 the available national-level data sets that contribute to the competent authority's knowledge of
473 farm animal populations. It has allowed users of these population level datasets to fulfil their
474 responsibility to understand the strengths, weaknesses, alternatives to and biases of the datasets
475 that are used to provide technical advice for biosecurity response and disease preparedness
476 planning. This project is necessary to highlight the importance of up-to-date, accurate population
477 information. These data are difficult and expensive to maintain and historical decisions made in
478 New Zealand have resulted in the current fragmented situation. We present this report as an
479 attempt to bring these issues to light and to encourage discussion regarding the collection of these
480 data, their maintenance and importance.

481 Recommendations on aligning farm level demographic information in New Zealand into a single
482 national farms register have been made, the authors believe that this approach is in line with the
483 importance of the data and the small size of New Zealand .

484

485

486

487

488

490 Introduction

491 Isolation and strict biosecurity regulations have prevented the incursion of many significant
492 agricultural pests and diseases into New Zealand and Australia. The agricultural industry is important
493 in the economies of both countries, but especially in New Zealand where it contributes more than
494 half of export revenue and almost a fifth of national GDP (Statistics New Zealand, 2014). Preventing
495 the arrival of agricultural diseases is crucial to maintaining access to international markets,
496 reputation, and protecting the economy and industry. Biosecurity investment by both governments
497 is substantial but the recent and economically damaging incursions of varroa mite, fire blight and
498 bacterial kiwifruit vines disease (PSA) in NZ show that significant biosecurity pests can still pass
499 borders undetected.

500 Farm demographic data are required for effective and efficient surveillance, investigation and
501 response to emerging and exotic diseases, and preparedness activities, which by definition happen
502 in advance of an outbreak and inform how agencies might prepare to protect biosecurity. The NZ
503 and Australian biosecurity regulatory agencies, Ministry for Primary Industries (MPI) and
504 Department of Agriculture and Water Resources (DAWR) respectively, have a duty to prepare for
505 potential disease incursions. Inaccurate livestock demographic information can impede their ability
506 to prepare and respond, for example on disease management planning (e.g. carcass disposal,
507 vaccine bank and laboratory capacity), recovery following disease outbreak, and policy decisions
508 (e.g. use of vaccinations and zoning of control areas). In the event of an outbreak, livestock
509 demographic information will be crucial to prioritise surveillance activities, to allocate sufficient
510 resources for responding, and inform animal movement control measures. During the investigation
511 phase of an outbreak, such information will be essential to quantify the risk of disease spread on a
512 property, to neighbouring properties, and to properties where animal movements have taken place.
513 Similarly information about the animal population in the region surrounding the infected property is
514 critical and typically this information is required within 12 hours of a suspect property being
515 identified. Prior to incursions, livestock populations and their geographic distribution are required
516 for exposure assessment during risk analysis, to describe pathways and associated probabilities of
517 disease introduction and establishment.

518 Foot-and-mouth disease (FMD) is a particular policy concern for New Zealand and Australia due to
519 the severe economic, biological and social implications an incursion would have. Incursion response
520 to an FMD outbreak would be facilitated by the availability of a complete, up-to-date census of
521 vulnerable animals. Both Australia and New Zealand currently have accurate, up-to-date national-
522 level databases of the geographic locations of rural holdings, however, counts of livestock on these

523 holdings are not reliably available. In addition, the accuracy of the information available on the land
524 use category for each rural land parcel is not known. The importance of this lack of information for
525 biosecurity and disease preparedness is recognized (Emelyanova et al., 2009; Ministry for Primary
526 Industries, 2012).

527 Models for FMD outbreaks are very sensitive to the time between infection and detection. The delay
528 is in turn sensitive to the level of monitoring of herds and flocks for disease outbreaks. Monitoring of
529 'lifestyle' block farms is the most difficult to maintain, in particular, among the increasing number of
530 such farms at the urban fringes. The number of these lifestyle properties could be in the order of
531 100,000 in New Zealand (according to the Ministry for Primary Industries maintained database,
532 FarmsOnLine) with unknown animal numbers and species, and no mandatory requirement for
533 reporting. Such issues represent a significant knowledge gap when surveying, analysing, and
534 responding to biosecurity risks. To help the NZ government initially, and the Australian government
535 subsequently, to prepare for and effectively respond to disease threats, a major goal of this project
536 was to develop a stochastic spatial model that can estimate national-level farm animal demographic
537 data using species distribution models (SDM). The models are built from a commercial livestock
538 database and environmental predictor data for New Zealand, some of which have been derived from
539 satellite imagery, with scripts that are readily available and can be updated and rerun with new
540 input data.

541 At the time of writing, the Surveillance and Incursion Investigation (Animals and Marine) team of the
542 NZ government's Ministry for Primary Industries (MPI) is examining the integrity of available national
543 level farm data, both for animal movements and counts, and developing methodologies to estimate
544 national-level farm demographic data. This project develops and tests tools for that objective
545 including the use of modelled and inaccurate data in disease simulation models. The tools will be
546 equally applicable in both Australia and New Zealand. In line with this objective, this project
547 assessed the availability of predictor data and the application of remotely sensed imagery for
548 livestock detection, the available livestock databases and the efficacy of species distribution
549 modelling techniques, use of the data in disease simulation models and validation of the modelled
550 data sets. The project also highlights capacity building and knowledge transfer to biosecurity staff of
551 the available data and databases. In line with this, this project provides nine key deliverables:

- 552 1. A review of the uses of farm-level animal populations and possible ways to value this data
- 553 2. A review of remotely sensed data techniques that identifies if they provide a current
554 solution to the data gap present in New Zealand

- 555 3. A review of the existing data sets available to inform animal populations and their strengths
556 and weaknesses.
- 557 4. A machine learning model that uses remotely sensed environmental information to predict
558 counts of livestock units and cattle
- 559 5. A model in a Bayesian framework that uses the same remotely sensed environmental
560 information to predict counts of livestock units and cattle
- 561 6. The use of both of these models to predict outcomes for farms and validate these
562 predictions against external data
- 563 7. Use of a within herd spread disease spread model to estimate how accurate farm-level data
564 must be to provide advice that can be used for disease preparedness.
- 565 8. Use of the two existing New Zealand farms databases in a disease spread model and a
566 comparison of the results of simulations and their impact on decision making.
- 567 9. Capacity development and information transfer by developing expertise and knowledge of
568 MPI biosecurity personnel around the existing data available, its strengths and weaknesses
569 in addition to the ability to generate input data for use in epidemic modelling for
570 preparedness work for infectious disease outbreaks where complete demographic data are
571 unavailable.

572 This is the final report for the two year project and outlines the methods and results of the nine
573 deliverables. Several deliverables have been (2, 3, 4 and 55) or will be submitted after further
574 revisions for publication in peer review journals and are therefore provided as stand-alone chapters
575 in this report. Deliverables 1, 6 and 8 may be submitted for peer reviewed publication with some
576 further additions and emendations. Deliverable 9 documents the impacts within MPI of this project.
577 Deliverable 7 represents some initial work that will build up to capability in New Zealand to use the
578 Australian Animal Disease Spread model (AADIS). We hope that this initial work leads to some
579 interesting modelling experiments in the future.

580

581 **Deliverable 1: A review of the uses of farm-level animal**
582 **count information for planning for TAD management and**
583 **the means by which the value of these data can be**
584 **described**

585 **Mary van Andel¹**

586 ¹ *Ministry for Primary Industries, Wellington, New Zealand 6140*

587 **Introduction**

588 According to the Food and Agriculture Organisation of the United Nations (FAO) a transboundary
589 animal disease (TAD) is defined as a disease of significant economic, trade and/or food security
590 importance for a considerable number of countries; that can easily spread into other countries and
591 reach epidemic proportions; and for which control/management, including exclusion, requires
592 cooperation between several countries (Otte et al., 2004).

593 An outbreak of such a disease in a disease free country can prove devastating to those countries
594 whose economies rely on the export of live animals and / or animal products, due to the trade
595 restrictions that are applied until proof of freedom can be re-established (World Organisation for
596 Animal Health, 2013). Adverse effects on the health and production of a country's national herd are
597 the responsibility of the competent authority which control, eradicate, manage and report on the
598 disease status of the country in question to the World Organisation for Animal Health (OIE).

599 According to a recent economic assessment performed by New Zealand's competent authority, MPI,
600 an outbreak of foot-and-mouth disease (a TAD of international concern and significance) would
601 result in a total loss in real GDP of up to \$16.2bn over 8 years (Forbes and van Halderen, 2014). A
602 high priority for New Zealand is thus to reduce the risk of an outbreak of FMD and to prepare as best
603 it can for such an outbreak should it occur. To this end, MPI maintains the veterinary service and has
604 responsibility for investigation, response and preparedness against exotic and emerging disease
605 outbreaks.

606 A necessary part of the description and control of disease is the quantification of the amount of
607 disease in a population (Thrusfield, 2007); a task which requires the size of the population to be
608 known. Chapter 5 of this report examined the availability of farm animal demographic data sources
609 in New Zealand and found that while New Zealand may be in a better situation than many countries
610 where farm boundary, farmer contact and animal count data are unavailable, the fragmented nature
611 of the databases and the lack of a single farm identifier creates a situation in which data are not able

612 to be used to the full potential benefit (Jewell et al., 2016). In a resource scarce environment,
613 changes to existing systems must be prioritized. The value of information (VOI) framework provides
614 a way to compare the costs of collecting further data with the value that the data has for decision
615 makers. The problem of maintaining up-to-date farm-level animal demographic data provides an
616 excellent opportunity to test these ideas and methods in the veterinary epidemiology space and
617 provide advice on the appropriate level of resources to use for maintaining farm demographic data.

618 In this review, we first briefly discuss the uses of farm-level animal populations in epidemiology and
619 disease spread modelling. We then provide specific examples of how demographic data are used to
620 inform preparedness and response activities for TAD outbreaks. This chapter is broken into two
621 broad parts:

- 622 1. Activities undertaken to inform policy on how best to respond in the event of an outbreak,
623 performed well in advance of an event;
- 624 2. Operational and intelligence functions that occur during an outbreak which must be
625 adequately prepared and practiced prior to an event occurring as well as being performed during the
626 outbreak.

627 We then discuss the implications of having bad data, and lastly introduce the value of information
628 (VOI) framework as a novel method for examining the value of accurate animal count information.

629 **Farm-level demographic data in epidemiology and disease** 630 **spread modelling**

631 Population characteristics influence the way that disease spreads and may be used to identify
632 individuals or areas at highest risk of infection (Ferguson et al., 2001). Disease control and planning
633 for these activities is performed on centrally held datasets. The accuracy of these data are seldom
634 understood or examined (Honhold and Taylor, 2006; Jewell et al., 2016). Recent studies have
635 explored how different models may be used to represent similar outbreak situations and have
636 compared model outputs within and between countries (Dube et al., 2007; Sanson et al., 2011;
637 Sanson et al., 2013; Roche et al., 2015). Sensitivity analyses have been reported for many model
638 parameterisations (Sanson et al., 2006a; Owen et al., 2011) but the effect of inaccuracy in the
639 population is seldom discussed. The population estimates will have varying effects on the model
640 depending on how it is used within the spread part of the model. For example, in the Australian
641 Animal Disease spread model (AADIS) (Bradhurst, 2015) the population of animals on each farm is
642 used to generate infectivity for each individual farm according to latent periods, within herd contact
643 rates and incubation periods specific to the species and numbers of each species present; this

644 creates a unique probability of transmission curve for each farm based on its population. In the
645 disease spread simulator InterSpread Plus (ISP) (Stevenson et al., 2013) a probability of transmission
646 curve is applied to each farm type based on the average size (in New Zealand's case across all of
647 New Zealand) of a herd of a particular type.

648 The 2001 FMD epidemic in the UK came under intense scrutiny by the public and academics. Some
649 authors have suggested that model based decisions may have been made differently if different
650 animal demographic data were used to build the models (Savill et al., 2007; Tildesley et al., 2010;
651 Tildesley and Ryan, 2012), particularly if within farm dynamics were included (Tildesley et al., 2008)
652 and the spatial scale of the model had been smaller (Shea et al., 2014). Disease spread is affected by
653 the number of animals present on farms, as well as animal and vector movements that link farms to
654 each other spatially. Modelled predictions of spread are affected by the scale of the model, for
655 example the use of polygon or farm-level area information as opposed to point features that
656 represent the weighted centroids of farms will affect model predictions of spread (Flood et al.,
657 2013). In the event of prioritisation of farms to receive surveillance visits, those with a boundary that
658 is shared with an infected place (IP) can be expected to have higher priority than those in close
659 proximity and not sharing a fence line.

660 Advances in modelling techniques and increased computational ability have resulted in the current
661 trend in veterinary epidemiology of using individual animals to generate infectivity at the farm-level
662 by the inclusion of intra-herd spread within disease spread simulators. Due to this high level of detail
663 at the farm-level, these models are being used to evaluate the impacts of management decisions at
664 a microeconomic level on individual farms and to address the effects of the vaccine strategies
665 described in 3.1.1 above. Livestock density strongly influences the farm-level reproductive number
666 used in these models (Porphyre et al., 2013) and can only be estimated for a particular farm by
667 knowing the count of animals present on the property. Porphyre et al (2013) further show that it is
668 the spatial variation of these farm-level reproductive numbers that best inform a decision to
669 vaccinate or not, along with knowledge of the circulating strain of virus.

670 **Preparing for outbreaks by performing research activities that** 671 **inform policy**

672 Information on the size and distribution (demography) of the susceptible populations at the farm-
673 level and their proximity to sources of infection are required to adequately prepare for TAD
674 outbreaks (Woolhouse, 2003) as these characteristics fundamentally influence how disease will
675 spread (Porphyre et al., 2013; Buhnerkempe et al., 2014). Forming policy frameworks for exotic

676 disease control (including defining and documenting a strategy to reduce disease spread, vaccine
677 bank resourcing and the prior identification of burial sites for animal disposals) relies on knowing the
678 farm animal demography of the area in question (Tomassen et al., 2002). Consider the following very
679 specific planning tasks that must be performed as part of preparation for a TAD outbreak:

680 **i. Planning for culling & burial**

681 Once a disease is identified on multiple properties, depopulation of the properties will be required in
682 the event of a severe TAD. Insufficient planning for associated disposal activities has numerous
683 impacts that may extend many years after the end of an outbreak and can include economic losses,
684 groundwater contamination, and air pollution from burial sites (Gwyther et al., 2011;
685 Joung et al., 2013). These major challenges were encountered during the 2001 FMD epidemic in the
686 UK and the Korean FMD outbreaks in recent years, with drinking water contamination problems
687 being encountered in both countries as a result of inappropriately positioned burial pits. To
688 adequately plan for these events, farm-level animal counts are needed to estimate if a particular
689 area has sufficient capacity and capability to dispose of large numbers of livestock carcasses during a
690 large outbreak. This is important in New Zealand, which does not have sufficient rendering and
691 commercial incineration capacity ([http://www.mfe.govt.nz/publications/environmental-
692 reporting/state-new-zealand%E2%80%99s-environment-1997-chapter-three-production-3](http://www.mfe.govt.nz/publications/environmental-reporting/state-new-zealand%E2%80%99s-environment-1997-chapter-three-production-3)) for
693 projected outbreaks, and therefore will rely heavily on on-farm burial, burial at landfills, and burning
694 on pyres¹. Alternative strategies include composting, high-temperature gasification or the
695 negotiation with government bodies regarding the transportation of carcasses across multiple
696 regions to reach suitable landfill sites. These negotiations must be informed by geospatial analyses
697 to compare livestock densities with local capacities to bury or burn carcasses to identify areas that
698 require special measures.

699 **ii. Planning for vaccination**

700 Accurate estimates of animal numbers are required to prepare sufficient resourcing to respond to an
701 exotic disease incursion; for example the numbers of vaccines required to vaccinate a particular area
702 are determined by what vaccination policies are adopted and how many animals are present.

703 Vaccine strategies require animal counts to make decisions on the best policy for a particular
704 outbreak situation or region. Examples include:

¹ Personal communication with Evelyn Pleydell (MPI disposal project lead) by email 12092016 "I would not recommend using any rendering plant where we have to dismember or pre-break carcasses, which basically leaves the 2 Wallace's plants in NI (Taranaki and Waikato) and I struggle to believe that the throughput would be of any major use..... particularly as there'd be a degree of set-up required to get it operational and then extensive C&D afterwards. " and regarding incineration speaking of the capacity of New Zealand's incinerators " 2000kg of waste per hour at the largest incinerator in Auckland = a maximum of 4 cows/hour. To my mind, not a throughput that's of much practical use in FMD when we can take truck loads to the largest landfills, particularly Hampton Downs between Auckland and Hamilton."

705 *Single species vaccination:*

706 Models of hypothetical FMD outbreaks have returned conflicting results regarding the benefit of
707 vaccinating cattle only (Roche et al., 2015). Laboratory transmission studies have shown that cattle
708 and sheep are equally susceptible to FMD virus infection, but cattle are more infectious than sheep
709 (de Rueda et al., 2014). From an epidemiological perspective, there might be merit in vaccinating
710 cattle in preference to sheep in order to reduce transmission rates in the population when
711 vaccination resources are limited. In New Zealand livestock industries are primarily based on sheep
712 and cattle farming. As cattle and pigs are better indicator species for FMD than sheep and goats,
713 cattle-only vaccination may mask infection in an area where diagnosis of disease was dependent on
714 clinical signs, which could delay eradication or interfere with proof of freedom surveillance testing
715 after an outbreak. This topic is of interest to a number of countries. If cattle- only vaccination were
716 to be shown to be equally effective as all species vaccination, it could result in substantial savings in
717 the number of vaccine doses required to confer benefits, and decrease the number of animals to be
718 destroyed in “vaccinate-to-die” policies (see below). Vaccination of a subset of species or a single
719 species is clearly an important area for future study. However detailed information on within and
720 between species spread of FMD will be essential to gain useful information on this complex
721 problem, becausebecause of the sizes of the populations involved and their contact dynamics will
722 have effects on model outcomes.

723

724 *“Vaccinate-to-live” (VTL) vs “Vaccinate-to-die” (VTD):*

725 The options for exporting countries employing vaccination for control of an FMD outbreak are
726 limited to either a “vaccinate-to-live” (VTL) or a “vaccinate-to-die” (VTD) policy (Anonymous,
727 2014).Anonymous, 2014) Under the former policy, vaccinated animals continue to be farmed
728 normally for the duration of their productive lives, and the country seeks to prove to the Office
729 International des Épizooties (OIE) that it is “free from FMD with vaccination”. The latter means that
730 the vaccinated animals are slaughtered as soon as there is culling capacity available, and the country
731 seeks to then prove that it is “free from FMD without vaccination”. Under a VTD policy, vaccinating
732 animals places a death sentence on them. The associated costs, including loss of genetic material,
733 compensation to farmers, time to depopulate and dispose the animals following the epidemic are
734 expected to be very high. The OIE allows countries to regain disease-freedom 3 months after culling
735 all the vaccinated animals (although importing countries may still take longer to restore trade) if a
736 country adopts a VTD policy. Under a VTL policy a country can only regain FMD-freedom 6 months
737 after the last case is identified and surveillance on all vaccinated animals is completed. The objective
738 of this surveillance is confirmation that no infection is present in the national herd.

739 Under the current OIE regulations the fast and aggressive use of vaccination, which maximises the
740 usefulness of its deployment (Roche et al., 2015), is unattractive to decision makers who must
741 minimise the length of the outbreak to rapidly return to trade. This must be weighed against the
742 impacts of a VTD policy. Macroeconomic analyses and accurate estimates of local heterogeneity in
743 animal population is crucial for guiding decisions on deployment of vaccination and weighing the
744 merits of a VTL compared with a VTD policy.

745

746 **iii. Testing and laboratory capacity**

747 The capacity of the national laboratory to process large volumes of samples for the TAD in question
748 will inform some of the decisions made around implementing control measures. This is based on
749 clinical signs vs the presence of DNA or serological disease markers when multiple infected places
750 (IPs) are identified. Similarly the decision to vaccinate or not vaccinate will be informed by the ability
751 of the country to return to trade at the earliest possible time, which is in turn conditional on the
752 testing of large numbers of serum samples as specified by the OIE for the disease in question. In the
753 case of FMD, while “vaccinate to live” policies are attractive from a compensation and disposal point
754 of view the delay in return to trade and the extensive DIVA (differentiating infected from vaccinated
755 animals) testing has to be considered. All these considerations will vary greatly based on the
756 population of animals affected by the outbreak.

757 **iv. Economic modelling**

758 In a country where primary sector exports make up a significant part of GDP, like New Zealand,
759 preparing for a catastrophic situation that an outbreak of FMD would present is essential. However
760 reaffirmation of the value of investment in these activities is needed in a resource scarce economy.
761 Economic modelling that includes both farm and country level details is required to provide this
762 overview (Forbes and van Halderen, 2014).

763 **Preparing for and performing operations and intelligence** 764 **functions in the face of a suspected or confirmed outbreak**

765 The activities in this section will be performed during an outbreak of TAD on an ongoing basis,
766 however it is essential that those who perform these activities are familiar with them so that they
767 can be performed quickly under stressful conditions. The best way to ensure this is for the
768 participants to regularly practice these activities and for preparation to focus on the availability of
769 the necessary data and development of appropriate processes and capacity before an outbreak is
770 identified. All these tasks depend on a knowledge of how many livestock are present so that

771 adequate human resources can be obtained to perform the tasks at hand in the shortest possible
772 time. An accurate understanding of how many livestock and what species are present is very
773 important information at a local disease control centre in terms of resourcing and targeting both
774 surveillance and response.

775

776 i. *Investigation at farm-level* - During an animal disease investigation at the farm-level, investigators
777 use farm contact databases to obtain contact details of farmers/managers, information on other
778 holdings of the same owner including runoffs (properties which present the same level of
779 epidemiological risk as the farm under investigation because of farming activities on non-contiguous
780 land), and creation of maps showing the investigation farm and its location in relation to other
781 landmarks, which assist in assessment of spatial risk factors. During the investigation phase of an
782 outbreak, animal demographic information is essential to quantify the risk of disease spread on a
783 property, neighbouring properties and to properties that have received animal movements from the
784 farm in question. These factors (among others not related to animal demography) enable the extent
785 and severity of the problem on a property to be quantified. The disease investigator on the farm
786 must collect the movement and demographic information, and clinically examine all animals on the
787 farm.

788

789 ii. *Surveillance and tracing prioritisation*: Control activities must be prioritised to fit within limited
790 resources. Animal demographic information in a geographic area that is at risk of disease allows
791 prompt prioritisation of sites at high risk of spreading disease (Mansley, 2004; Bessell et al., 2010a).
792 Demographic data must be used to prioritise surveillance activities very early during a response
793 (Bessell et al., 2010b); for instance those properties with intensively housed pig populations within a
794 3km surveillance zone around a property that has been infected with foot and mouth disease may
795 be visited first due to the high risk of aerosolised virus attributed to pigs, and the risk of wind- borne
796 spread from these facilities. Bessell et al (2010a) identified farm-level populations and the
797 characteristics of the population in the surrounding 10km as important for estimating the risk of
798 infection with FMD. In the case of a disease outbreak, not all farms will present the same risk of
799 onward spread of disease. This prioritisation is critical as delays in identification of infected premises
800 have very large monetary and disease spread consequences (Carpenter et al., 2011). Similarly with
801 contact tracing, those farms with the presence of large numbers of virus excretors or those with
802 many contacts should be prioritised for immediate visits (Eames and Keeling, 2003). In the event of
803 an outbreak, MPI would make use of call centres and field veterinarians to get up-to-date
804 information on each property at risk of infection (a property might be at risk because of its proximity

805 to a known infected property or because of a pattern of movements or linkages that connect it with
806 a source of infection). The management of exotic disease outbreaks presents a significant challenge
807 to resources which makes prioritisation and allocation of effort essential. The presence of reliable
808 animal count information at the farm-level would allow field activities to be performed in a shorter
809 time frame, improving efficiency, as they will supply a more reliable source of information on which
810 to base prioritisation for purposes of disease management and control to prevent spread and
811 ultimately limit damage arising from an incursion.

812

813iii. *Size of controlled areas and logistics that go with controlled areas, i.e. roadblocks:* The immediate
814 establishment of movement controls and controlled areas can greatly limit the spread of an exotic
815 disease outbreak. These areas will depend on the animal population surrounding the property of
816 concern and the geographic position of high risk trace properties. This information is required within
817 12 hours of a suspicious disease having been identified on a farm.

818

819iv. *Risk assessment:* Animal numbers and their geographic distribution are required for exposure
820 assessment during risk analysis (Zepeda et al., 2001). This step of the risk analysis process describes
821 the pathways and their associated probabilities that might lead to disease introduction and
822 establishment. Moreover, the population composition at the farm-level has been identified as
823 important for estimating the risk of onwards infection with contagious livestock diseases (Keeling et
824 al., 2001; Haydon et al., 2003; Carpenter et al., 2004; Thrusfield et al., 2005; Bessell et al., 2010a).

825

826v. *Tasking and scheduling:* As has been explained for surveillance and tracing prioritisation, all field and
827 operational tasks must be scheduled to minimise risk of onward spread of disease. This includes the
828 scheduling of farms where animals are to be culled, buried and facilities and equipment disinfected
829 as well as the disinfection of non-animal vectors (e.g. stock conveyances).

830

831vi. *Reporting and providing advice on control strategies:* The epidemiology team within a TAD response
832 reports to decision makers on estimated progress of control activities and advise on changes that
833 should be made in the field to improve outcomes. These reports will include estimation of a farm
834 level R_e^2 , for which the distribution of the population at risk is necessary, providing an indication of
835 whether an outbreak is under control (Ferguson et al., 2001; Paine et al., 2010) and informing policy

² The basic reproductive number represents the number of new infections which arise, on average, from one infected individual when the entire population is susceptible (i.e. at the beginning of an epidemic) (Dohoo, Martin et al. 2009)

836 adjustments in the face of an outbreak. Surveillance activities, both routine and response, collect
837 counts of test results for diseases of interest. The ability to display these counts as a proportion of
838 diseased animals in a population is more meaningful than having counts of test results only, and
839 show the progress of response activities in the field. Changes in disease prevalence may only be
840 identified if the denominator (a count of unaffected animals) is available; likewise studies that seek
841 to identify the magnitude and strength of association of risk factors with disease are only meaningful
842 if such data are available.

843

844ii. *Early decision making:* Information is very scarce in the early stages of a fast moving TAD outbreak.
845 However, we know that in the case of FMD, to get the maximal benefit from operational strategies
846 such as movement control (Carpenter et al., 2011) and vaccination (Roche et al., 2015) they must be
847 implemented as early as possible. There has been recent interest in the literature in identifying
848 outbreak characteristics that might indicate the final size of an outbreak; species specific animal
849 population density in the 5km surrounding the infective place has been identified as a statistically
850 significant factor in the prediction of final outbreak size (Tomassen et al., 2002; Sarandopoulos,
851 2015).

852

853iii. *Study design:* In the course of TAD control and eradication, the design of appropriate studies may be
854 required to generate information on particular aspects which will contribute to control efforts
855 (Probert et al., 2015). Study design relies on the size of the population, the dynamics of disease
856 spread within the population and the demographics of the population (Slater, 2001). In addition, the
857 existence of accurate data of this sort will assist more complex assessments of an epidemic. For
858 example, it will provide an explanatory variable in regression modelling which will allow assessment
859 of risk factors for infection, or assist in spatial epidemiological approaches (i.e. observational
860 studies).

861

862ix. *Plume modelling:* Disease modelling may be used during an FMD outbreak to gain further
863 information on properties at risk based on their proximity to a source of infection. Populations of
864 animals present on the source farm, populations of animals present on farms under the modelled
865 plume and the prevailing weather conditions are required to complete this analysis (Sanson, 1994;
866 Sorensen et al., 2000; Donaldson et al., 2001; Alexandersen et al., 2002; Alexandersen and
867 Donaldson, 2002; Hess et al., 2008; Bessell, 2009). This information is used to prioritise surveillance
868 and target communications to areas which may be at higher risk.

869

870x. *Predictive modelling*: Disease modelling can be used during a disease outbreak to gain information
871 on proposed control strategies and to make forecasts about local and national eradication progress
872 (Morris et al., 2001). This was done in the 2001 UK epidemic when ISP (Stevenson et al., 2013) was
873 run 2 or 3 times a week for the duration of the outbreak (Stevenson, 2003) (page 191) using the real
874 epidemic data as the start point of the simulation. This method provided useful positive and
875 negative spatial predictive values at 0 to 14 days after the start of the simulation. Predictive accuracy
876 is expected to be greater when the model is parameterised to reflect regional heterogeneity
877 (Stevenson, 2003) (p.208).

878 **Cost of inaccurate data**

879 Although we are able to describe in detail the use of farm-level animal demographic data for
880 essential TAD preparedness and response, the accuracy of this data is very seldom examined. There
881 are publications noting that some of the preparedness work, for instance research for vaccination
882 strategy policy (number 1.ii), may be performed with data aggregated from the farm-level to slightly
883 larger spatial scales (Tildesley et al., 2010; Tildesley and Ryan, 2012). However, more worryingly, the
884 inaccuracy of farm-level data has been identified as a serious flaw in the management of and the
885 interpretation of model results from the UK 2001 FMD epidemic (Keeling et al., 2001; Savill et al.,
886 2007). In the New Zealand context, little attention has been paid to documenting the impact of
887 incomplete farm demography on disease management, preparedness and control. The reason for
888 this seems to be that New Zealand was very fortunate to have a single national farm database
889 (AgriBase) in the 1990s (Sanson and Pearson, 1997) when very few countries had such a resource.
890 Since then, changes in government funding and policy have seen the creation of a second farms
891 database, FarmsOnLine (FOL) managed by MPI, which does not contain reliable animal count
892 information (Jewell et al., 2016). The resulting division of government resources has decreased the
893 ability to keep AgriBase up-to-date with current information on animal populations, although
894 activities continue between industry representative bodies and local councils which use AgriBase for
895 purposes other than biosecurity. A sensitivity analysis of the effect of completeness of the farm
896 population for use in a disease spread simulation model for FMD was undertaken in the late 1990s.
897 This analysis was performed in InterSpread Plus (Stevenson et al., 2013). The farm file was randomly
898 thinned to approximate lack of information about how many farms were present in New Zealand
899 and the effect that this data inaccuracy would have on a simulated epidemic of FMD. Results
900 showed that when data were randomly thinned to 72% of the “true” data, the number of predicted
901 infected premises (IPs) was 36% lower than when 90% of the “true” population are present in the
902 model (Mackereth, 1998). This shows the importance of completeness of farm location data but

903 does not address the question of the accuracy of the population data present for each of the
904 locations.

905 Aside from the technical disease management risks detailed in sections 3.3 and 3.4 the competent
906 authority exposes itself to multiple high-level risks by not having these data sources up to date.

907 These include:

- 908 ● Reputational risk for not learning quickly from past experiences and delays in implementing
909 recommendations provided by various reviews/audits (Ministry for Primary Industries,
910 2013);
- 911 ● Delayed and less efficient execution of response operations due to uncertainty and/or time
912 required to case properties for animal numbers which affect operations as detailed in
913 section 3.4;
- 914 ● Active surveillance and modelling give less accurate and informative outcomes which may
915 reduce MPI's ability to assess risks and impacts and making sound decisions in policy and
916 operation as detailed in section 3.3 and 3.4;
- 917 ● Compromised quality of preparedness and readiness work which may not be tailored to the
918 real needs as detailed in section 3.3;
- 919 ● Disappointment or inaction from industry due to incorrect expectation that the current
920 database systems are supplying all the required demographic information; and
- 921 ● Reduced efficacy and effectiveness of government industry agreements as farm-level
922 demographic data required for an agreed level of response and preparedness are not
923 available, or data generated through partnership are not properly managed and utilized.

924

Discussion: Estimating the value of information of animal population data

925

926

It is clear that animal demographic information are required and that there should be some confidence about their accuracy if the competent authority is to fulfil its responsibilities in the area of preparedness and response to TAD outbreaks. As these databases are costly to maintain a way to establish the value of this information is required. The value of information (VOI) framework provides decision makers with a mechanism for assessing the trade-off between increased accuracy and the costs of achieving that accuracy (Yokota and Thompson, 2004; Keisler et al., 2014). Yokota and Thompson, 2004; Keisler et al., 2014). Value of information (VOI) has been applied in the context of epidemiology in the areas of import risk analysis (Disney and Peters, 2003), and decision making under foreign animal disease outbreak conditions (Cox et al., 2005; Shea et al., 2014; Probert et al., 2015). Disney and Peters, 2003), and decision making under foreign animal disease outbreak conditions (Cox et al., 2005; Shea et al., 2014; Probert et al., 2015). Further applications have been published for decision making in human global health (De Gourville et al., 2006), fisheries management (Forsberg and Guttormsen, 2006; Hansen and Jones, 2008), the utility for the public of labelling of meat products (Klain et al., 2014) De Gourville et al., 2006), fisheries management (Forsberg and Guttormsen, 2006; Hansen and Jones, 2008), the utility for the public of labelling of meat products (Klain et al., 2014) and management of natural resources (Williams et al., 2011), invasive species (Moore and Runge, 2012) and the abundance of endemic species (Williams et al., 2011) Moore and Runge, 2012) and the abundance of endemic species (Williams et al., 2011).

944

Approach to Estimating the impacts of inaccurate data on disease spread modelling with estimation of the value of information on farm populations as a final goal:

945

946

The question that we have is somewhat simpler than the questions examined by Shae and colleagues in 2014. It can be summed up in the following way:

947

948 The competent authority uses disease spread modelling to compare the relative benefits of different
949 control strategies. As far-reaching decisions are based on these relative benefits, the sensitivity of
950 the models used to variations in the number of farms in the simulation and the number of animals
951 present on each farm (if the models use both of these inputs). Furthermore, it is of interest to know
952 if (for the purposes of disease spread modelling) estimation of a distribution of animals in a herd or
953 within a herd type (when animal numbers are used to drive disease spread in the model) provides
954 the same results as the use of “real” herd sizes in a disease spread model. The use of a distribution
955 to generate animal numbers on farms might provide some assurance that when figures are within a
956 range of the “real” population, then the model results are robust.

- 957 1. Run a simulation with the “real” dataset. Identify the optimal control strategy.
- 958 2. Run a simulation with one or more “alternate” datasets. Identify the optimal control
959 strategy.
- 960 3. Compare the optimal strategies. If they are the same, identify differences in resourcing
961 requirements
- 962 4. If optimal strategies differ, apply the control strategy identified in point 2 to the simulation
963 in point 1 with the benchmark dataset and compare the difference in the fixed costs and other
964 outbreak outcomes.
- 965 5. If the optimal strategies are the same but resourcing requirements differ, apply the
966 resources suggested in point 2 to the simulation in point 1 and compare the difference in fixed costs
967 and other outbreak outcomes.

968 If the most advisable course of action and the expected resource requirements do not change based
969 on different farm herd data sets or different herd sizes within each farm, then the model under
970 examination is not sensitive to the changes in farm counts of herd sizes. This does not mean that the
971 competent authority does not require herd sizes, but rather that these data are not not highly
972 influential within the model examined.

973 The experiment outlined above will allow an economic value to be placed on different “starting
974 states” for the disease spread model chosen for the simulations. The economic value is related the
975 decisions made during the outbreak based on the data on animal populations available at the
976 outset.

977 A second experiment will consider the information collected in the initial stage of a disease response
978 where tracing and casing activities are performed to identify which properties must be prioritised for
979 disease surveillance. During this experiment three steps could be taken

980 1. Delay some operations in the model so that the time lag while information is collected is
981 incorporated in the disease control part of the model. Use a delay of a week and compare it with a
982 delay of 2 weeks or no delay at all to starting surveillance visits.

983 2. Examine different geographic scales for the effect of improved information. For example, run
984 one scenario with “starting state real” data from experiment 1 above. Allow animal counts to be
985 updated in the second week of the outbreak in a pre-defined area and examine how this changes
986 the outbreak outcome.

987

988

989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004

1005

1006

1007
1008

1009

1010

1011
1012

Conclusions

This review shows that farm-level animal demographic information are essential for TAD response and preparedness. In Chapter 9 we document the foundation work for the 2 experiments described above. We parameterise and perform a sensitivity analysis using New Zealand data in the Australian Animal Disease Spread Model (AADIS). Our sensitivity analysis was only performed for the silent spread period (the simulation was run with no control measures applied and no surveillance to detect disease) but it is a useful exercise which allows the experiments described in this chapter to be set up. In Chapter 10 we examine a simulated situation in which a decision maker makes the best management choice to constrain the length of the hypothetical FMD outbreak using one data source (AgriBase) in the InterSpread Plus framework and compare that with a situation that uses another data source (FOL) in the same framework. We hope that Chapter 9 and 10 as well as our future, planned analyses will be useful as they will explicitly show the losses that could be experienced if herd level farm animal population information is not available to the competent authority with responsibility for TAD preparedness and response. It is hoped that this economic information can be used to communicate the urgency of the need for up-to-date and accurate animal population information to decision makers.

1.

1013 **Deliverable 2: Review of Remotely Sensed Data Techniques**

1014 **Tracey Hollings¹, Mary van Andel², Mark Burgman¹, Andrew Robinson¹**

1015 ¹ *Centre of Excellence for Biosecurity Risk Analysis, School of Biosciences, University of Melbourne,*
1016 *Melbourne Australia 3010*

1017 ² *Ministry for Primary Industries, Wellington, New Zealand 6140*

1018

1019 **This paper is currently in review at Remote Sensing of Environment entitled "A review of the*
1020 *application of remotely sensed imagery for estimating population sizes of large-bodied wild and*
1021 *domestic animals"*

1022 **Introduction**

1023 Policy and planning for the livestock sector is often impeded by the lack of reliable and accessible
1024 information on the distributions and abundance of livestock species (Robinson et al., 2007b).
1025 Accurate estimates of livestock numbers are important for many different applications including
1026 epidemiological (e.g. disease risk and spread, disease emergence, vaccination programme planning,
1027 estimating antimicrobial consumption), environmental (e.g. Greenhouse inventories, grazing
1028 pressure, nutrient loading), and social applications which might include estimates of production, and
1029 contributions to nutrition and food security. Population estimates may also be used to validate
1030 predictive models which may be implemented where accurate livestock population information is
1031 lacking or not on the required scale. Recent advances in the availability and resolution of remotely
1032 sensed data may allow augmentation or replacement of data that have been gathered by traditional
1033 means. Remotely sensed imagery has been used extensively for classifying habitats and vegetation
1034 cover (e.g. Sheeren et al., 2009; Recio et al., 2013; Peña et al., 2014), detecting changes in land use
1035 or land cover (e.g. Desclée et al., 2006; Koller and Samimi, 2011; Hussain et al., 2013), and classifying
1036 urban environments (Jensen and Cowen, 1999). More recently, remotely sensed imagery has been
1037 used to estimate animal populations, through either the detection of individual organisms (e.g.
1038 Laliberte and Ripple, 2003; Groom et al., 2011; Yang, 2012; Fretwell et al., 2014), or environmental
1039 proxies related to their presence (e.g. Robinson et al., 2007b; Lynch et al., 2012; LaRue et al., 2014).

1040 Identifying and monitoring populations from remotely sensed imagery may provide accurate
1041 estimates of animal populations (Fretwell et al., 2012; Pettoirelli et al., 2014) and has the advantage
1042 of being able to address questions on spatial scales that could not be undertaken using solely
1043 ground-based methods (Pettoirelli et al., 2014). Although animal population estimation from

1044 remotely sensed imagery was proposed more than 30 years ago (Löffler and Margules, 1980), one of
1045 the biggest hurdles in conservation, environmental and agricultural applications is the perception
1046 that the level of image resolution is not sufficiently informative (Turner et al., 2003). Other studies
1047 have contradicted this perception, by using surrogates based on an animal's ecology to predict
1048 population abundances from satellite imagery. For example, Löffler and Margules (1980) were able
1049 to approximate the distribution of wombats based on the detection of burrows, Velasco (2009)
1050 detected active marmot mounds in Mongolia, and several studies have estimated penguin
1051 populations from guano-stained ice (Fretwell et al., 2012; Lynch et al., 2012; LaRue et al., 2014).
1052 Satellite-derived habitat suitability maps have also been used to estimate species richness (e.g. Seto
1053 et al., 2004; Leyequien et al., 2007). However, only recently has the resolution of remotely sensed
1054 imagery and image processing tools and software been sufficient to detect individual, large-bodied
1055 animals.

1056 Remote sensing has advantages over direct observation for the census of animals because it can
1057 cover large areas, has short revisit intervals, is unobtrusive so as not to disturb animals, requires less
1058 human intervention, and multi-spectral data are available at high resolution (Yang et al., 2014).
1059 Estimating populations using direct detection techniques from remote sensing imagery is still
1060 relatively rare though proof-of-concept studies are emerging (e.g. McNeill et al., 2011; Conn et al.,
1061 2013; Groom et al., 2013), showing some success with moderate and high resolution satellite
1062 imagery (Pettorelli et al., 2014). Population estimation by direct detection has been used mainly in
1063 the context of wildlife conservation, in particular for inaccessible or remote areas, and only very
1064 recently for livestock species including cattle, sheep, poultry, deer and horses. Computer pattern
1065 recognition and object counting techniques are widely used in many scientific disciplines, including
1066 microbiology and medical science, and are now being adapted for identifying and counting animals
1067 (Laliberte and Ripple, 2003).

1068 In this chapter, we review current spatial tools and technologies for their potential to estimate
1069 animal populations from remotely sensed data. We outline recommended methods and data
1070 requirements to generate reliable estimates and consider how methods in the future could take
1071 advantage of better resolution and lower costs of imagery, greater computing power, and more
1072 advanced algorithms and programming.

1073

1074

1075 **Table 4.1:** A list of representative peer reviewed articles applying automated and semi-automated
 1076 techniques to remotely sensed imagery to estimate the abundance of mammals

Reference	Target species	Location	Method	Remote sensing imagery type and resolution	Population size of predictions or geographic extent	Measure of accuracy
(Yang et al., 2014)	Wildebeest (<i>Connochaetes taurinus</i>), zebra (<i>Equus quagga burchellii</i>), and gazelle (<i>Eudorcas thomsonii</i>)	Maasai Mara National Reserve, Kenya	Artificial neural network applied to spectral reflectance	Very high resolution satellite imagery	Two pilot study areas 1km x 1km each; hundreds of animals	Average count error of 8.2%, omission error of 6.6% and commission error of 13.7%
(Terletzky and Ramsey, 2014)	Cattle (<i>Bos taurus</i>) and horses (<i>Equus caballus</i>)	Utah, USA	Short-time interval image differencing using principal components analysis	Aerial imagery	Total of 158 animals across eight pastures	82% of the animals correctly identified, mean percent commission was 53%, and mean percent omission was 18%.
(Oishi and Matsunaga, 2014)	Cattle; sika deer (<i>Cervus nippon</i>); humans	Hokkaido, Japan	Image differencing	Aerial imagery	20 km ²	Not applicable
(Laliberte and Ripple, 2003)	Caribou (<i>Rangifer tarandus</i>) and cattle	Caribou: Alaska, USA; Cattle: Oregon, USA	Thresholding	Caribou: Black-and-white aerial imagery; Cattle: Very high resolution satellite imagery	60-406 individuals of caribou; >50 cattle	Mean count error for aerial photos of caribou 10.2%
(LaRue et al., 2015)	Polar bears (<i>Ursus maritimus</i>)	Rowley Island, Canada	Supervised spectral classification ; image differencing	Very high resolution satellite imagery	Approx. 1,000 km ²	Supervised classification was uninformative ; Automated image differencing correctly identified nearly 90% of bear locations.
(Fretwell et al., 2014)	Southern Right Whales (<i>Eubalaena</i>)	Golfo Nuevo, Peninsula Valde's in	Various; thresholding the best performing	Very high resolution satellite imagery	113 km ² <100 individuals	Thresholding: found 84.6% of all manually

	australis)	Argentina				digitized whales and 89% of the objects manually classed as probable whales, with 23.7% false positives. Supervised classification: no meaningful results
(Mejias et al., 2013)	Dugongs (Dugong dugon)	Shark Bay, Western Australia	Morphological-based detection; Shape profiling on saturation channel	Aerial imagery from unmanned aerial vehicle (UAV)	100 images taken at 100 ft	Recall values of 48.57% and 51.4%, and precision values of 4.01% and 4.97% for the two algorithms respectively
(Barber-Meyer et al., 2007)	Emperor penguins (<i>Aptenodytes forsteri</i>)	Ross Sea, Antarctica	Supervised classification	Very high resolution satellite imagery	Predict relative abundances for two categories: <3,000 or >5,000 adult birds; 12 images	Absolute percent deviation between predicted adult birds and known adult birds averaged 53% (SE = 15) and ranged from <1 to 128%
(Groom et al., 2011)	Lesser Flamingo (<i>Phoeniconaias minor</i>)	Kamfers Dam, Northern Cape, South Africa	Object-based image analysis methods : quadtree image segmentation and sequential object brightness thresholding	Aerial imagery	81 664 flamingos; 31 images	99% compared to human visual interpretation . Under-estimation by the object-based method of less than 0.5% is indicated
(Groom et al., 2013)	Five marine bird species: Common Scoter (<i>Melanitta nigra</i>); Great	Two offshore wind farms in the Irish Sea	Object-based image analysis methods	Aerial imagery	(total area ca. 670 km ²); 18 image frames;	The overall success level across these frames is 92.5%.

	Cormorant (<i>Phalacrocorax carbo</i>); Diver species group (<i>Gavia sp.</i>); Sandwich Tern (<i>Sterna sandvicensis</i>); Manx Shearwater (<i>Puffinus puffinus</i>)					
(McNeill et al., 2011)	Adélie penguins (<i>Pygoscelis adeliae</i>)	Ross Sea, Antarctica	Semi-automated software written in Matlab; linear discriminant analysis to separate the background; morphological image processing operators to select the breeding penguins	Aerial imagery	10 images	True positive and false negative were >85%
(Strong et al., 1991)	Lesser snow geese (<i>Chen caerulescens</i>)	Klamath Basin National Wildlife Refuge near Tule Lake, California,	Linear mixture model of pixel reflectance values	Multispectral scanner data and aerial photography	48 flocks of geese	Mixture model predictions and manual counts of the number of geese in a flock were highly correlated ($R^2 = 0.95$, $SE = 124.8$). mixture model prediction of the total number of geese in all flocks was 11.8% greater than the manual count.
(Trathan,	Macaroni	Bird	Image	Aerial	3 penguin	Automated

2004)	penguin (<i>Eudyptes chrysolophus</i>)	Island, South Georgia	Processing Toolbox of the Matlab; image segmentation, thresholding	photography	colonies with total of >50,000 birds; island size 5.0 km long by 0.8 km wide	image- analysis routines produced estimates that were highly correlated with ground counts.
(Descamps et al., 2011)	Greater Flamingo <i>Phoenicopterus roseus</i>	France, Turkey and Mauritania	Stochastic approach based on object processes; a birth-and- death algorithm	Aerial photographs	~2000 to greater than 10,000	Less than 12.9% of the birds were missed by the model; The overall error was 5% in 11 of 15 samples, and 2% in 9 cases out of 15.

1077

1078

1079

Automated and semi-automated methods for the detection of animals from remotely sensed imagery

1080

1081 At the time of writing, all published studies of automated estimation of animal numbers from
1082 remotely sensed imagery are proof-of-concept work using relatively small areas and in relatively
1083 homogenous environments (e.g. Laliberte and Ripple, 2003; Groom et al., 2013; Terletzky, 2013;
1084 Fretwell et al., 2014; Table 4.1). The first attempt to automate counts of individual animals from 1m
1085 resolution satellite imagery was made by Laliberte and Ripple (2003), using ImageTool software
1086 developed for medical imaging analysis. The last few years have seen the emergence of technologies
1087 in ecology for the automated detection of animals in areas which are difficult to access or where
1088 ground counts are ineffective, e.g. emperor penguins (Fretwell et al., 2012) and southern right
1089 whales (Fretwell et al., 2014). The capability of image processing software to incorporate texture,
1090 shape and context into analyses has reduced error rates and improved object classification (Laliberte
1091 and Ripple, 2003; Peña-Barragán et al., 2011).

1092

Image segmentation

1093

Discrimination of animals in imagery depends more on their contrast with the environment than on
1094 spatial and spectral resolution (Laliberte and Ripple, 2003). Spatial separation of individuals can also

1095 affect population count estimates (Cunningham et al., 1996). Distinguishing objects of interest from
1096 the background in image processing is done by image segmentation. Thresholding is the simplest
1097 and most common image segmentation method, whereby pixels are categorised into foreground or
1098 background based on their intensity value relative to a threshold value or band of values.
1099 Thresholding has had some success on small scales, with population estimations usually highly
1100 correlated with ground or manual counts (Laliberte and Ripple, 2003; Trathan, 2004; Fretwell et al.,
1101 2014; Table 4.1) and in some cases outperformed other automated methods including supervised
1102 and unsupervised classifications (Fretwell et al., 2014). Filtering and image enhancements before
1103 processing can be beneficial for some imagery, particularly for images containing small animals that
1104 may have been missed due to the small number of pixels they occupy. However, the benefits that
1105 arise from filtering are situation dependent as it has the potential to misclassify the background or
1106 animals, as was the case for caribou (Laliberte and Ripple, 2003), or to distort the spectral
1107 information of pixels depicting animals (Yang, 2012). Objects in the environment also have the
1108 potential to affect automated counts through misclassification of pixels (Laliberte and Ripple, 2003).

1109 **Supervised classification**

1110 One of the most common object identification techniques in the analysis of remotely sensed
1111 imagery is supervised classification (e.g. Barber-Meyer et al., 2007; Fretwell et al., 2012; Fretwell et
1112 al., 2014; LaRue et al., 2014). Despite its popularity it has not been as successful at detecting
1113 individual animals as other methods (Fretwell et al., 2014; LaRue et al., 2015; Table 4.1). Users
1114 classify objects or areas of interest that the image processing software can then use as training sites.
1115 The mean and variance of the spectral signatures of the training pixels are then used to classify the
1116 remaining pixels in the image. Regression methods estimate populations by relating classified pixels
1117 or areas of animals to population size (Barber-Meyer et al., 2007; Fretwell et al., 2012). The output
1118 quality of supervised classification depends heavily on the user's knowledge and ability to classify
1119 areas correctly, distinctiveness of the spectral signature of chosen classes, and the accurate
1120 representation of the variability of a class within the training data (Hussain et al., 2013). This
1121 technique may also be used in combination with unsupervised classification, which can be used as a
1122 preparatory step to analyse spectral classes of the image to determine how well separate classes can
1123 be defined.

1124 **Spectral reflectance**

1125 Studies have assessed the spectral reflectance of different species to support the identification of
1126 livestock species and other animals in satellite imagery. Older studies had little success in

1127 determining a unique spectral signature for deer (Trivedi et al., 1982; Wyatt et al., 1985). More
1128 recently, others have found spectral separability between large livestock species including sheep,
1129 pigs, horses and cows, and between mammals and landscapes (Bortolot and Prater, 2009; Terletzky
1130 et al., 2012; Yang, 2012), but poor spectral separability between mammals and shadows (Yang,
1131 2012). The referenced studies all use different source data, methods to detect spectral separability
1132 and assessment of spectral reflectance overlap between different animal species, and animals and
1133 background.

1134 Spectral signatures, obtained from training data, can be applied to new imagery for the classification
1135 of pixels into predefined classes (Turner et al., 2003; Bortolot and Prater, 2009). This has been
1136 demonstrated with high accuracy (>90%) in some studies using hyperspectral data, even with
1137 relatively coarse image resolution of 2m (Bortolot and Prater, 2009). No studies have conducted this
1138 type of analysis over spatial scales larger than a few square kilometres and most have been proof-of-
1139 concept studies (Bortolot and Prater, 2009; Yang, 2012; Terletzky, 2013). Hyperspectral imaging has
1140 been used to identify vegetation species in forest canopies for different types and growth stages of
1141 horticultural crops (Peña-Barragán et al., 2011; Peña et al., 2014), and differences in animal species
1142 (Bortolot and Prater, 2009).

1143 **Short time image differencing and change detection**

1144 Change-detection methods using multi-temporal imagery have most commonly been used to
1145 determine land use changes such as deforestation due to the consistent, repetitive nature of
1146 imagery (Singh, 1989). Recently several studies have used these techniques to estimate animal
1147 populations. If spectral differences of pixels with and without animals can be quantified then the
1148 movement of animals would change the spectral reflectance of the pixel between images taken at
1149 different times. Terletzky and Ramsey (2014) published one of the first studies of standard change
1150 detection techniques to estimate numbers of large animals. Their method was semi-automated,
1151 based on a principal components analysis of two aerial photography images captured on the same
1152 day. Polygons representing two livestock species, namely cattle and horses, were generated by using
1153 the difference in the first principal component of the images and heuristic thresholding. Using this
1154 method, 82% of animals were correctly identified, though a relatively high mean commission error
1155 was recorded of 53%. Commission error was attributed to small mismatches in the alignment of the
1156 separate images, misidentification of shadows, and animals grouped together. Oishi and Matsunaga
1157 (2014) also describe automated methods for detecting animal movement through snow from
1158 overlapping aerial photographs. Their three algorithms for image registration and detection of
1159 moving animals successfully detected five of six cattle and a deer in two unmanned aerial vehicles

1160 (UAV) images covering approximately 0.6km² each. A walking human could also be differentiated
1161 from more than 2,100 objects. Misclassification increased with more stringent thresholding, but
1162 applying the algorithms reduced the person hours required to survey animals by visual inspection by
1163 over 90%.

1164 Manual change detection techniques have been used to count polar bears from high resolution
1165 satellite imagery (Stapleton et al., 2014). Objects that could be polar bears were referenced against
1166 images taken at a different time to eliminate features that remained constant. This method had a
1167 reasonably high accuracy (n= 94; 95% CI=92-105) when compared to counts from established aerial
1168 survey methods (n=102, CI=69-152) (Stapleton et al., 2014). Observers experienced with the species
1169 biology had higher detection probabilities of 96% compared with 42% for a less experienced
1170 observer (Stapleton et al., 2014). Automated methods were recommended for large-scale studies of
1171 more than a few square kilometres, largely due to the extensive time investment required for
1172 manual counts.

1173 Change detection in landscape analysis is often applied to low- and medium resolution imagery. It
1174 has been unsuccessful for very high resolution (VHR) imagery (sub-meter resolution), which has
1175 been attributed to difficulties in accurately geo-referencing images. Higher variability in spectral
1176 reflectance also makes it more difficult to discriminate between real changes and background noise
1177 (Hussain et al., 2013). Despite the lower resolution requirements, the cost of imagery may still be a
1178 constraint, as images should be obtained at the same time of day and less than a week apart to
1179 avoid changes in sun angle and vegetation (Terletzky and Ramsey, 2014).

1180 **Object-based image analysis**

1181 High resolution imagery has relatively high internal variability, and where individual pixels are no
1182 longer characteristic of the object of interest, such imagery can give rise to a 'salt and pepper' effect
1183 with pixels erroneously classified into maximum or minimum values, and a decrease in the
1184 classification accuracy of pixel-based methods (Ke et al., 2010; Yang, 2012). The majority of animal
1185 identification studies have used pixel-based approaches, but object-based image analysis (OBIA)
1186 methods have been suggested as valuable alternatives. The unit for analysis in OBIA is the object,
1187 and neighbouring objects can provide context in spatial relationships, texture, and shape (Hussain et
1188 al., 2013). Object-based methods have been mainly used for land cover and land use classification
1189 (e.g. Mathieu et al., 2007; Ke et al., 2010; Peña-Barragán et al., 2011) where they are generally
1190 considered superior to traditional pixel-based methods because they reduce spectral overlap
1191 between classes, can incorporate expert knowledge, consider both spectral and spatial information,
1192 and produce greater classification accuracy (Yang, 2012). This superiority may depend on the image

1193 resolution and choice of algorithm (Duro et al., 2012). OBIA can use contextual information in the
1194 classifier algorithm, such as crop patterns, management techniques and crop calendars, and
1195 combined with machine learning techniques can increase the accuracy of crop discrimination even
1196 for crops which have similar spectral characteristics (Peña et al., 2014). Such techniques have been
1197 used for fine scale habitat resource mapping for invasive species (Recio et al., 2013) and could be
1198 applied to assess spatial variability in pasture quality.

1199 Object-based approaches have been successful in the detection of individual animals (Groom et al.,
1200 2011; Groom et al., 2013; Chrétien et al., 2016). An automated OBIA using quadtree image
1201 segmentation and sequential object brightness thresholding to estimate lesser flamingo
1202 (*Phoeniconaias minor*) abundance demonstrated that the accuracy of detection was >99% when
1203 compared with independent visual counts at a 525ha site, with more than 81,000 individuals
1204 identified. The whole process was implemented as an algorithm (Groom et al., 2011). Generally the
1205 algorithm underestimated flamingos compared to visual counts and had higher detection rates and
1206 lower omission and commission rates over water than on land due to the greater contrast of
1207 individuals with the background (visual counts as a percentage of automated counts, water: range -
1208 0.34% to -5.68%; land: range -0.99% to -21.82%). OBIA approaches were presumed superior to pixel-
1209 based approaches because the variability in brightness between adjoining images, and the lower
1210 contrast between birds and background on the land than in the water, limited the use of pixel-based
1211 approaches across the entire site. One of two studies that have compared pixel and OBIA methods
1212 for the detection of animals, namely migrating mammals in African savannahs, failed to find a
1213 significant difference in performance of the methods with respect to population estimation and the
1214 correct identification of pixels (Yang, 2012). The average omission error was 12% and commission
1215 errors 11-15% for pixel-based approaches compared with omission errors of 13-16% and
1216 commission errors of 7-13% for the object-based approaches. The second study found OBIA
1217 methods using thermal infrared multispectral data superior to pixel-based approaches which proved
1218 ineffective at detecting white-tailed deer. In contrast OBIA methods had a detection rate of 0.5
1219 when compared with ground counts and was considered to be comparable to aerial surveys
1220 (Chrétien et al., 2016).

1221 **Manual methods for the detection of objects from remotely** 1222 **sensed imagery**

1223 Manual counting of animals in remotely sensed imagery has been used for decades (e.g. Kadlec and
1224 Drury, 1968; Leonard and Fish, 1974; Löffler and Margules, 1980) but can be costly, time consuming,

1225 subjective and unrepeatable (Terletzky and Ramsey, 2014) and is therefore difficult to apply
1226 successfully over spatial scales larger than a few square kilometres. Despite this, manual counts from
1227 satellite imagery have been shown to be reasonably accurate (e.g. LaRue et al., 2011; Lynch et al.,
1228 2012; McMahon et al., 2014), with highly significant correlations demonstrated when assessed with
1229 ground counts ($r = 0.98$, $n = 1000$, LaRue et al., 2011; $r\text{-squared} = 0.91$, $n = <2000$; McMahon et al.,
1230 2014) and this is likely to improve as the resolution of the source material improves. However those
1231 studies that have been ground-truthed often report underestimating populations (e.g. McMahon et
1232 al., 2014), due to visibility bias in aerial photographs whereby animals are hidden from observers
1233 (Jachmann, 2002). Environmental homogeneity, body size and ability to differentiate individuals from
1234 their background all have a substantial impact on accuracy and omission and commission error
1235 rates (Jachmann, 2002; Terletzky and Ramsey, 2014). Observer confidence can also affect
1236 misclassification rates of species (McClintock et al., 2015). Issues with manual counting make the
1237 development of automated and semi-automated techniques necessary before large geographic
1238 areas can be reliably surveyed. Although there is general recognition that visibility bias can affect
1239 outcomes (see, e.g., Jachmann 2002, McMahon et al. 2014), comparatively little work has been done
1240 to develop models that will handle imperfect detection, White (2005) and Miller et al. (2011) being
1241 notable exceptions.

1242 **Crowd sourcing applications and software**

1243 One alternative approach for livestock estimation with remotely sensed imagery is to use
1244 crowdsourcing platforms. Crowdsourcing as outlined by Papadopoulou and Giaoutzi (2014) is
1245 “nowadays extensively used to describe a process, including methods and techniques of data
1246 collection and info generation, that involves large groups of users, who are not organized centrally
1247 and generate shared content.” The idea is a relatively new approach to knowledge acquisition and
1248 has the following steps: 1. release of the problem and relevant information to the crowd, 2.
1249 solutions provided by the crowd, 3. the evaluation and selection of the most appropriate solutions,
1250 and 4. the uptake of the solution by the organisation who originally proposed the problem
1251 (Papadopoulou and Giaoutzi, 2014). Crowdsourcing has been particularly beneficial for both the
1252 collection and generation of spatial data (Papadopoulou and Giaoutzi, 2014). As an example, Geo-
1253 Wiki is an internet crowd sourcing platform where volunteers improve the quality of land-cover
1254 maps using datasets in Google Earth (<http://www.geo-wiki.org/>). Livestock Geo-Wiki has been
1255 proposed for the validation of livestock systems and distributions (Robinson et al., 2014). Two
1256 examples of crowdsourcing platforms are Mechanical Turk
1257 (<https://www.mturk.com/mturk/welcome>) and Tomnod (www.tomnod.com). The latter is

1258 specifically for satellite imagery applications and was used recently in the search for missing flight
1259 MH370, whereby volunteers were recruited to search satellite imagery for plane debris.

1260 Public participatory GIS (PPGIS) is another form of crowdsourcing that uses published maps to allow
1261 people to interact and engage with real environmental issues (Papadopoulou and Giaoutzi, 2014)
1262 and was recently used in New Zealand for vegetation mapping in a study covering 3.2 million
1263 hectares (Brown, 2012). Participants were recruited through mail surveys, advertising and
1264 interaction with visitors at conservation sites. The participant's mapped locations of native
1265 vegetation were then compared with the satellite derived NZ Land Cover Database (LCDB) to assess
1266 the accuracy of the publically generated data. Spatial error, measured as locations with no native
1267 vegetation within 1000 m, was low at around 6% and those who volunteered had lower error rates
1268 than people who had been randomly sampled (Brown, 2012).

1269 Crowd sourcing could be used for livestock estimation. However, consideration would need to be
1270 given to cost and privacy issues and the application may be more realistic for the validation of
1271 predictive modelling.

1272 **Limitations and constraints for direct counts of animals**

1273 Estimation of populations by direct count from satellite imagery over large spatial scales has not
1274 been successfully undertaken for any animal species. Individual animals of any size cannot yet be
1275 identified accurately from remotely sensed spatial data (Pettorelli et al., 2014), except in relatively
1276 small areas of no more than a few square kilometres (e.g. Laliberte and Ripple, 2003; Terletzky,
1277 2013; McMahon et al., 2014) and homogenous environments (e.g. Groom et al., 2011, white
1278 flamingoes against water and land; Fretwell et al., 2012, penguins in Antarctica). The same
1279 constraint applies for objects that imply the presence of different livestock species such as water
1280 troughs, gates, stockyards, and cattle stops, or absence of livestock, such as home gardens or high
1281 density buildings. The major limitations and constraints are the relatively low accuracy of automated
1282 detection techniques across large spatial scales, and the cost of high resolution data. Many of the
1283 algorithms are experimental and not suitable for application across broad geographical areas
1284 (Kuemmerle et al., 2013), but they have been successful on small scales with several species and
1285 imagery types (Laliberte and Ripple, 2003; Table 4.1) showing great future potential.

1286 Logistical costs of obtaining, processing and analysing remote imagery, staff training and software
1287 are still prohibitively expensive for many governments and organisations (Pettorelli et al., 2014).
1288 Despite the declining costs of high resolution imagery and increasing computational power, the costs

1289 and technical capabilities needed are still substantial, especially when considered with the software
1290 and hardware required to analyse even small quantities of remotely sensed imagery (Turner et al.,
1291 2003). The age of the data is also a consideration for high resolution imagery. More recent satellite
1292 imagery is available, but at either high cost or low resolution. Caution must be applied to the
1293 analysis of imagery taken from different satellite sensors, with different pixel sizes, scales, or
1294 numbers of bands and spectral information (Hussain et al., 2013). Problems like these may be
1295 unavoidable when working over large spatial scales where many different image sets are required.

1296 Corroborating the results of imagery analysis with ground-truth data are a key requirement because
1297 factors including atmospheric conditions which can influence spectral results may affect the output
1298 (Turner et al., 2003). All methods that detect individual animals from remotely sensed imagery will
1299 also suffer from detection errors from animals in sheds or obscured by vegetation and seasonal
1300 effects when young animals are present. It may be difficult to differentiate individuals of a species
1301 and other objects based purely on size and location (e.g. in a paddock).

1302 **Estimating environmental predictors and proxies from satellite** 1303 **imagery**

1304 The remote sensing techniques used for plant diversity and distributions cannot be applied to fauna
1305 in the same way due to their mobility and cryptic nature, thereby requiring proxies and surrogates of
1306 their presence (Leyequien et al., 2007). Satellite-derived predictor data are increasingly used in
1307 species abundance and distribution modelling, but are generally underutilised by modellers
1308 (Buermann et al., 2008). Significant attention in ecological modelling has been applied to the
1309 modelling techniques and the quality of occurrence data, but relatively little focus has been placed
1310 on the environmental predictor data used to build such models (Buermann et al., 2008). Ecological
1311 models built on both remote sensing and climatic variable datasets have performed better across
1312 various species than either set alone and this is especially so for species considered habitat
1313 generalists and with large home ranges (Buermann et al., 2008), as is the case for most livestock
1314 species.

1315 Recently studies have begun to take greater advantage of environmental proxies derived from
1316 remotely sensed imagery to both estimate and downscale wildlife and livestock populations,
1317 densities and distributions (e.g. Prosser et al., 2011; Fretwell et al., 2012; LaRue et al., 2014;
1318 Robinson et al., 2014). For example, several studies mentioned previously have used the area of
1319 guano on ice sheets as a surrogate for nest density to estimate penguin abundance (Fretwell et al.,

1320 2012; Lynch et al., 2012; LaRue et al., 2014; Lynch and LaRue, 2014) and Prosser et al. (2011) used
1321 phenology and elevation information collected from the MODIS (Moderate Resolution Imaging
1322 Spectroradiometer: an instrument aboard both the Terra and Aqua satellites), in regression
1323 modelling for poultry. Other studies have used satellite derived environmental proxies such as
1324 elevation, land cover and productivity to estimate species richness (Seto et al., 2004; Coops et al.,
1325 2009) and for fine-scale landscape resource modelling of invasive species habitat use (Recio et al.,
1326 2013).

1327 Primary productivity, climate and habitat variables are readily available and biologically relevant for
1328 most livestock species, and many can be derived from satellite imagery (Turner et al., 2003). The
1329 range of predictor datasets that can be extracted from remotely sensed imagery includes vegetation
1330 density and seasonality, air temperature, humidity, vapour pressure deficit, surface wetness, soil
1331 moisture, roughness and topography, and elevation (mean and SD) (Goetz et al., 2000; Buermann et
1332 al., 2008). Land surface temperature from thermal bands and surface rainfall from cold cloud
1333 duration have both been used for livestock population estimation in agronomic models (White et al.,
1334 2001).

1335 Given current availability in high resolution, vegetation index mapping may be one of the greatest
1336 potential applications for remotely sensed imagery in ecological modelling. Pasture growth,
1337 production/biomass, and spatial and seasonal variability have all been successfully assessed from
1338 satellite imagery (e.g. Taylor et al., 1985; Vickery et al., 1997; Moore et al., 1999; Edirisinghe et al.,
1339 2011; Recio et al., 2013). Normalised Difference Vegetation Index (NDVI) is one of the few satellite
1340 derived environmental proxies that has been used extensively in ecological modelling (Buermann et
1341 al., 2008). NDVI is a measure of the photosynthetic activity of vegetation derived from multispectral
1342 data and provides a 'greenness' index, which can indicate vegetation condition and quantity (White
1343 et al., 2001; Turner et al., 2003). This could be used to measure pasture growth, primary productivity
1344 and potentially even grazing intensity (e.g. Kawamura et al., 2005; Paudel and Andersen, 2010;
1345 Jansen et al., 2016), particularly relevant for grazing livestock species. Enhanced vegetation index
1346 (EVI) is similar but enhances the vegetation signal by incorporating the blue band to reduce
1347 atmospheric influences and canopy background, giving higher sensitivity in areas with high biomass
1348 (Huete et al., 2002). Growing season length can be calculated by the period of time EVI is above a
1349 given value. Both NDVI and EVI are calculated from data collected from MODIS satellites. These
1350 satellites collect imagery across the entire earth every 1-2 days, recording up to 36 spectral bands at
1351 1000m, 500m and 250m resolution. For the vegetation index measure, Robinson et al. (2014) in their
1352 Gridded Livestock of the World (GLW) models considered the maximum NDVI rather than the mean
1353 to be the most suitable model predictor for livestock based on the assumption that areas with low

1354 maximum cover would rarely if ever be suitable for grazing livestock species, but those with a low
1355 mean may be seasonally suitable.

1356 There has been limited success in object proxies for detecting the presence of different livestock
1357 species such as water troughs, fences, and shearing or milking sheds. In addition, intensification of
1358 animal production is on the rise in many countries, with increasing off-land production (animals
1359 intensively housed in buildings). In those areas, animal detection becomes impossible, unless one
1360 shifts attention to the detection of infrastructure. For example, in the US, a hog farm is typically a
1361 building with specific size, with a big silo for stocking the feed nearby, and a lagoon to collect the
1362 manure, all of which could potentially be detected using remote sensing. Adding such contextual
1363 information and requirements to automated detection algorithms could improve detection rates for
1364 specific building types. The purpose of farm buildings and the animals housed in them could also be
1365 differentiated from the roof shape and objects on the roof such as vents or chimneys, if the
1366 resolution is high enough (Petráš, 2012).

1367 Several different approaches for the detection of buildings through automated techniques have
1368 been published and while many show reasonable success, others have issues including high
1369 computational cost or difficulty detecting dark roofs (Nosrati and Saeedi, 2009). Automated
1370 methods that used contextual, structural and spectral information to detect buildings from satellite
1371 imagery had a detection rate of 73% and 59% quality in a small subset of imagery comprising 20 city
1372 blocks (Jin and Davis, 2005). For hog farms with defined spatial patterns, a template or pattern
1373 matching process may be effective, whereby automated detection of clearly defined building shapes
1374 occurs through recognition of polygon shapes (Petráš, 2012). Such an approach has been successful
1375 for rural buildings with a matching rate of 78%, but a misclassification rate of 15%, and difficulties
1376 may arise matching shapes which are not right angles (Rainsford and Mackness, 2002). An
1377 algorithm developed to identify animal farm buildings in Italy correctly detected 70% of the 75
1378 manually detected buildings but 24% of detected buildings were false positives. The algorithm was
1379 very successful for detecting rectangles, but subsequent filtering to reduce the false positives also
1380 removed genuine positives (Petráš, 2012). Application of these techniques for building detection has
1381 not been successful over large spatial areas but may have significant future potential, particularly for
1382 estimating intensively farmed livestock populations.

1383 Remote sensing and associated technologies have enormous potential for epidemiological
1384 applications including the monitoring of disease vectors (Goetz et al., 2000). The relative abundance
1385 and/or distributions of vectors have been predicted by studies using information on environmental
1386 conditions derived from remotely sensed imagery and reported disease incidents (Beck et al., 1994;
1387 Hay et al., 1998; Kalluri et al., 2007). Remote sensing provides data for monitoring and modelling the

1388 environmental conditions that are beneficial to the reproduction, dispersal and survival of disease
1389 vectors (Goetz et al., 2000; Kalluri et al., 2007).

1390 **Species distribution models and downscaling**

1391 Species distribution and downscaling models represent the relationship between species and their
1392 environments and have been rapidly rising in popularity for ecological and conservation applications
1393 (e.g. Elith et al., 2006), and more recently for agricultural (e.g. Van Boeckel et al., 2012; Robinson et
1394 al., 2014) and disease research (Bogh et al., 2007; Stevens et al., 2013). Most of the models can be
1395 used to either estimate populations for which there are no data on an area or to downscale
1396 population data at large scales to smaller localised populations. Such models have become more
1397 advanced in recent decades, in parallel with improved data availability, increased computing
1398 capacity, and software and statistical developments (Guisan and Thuiller, 2005). Several recent
1399 studies have used species distribution models for livestock species and commercial poultry
1400 population estimation over large spatial scales by disaggregating census data from statistical
1401 relationships with predictor variables, many of which have been derived from remotely sensed
1402 imagery and include climatic time-series data (the time-series capturing the seasonality through
1403 Fourier processing), other environmental variables such as elevation and slope, and anthropogenic
1404 variables like human population density, and accessibility – estimated as travel time to various
1405 population centres (e.g. White et al., 2001; Neumann et al., 2009; Prosser et al., 2011; Temme and
1406 Verburg, 2011; Van Boeckel et al., 2011; Van Boeckel et al., 2012; Robinson et al., 2014). Overall
1407 these studies report a range of accuracies in predictions varying with species, modelling type,
1408 regions and available data. They have been undertaken at a range of spatial scales from local and
1409 regional, to global in the case of the Food and Agricultural Organisations (FAO) Gridded Livestock of
1410 the World (GLW) (Robinson et al., 2007b). There are many different types of species distribution
1411 models, from machine learning techniques (e.g. GARP, MAXENT) to correlative methods (e.g. GLM,
1412 BRT). A growing literature evaluating different modelling techniques has arisen in response to the
1413 large number of models available, types of data and research questions (Elith and Graham, 2008).

1414 The GLW includes distributions worldwide at 1km resolution for major livestock species: cattle,
1415 goats, buffaloes, sheep, pigs and poultry (Robinson et al., 2014) and is suitable for both estimating
1416 and downscaling population data. The GLW data has applications in animal health research
1417 (Franceschini et al., 2009) including for predicting the risk of avian influenza infection (Gilbert et al.,
1418 2014) and the incidence and distribution of foot and mouth disease (FMD) (Sumption et al., 2008).
1419 The GLW uses explanatory variables from environmental, demographic and climatic data, several of

1420 which are derived from satellite imagery (e.g. rainfall, temperature, vapour pressure deficit,
1421 vegetation cover and human population density). Stepwise multiple regression analyses have been
1422 conducted on a regional basis due to the inconsistency of predictors across different geographical
1423 and ecological regions. For a detailed overview of the GLW methods see Robinson et al. (2014).
1424 Several studies have adopted the GLW methodology to estimate poultry stock in Asia (e.g. Prosser et
1425 al., 2011; Van Boeckel et al., 2011; Van Boeckel et al., 2012). Avian densities are modelled against
1426 agro-environmental predictor variables, which affect the geographic distribution of bird stock. The
1427 statistical relationships between these variables can downscale known population numbers at
1428 administrative levels to distributions as a raster layer across the countries of interest (Van Boeckel et
1429 al., 2011).

1430 Gathering and processing the initial data for models constitutes the largest resource and time
1431 component (Robinson et al., 2014). Due to the commercial nature of the livestock industry, the
1432 ecological and environmental requirements and limitations for different species are well
1433 understood, and relevant training and predictor data are often readily available. A key for analysis is
1434 to ensure that the environmental and species data are on the same scale and that the scale is
1435 relevant to the study questions (Turner et al., 2003). Most livestock distribution models are set up to
1436 enable them to be rerun when new and/or more accurate data become available (e.g. Robinson et
1437 al., 2014).

1438 Finally, the goal of the livestock modelling exercises outlined above is to construct models that
1439 predict unknown quantities, rather than models that estimate parameters or the influence of a
1440 particular predictor variable. Recent developments in statistical research have identified modelling
1441 approaches that have considerably better statistical performance when prediction is the goal, for
1442 example, the LASSO (least absolute shrinkage and selection operator) (Tibshirani, 1996; see Hastie et
1443 al., 2009 for a readable exposition and; Verbesselt et al., 2009 for an example application) and
1444 Random Forests (Breiman, 2001). The statistical approach of 'Random Forests' are now routinely
1445 used to downscale human population data (Stevens et al., 2015), and apparently show better
1446 performances than stratified linear regression for downscaling of livestock data too (Nicolas et al.,
1447 2016).

1448 **Software**

1449 ArcGIS and its associated tools (including supervised classification) are used in the majority of
1450 projects that aim to identify objects from remotely sensed imagery in ecological and environmental
1451 applications. Its popularity is likely due to the inclusion of geospatial and mapping tools without the

1452 requirement of alternative software applications, as well as users' familiarity with the program
1453 within these disciplines. Some studies have combined the use of ArcGIS with specific image
1454 processing software, such as Fretwell et al. (2014) who used ENVI 5 and ArcGIS to analyse satellite
1455 imagery in the first successful study of remote imagery to count whales. ENVI is a commercially
1456 available image processing program that is integrated with ArcGIS, including automated feature
1457 extraction tools. Other studies have borrowed image processing software from biomedical and
1458 health disciplines for image analysis including Laliberte and Ripple (2003; ImageTool) and Groom et
1459 al. (2011; Definiens Developer). Programs for object-based analysis of imagery include SPRING, ENVI,
1460 eCognition and Erdas Objective. ENVI and eCognition are the most commonly used (Yang, 2012). A
1461 common set of classification tools, including supervised classification, nearest neighbour classifiers,
1462 thresholding, and others, appear in most relevant software packages. There has been no direct
1463 comparison of the efficacy of ArcGIS and other image processing software for the identification of
1464 animals or objects in remotely sensed imagery. Ultimately the software choice will come down to
1465 cost, personnel considerations and the goal of the project.

1466 Valuable and underused alternatives may exist in computer pattern recognition and image
1467 processing software from other disciplines where it has been extensively and successfully used, such
1468 as biomedical science and engineering (Laliberte and Ripple, 2003). Many image processing
1469 software programs are available including Microsoft Visual Studio, C#, C++, OpenCV, Mathematica,
1470 Python, Matlab and ImageJ. The Image Processing Toolkit in Matlab was used for the estimation of
1471 penguin population size from aerial photography (Trathan, 2004). Aside from this one example
1472 Matlab is rarely used in environmental and ecological science for image processing despite its wide
1473 application in other disciplines such as engineering. A disadvantage of Matlab is its processing speed,
1474 which is relatively slow when processing large quantities of images. But, once the algorithm is
1475 implemented effectively it could be converted to a different programming format such as C or C++
1476 to speed up processing. Another program, ImageJ (e.g. Figure 4.1), is relatively easy to use and many
1477 people have developed macros that can be used or modified.

1478 R is a frequently used program for the implementation of statistical models including species
1479 distribution models. R is an open-source programming environment with associated packages for
1480 statistical computing and graphics (R Core Team, 2014). Packages in R now also have the capability
1481 to manipulate and analyse spatial data, and as a result, running GIS applications in R is becoming
1482 more commonplace. One of the major advantages with R is that the models can be easily updated
1483 and rerun when new data become available and modified for new scenarios, such as currently
1484 implemented in the GLW (Robinson et al. 2014).

1485 a)

1486

1487

1488

1489

1490

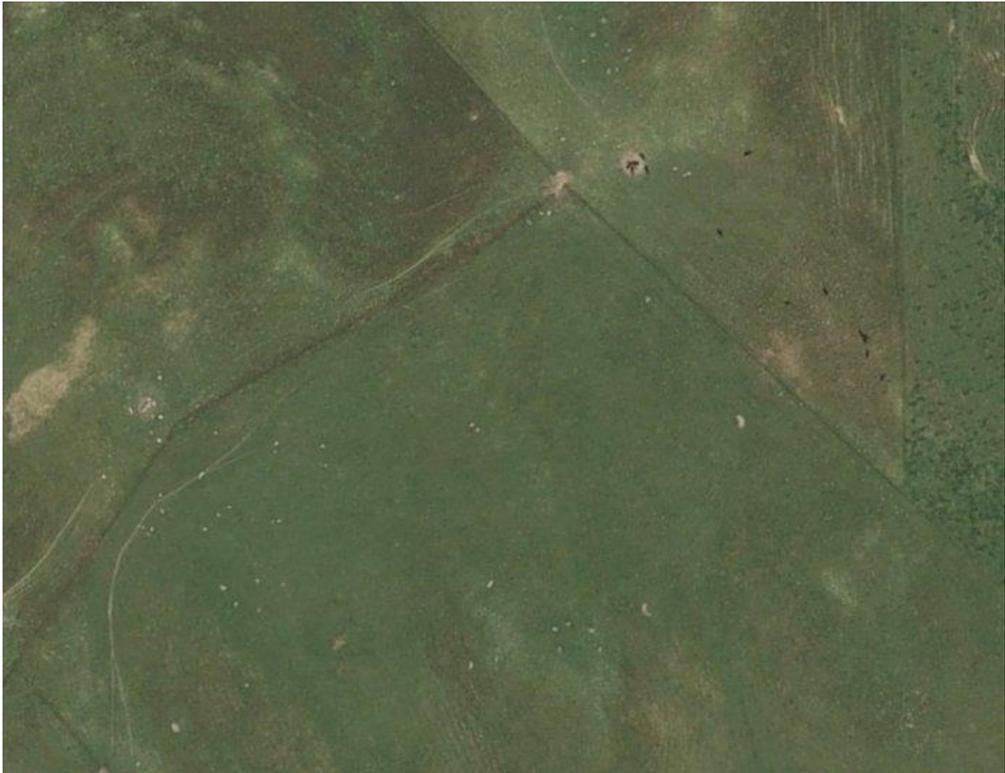
1491

1492

1493

1494

1495



1496 b)

1497

1498

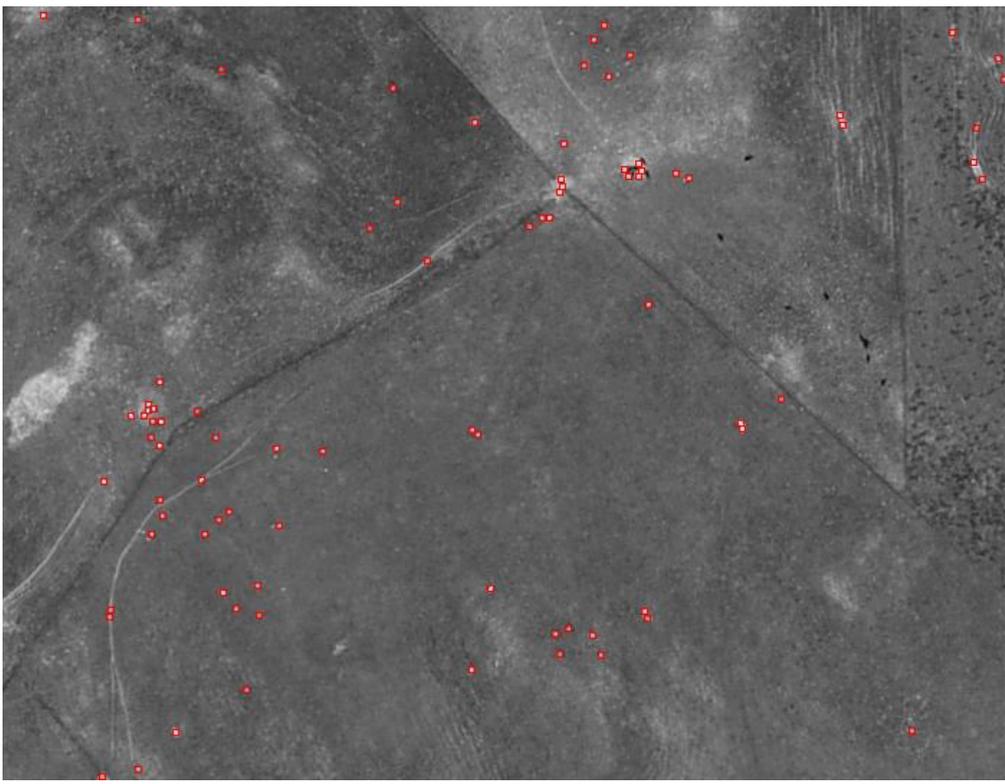
1499

1500

1501

1502

1503



1504 c)

1505

1506

1507

1508

1509

1510

1511

1512

1513

1514



1515

1516 d)

1517

1518

1519

1520

1521

1522

1523

1524

1525

1526



1527

1528 **Figure 4.1::** Using Particle Tracker (Sbalzarini and Koumoutsakos, 2005) in ImageJ designed for video

1529 *imaging in cell biology to detect sheep; a) original image of farming area in NZ at 0.4m resolution; b)*

1530 *particle tracker program for automated detection of 'particles' to detect sheep; c) a representative*

1531 *area of peri-urban small scale farming in NZ at 0.4m resolution; d) This relatively less homogenous*

1532 *landscape has less success with particle tracker plugin for detecting sheep. Images obtained from*
1533 <http://data.linz.govt.nz>

1534 **Future directions**

1535 Species distribution models (SDM) with a selection of predictor data derived from the analysis of
1536 remotely sensed imagery is the recommended tool in the short to medium term for estimating
1537 livestock distribution, and is investigated in Chapters 6 and 7. The advantage of SDM and
1538 downscaling techniques is that they can be continually updated with new data, can predict over
1539 large spatial scales, and are relatively easy to understand and modify. The identification of individual
1540 animals will be increasingly feasible with technological advances, and higher image resolution of less
1541 than 2m and lower costs (Turner et al., 2003). This is demonstrated with an increasing number of
1542 studies completing proof-of-concept work for estimating animal counts, and a rapidly expanding
1543 knowledge and technological base in image processing. Difficulties such as the detection of different
1544 livestock species, e.g. cows versus horses, in non-homogenous environments will need to be
1545 overcome.

1546 Two data types could be considered in future to support livestock estimation models; LIDAR data
1547 and short-time repeat image coverage. LIDAR data may offer advances in livestock estimation by
1548 detecting livestock related objects, including such things as fence height, buildings, and even
1549 individual animals. LIDAR sensors send out electromagnetic radiation pulses and measure the
1550 returning radiation from the target, providing the unique capability to measure 3D vertical space
1551 (Pettorelli et al., 2014). LIDAR data has previously been used for measuring vegetation (e.g. Lefsky et
1552 al., 2002; Streutker and Glenn, 2006) and urban structures (e.g. Rottensteiner, 2003; Zhang et al.,
1553 2006), forest species composition (e.g. Ke et al., 2010), and at high resolutions can detect the
1554 position of branches and leaves (Turner et al., 2003). It has also been proposed as a way of remotely
1555 assessing the body condition of cattle (CSIRO, 2015). The increasing accessibility and short-time
1556 repeat coverage from satellite imagery may offer solutions in the detection of individual livestock.
1557 This data can be utilised through short-time interval change detection analysis, in which animals can
1558 be detected by the change in spectral reflectance of pixels (Terletzky and Ramsey, 2014) resulting
1559 from animal movements relative to the static background.

Case study: Estimating livestock populations for biosecurity applications in NZ

New Zealand's agricultural industry is pivotal to the national economy, contributing more than half of export revenue and generating almost a fifth of national GDP. New Zealand's isolation combined with some of the strictest biosecurity regulations in the world have allowed it to remain free of significant agricultural pests and diseases affecting productivity and export capacity. Biosecurity investment by the NZ government is substantial but a single, accurate, up-to-date database of farm livestock populations is currently lacking, though crucial for preparedness and response to agricultural diseases. The estimated cost of a foot-and-mouth disease (FMD) outbreak in NZ involving several hundred farms from initial detection is more than NZ\$16 billion over eight years (Forbes and van Halderen, 2014).

Preparedness for a disease outbreak consists of border protection and plans for eradication response (Sanson, 1994). Containment of an incursion such as FMD requires knowledge of where livestock disease hosts are held. This is key as rapid eradication will substantially reduce the economic impact of FMD (Belton, 2004; Carpenter et al., 2011). In 2013 the number of livestock susceptible to FMD held in NZ was estimated at 6.5 million dairy cows, almost 31 million sheep, 3.7 million beef cattle, 316,000 pigs, and 1 million deer (Stats NZ, 2013). These totals exclude livestock held for purposes other than commercial farming, present on the estimated 60,000 - 80,000 small-scale peri-urban farming blocks across NZ which have multiple livestock species, often at low densities and kept for non-commercial reasons. While farm boundary and ownership information is available for all of these properties, the species present and their numbers are not recorded for the majority of the population. Given the importance of reliable intelligence relating to livestock populations, this is an area of ongoing work for the government's biosecurity preparedness and planning, enhancing capacity for effective containment or eradication of disease vectors in the event of a biosecurity incursion.

Given current data and technology, identification of individual animals from remotely sensed imagery is currently not yet feasible for livestock estimation. Instead, modelling of environmental predictors against available farm-level databases may be practical for estimation of livestock in NZ and elsewhere in the medium term. Long term actions must focus on alignment of data sources to provide a unified picture of the animal population (Jewell et al., 2015). Estimation of objects and environmental data from remotely sensed imagery that indicate the presence of certain livestock species can be used effectively to inform these ecological models. Unfortunately these correlative

1592 predictive models are limited in their capacity to predict uncharacteristic and stochastic variations in
1593 distribution and abundance, which may be of substantial importance to an outbreak response. Such
1594 models can be effectively utilised for planning, policy and simulations as outlined in Deliverable 1 in
1595 the previous Chapter, but even small error rates in models used for real outbreak responses could
1596 have severe implications. Satellite and aerial derived environmental predictor data are likely to
1597 provide greater predictive power for these models than will climatic variables alone (Buermann et
1598 al., 2008).

1599 Several data layers derived from remotely sensed imagery are already available. Primary
1600 productivity is generally derived from a vegetation index produced from multispectral data such as
1601 NDVI (White et al., 2001; Turner et al., 2003). NZ is covered by 5 MODIS (Moderate Resolution
1602 Imaging Spectroradiometer) tiles and all imagery and derived products including EVI and NDVI are
1603 freely available to download.

1604 The NZ Land Cover Database (LCDB) is derived from satellite imagery and classifies land cover into 33
1605 target classes across the whole of NZ's surface area with a resolution of 1 hectare. Four classes
1606 would be relevant to grazing livestock density and abundance models: high producing exotic
1607 grassland, low producing grassland, depleted grassland and tall tussock grassland. The accuracy of
1608 LCDB (version 3) is just over 96% (Landcare Research, 2015). It is free to download from a NZ
1609 government website (<https://iris.scinfo.org.nz>).

1610 Satellite imagery can be used to identify objects that indicate the presence of certain livestock and
1611 some GIS layers are currently being developed by Land Information NZ (LINZ) for objects including
1612 cattle stops, stockyards, fence-lines and water troughs. These datasets are incomplete but are being
1613 updated regularly from satellite imagery.

1614 Species distribution models are assessed in Chapters 6 and 7 with the inclusion of some of these
1615 satellite derived datasets. The area of highly productive exotic grassland from the LCDB was the
1616 consistently highest ranked predictor variable in all models, showing the significance of considering
1617 remotely sensed imagery in ecological modelling (Hollings et al., unpublished data).

1618 **Conclusions**

1619 Planning and policy applications for livestock require up-to-date spatial distribution and abundance
1620 records (Robinson et al., 2007b; Prosser et al., 2011). Governments manage and respond to risks by
1621 having timely and accurate information but few if any countries maintain up-to-date agricultural
1622 census data. Remotely sensed imagery may offer opportunities, though the direct counting of
1623 animals is still problematic. Future developments in remote sensing analysis will improve direct

1624 detection capabilities, including the development of algorithms, the crossover of programs from
1625 other disciplines, availability and cost of data. The eruption of recent studies in the field of direct
1626 animal detection shows the potential of the methods and with advances in technology applications
1627 over large spatial scales should be feasible. Extracting environmental proxy data for use in species
1628 distribution models and downscaling has been used successfully and remains the most consistent
1629 method for obtaining livestock density and abundance estimates.

1630

1631 **Deliverable 3: Overview of New Zealand national level data**
1632 **sources available at the farm-level**

1633 **Mary van Andel¹, Chris Jewell²**

1634 ¹ *Ministry for Primary Industries, Wellington, New Zealand 6140*

1635 ² *Lancaster University, Lancaster, UK, LA1 4YG*

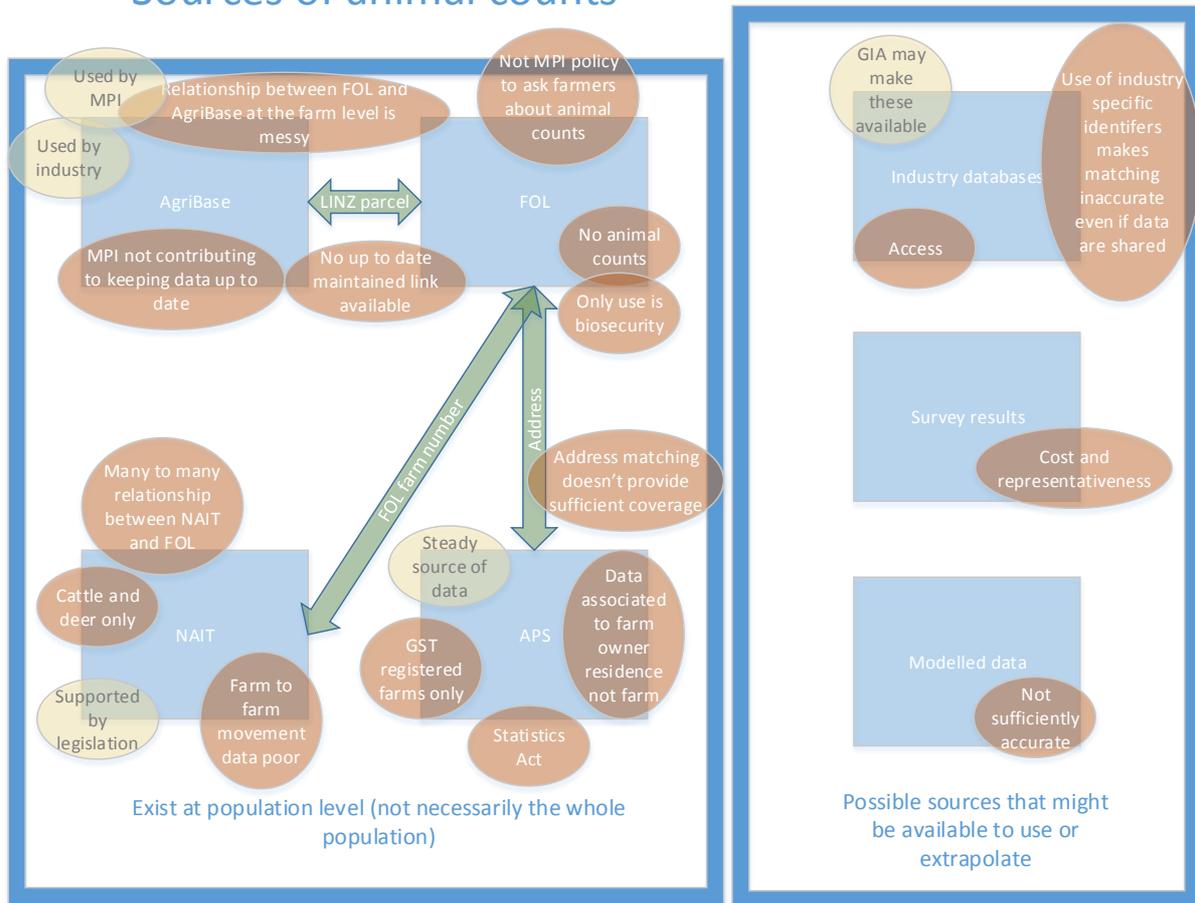
1636

1637 **This chapter has been accepted for publication in the NZ Veterinary Journal entitled "Compatibility*
1638 *between Livestock Databases Used for Quantitative Biosecurity Response in New Zealand"*

1639 **Introduction**

1640 Datasets for estimating animal counts which are available at the national level in New Zealand
1641 include AgriBase, FarmsOnLine (FOL), the data generated by the National Animal Identification and
1642 Tracing Scheme (NAIT), the Statistics New Zealand Animal Production Survey (APS) and various
1643 sources of industry data which are fragmented and not available to MPI. FOL, NAIT and AgriBase are
1644 discussed below and the APS data are discussed in Chapter 8 where these data are used to validate
1645 results of the demographic models presented in Chapters 6 and 7. NAIT data are more extensively
1646 analysed in the appendix 14.6. A representation of the relationships which currently exist between
1647 the different data sources is shown in Figure 5.1.

Sources of animal counts



1648

1649 **Figure 5.1:** Schematic diagram of the relationships between existing and possible national level
 1650 sources of animal counts, how data sets can currently be joined and challenges and strengths
 1651 associated with the data sources.

1652

1653 In response to the threat of epidemics, national-level livestock databases provide a vital resource for
 1654 understanding the spread of infectious disease between farms and for supporting policy decisions
 1655 (Stevenson et al., 2007). The use of these data in quantitative analysis has become an integral part of
 1656 disease outbreak investigation and planning (Sanson, 1993). Such analyses range from statistical
 1657 approaches such as disease mapping (e.g. Métras et al., 2015) and case-control comparison (e.g.
 1658 Ellis-Iversen et al., 2011), to social network analysis (e.g. Firestone et al., 2012), and complex
 1659 epidemic and ecological models comprising both mathematical and statistical elements (e.g. Jewell
 1660 et al., 2009b). Mathematical epidemic models are particularly important as they are able to
 1661 assimilate many sources of epidemiological and demographic data, for example the number of

1662 animals on farms, the spatial locations of farms, and the presence of animal movement networks
1663 between farms (Jewell et al., 2009a; Stevenson et al., 2013; Brooks-Pollock et al., 2014). As such,
1664 models provide a principled framework in which to explore the complex interplay between farm
1665 characteristics, proximity, and human behaviour in influencing disease spread. Model-based data
1666 assimilation may be used before an outbreak occurs to identify optimal interventions for reducing
1667 disease transmission and mitigating an outbreak (Tildesley et al., 2006), and for providing the basis
1668 of simulated outbreak exercises such as Exercise Taurus 2012 (Ministry for Primary Industries, 2012).
1669 Disease models may also be used during an outbreak to evaluate disease control strategies, design
1670 local management tactics, and calculate the likely impact of an outbreak in terms of welfare and
1671 socioeconomic pressures (Volkova et al., 2010; Flood et al., 2013). Moreover, recent developments
1672 in statistical methodology have provided the opportunity to fit epidemic models to data as an
1673 outbreak emerges, providing a best practice approach for quantitative investigation of
1674 disease spread determinants (Ster et al., 2012; Brooks-Pollock et al., 2014; Jewell and Brown, 2015).
1675 The trade-off for such analytic power is the requirement for highly accurate data, though
1676 determining the sensitivity of models to data inaccuracy is a complex, ill-posed problem which may
1677 only be avoided by effective data collection (Eames and Keeling, 2003; Kitching et al., 2006). It must
1678 be kept in mind that extreme caution should be used with extreme caution when employing models
1679 in the face of an epidemic as model utility will depend on many things including construction,
1680 assumptions and, as this project discusses, data.

1681

1682 Animal demographic information may be obtained from multiple discrete sources in New Zealand,
1683 maintained by both government and private industry bodies. Two primary national level databases
1684 exist: AgriBase is managed byASURE Quality Limited, a New Zealand state-owned enterprise which
1685 provides service delivery for MPI, whereas FarmsOnLine (FOL) is maintained by MPI. Separate
1686 demographic databases are maintained by Statistics New Zealand's 5-yearly census, and a number of
1687 private industry groups collect data for control programmes and research purposes. In addition to
1688 these demographic data sources, the National Animal Information and Tracing System (NAIT)
1689 records movements of cattle and deer. NAIT is a subsidiary company of OSPRI³ New Zealand
1690 (www.ospri.co.nz), which is a non-profit making company owned by cattle and deer industry bodies
1691 in New Zealand.

³ OSPRI New Zealand is the company which was formed in 2013 when the ANimal Health Board (the former names of TBFree New Zealand) and NAIT merged together. The acronym OSPRI attempts to reference the main functions of the holding company which is set up to deliver Operational Solutions to the PRImary industries.

1692 In New Zealand, the recent epidemic of bovine anaemia associated with *Theileria orientalis* in cattle
1693 (Lawrence et al., 2016) provided the motivation for investigating the extent to which the national
1694 livestock databases link together to provide a central integrated data resource. During this outbreak
1695 the real-time Bayesian forecasting model of Jewell and Brown (2015) highlighted the importance of
1696 being able to link complementary sources of data held by different organisations. However, those
1697 authors reported difficulties matching case reports to farm identities in their demographic data, and
1698 an inability to accurately link cattle movements to herd location data.

1699 In the light of this work, the aim of the present study was to characterise the three most important
1700 livestock databases in New Zealand, FOL, AgriBase and NAIT, investigate their degree of
1701 compatibility, and highlight key aspects preventing them linking to form a comprehensive view of
1702 the national livestock industry.

1703 **Materials and Methods**

1704 **FarmsOnLine (FOL)**

1705 This is a non-mandated system constructed by MPI in 2010 to provide farm-level information for
1706 biosecurity purposes (<https://farmsonline.mpi.govt.nz/>). It was initiated to fulfil a recognised need
1707 to have a single database of all rural properties with their land boundaries and contact details. The
1708 objective of FOL is to include information on all farms, milk and meat processing facilities, and sale
1709 yards in the dataset. This information is not publicly available, and is protected by the Biosecurity
1710 Act 1993. Each record includes a geospatial polygon representing the farm boundary, as well as
1711 occupier information, farm manager contact details, the farm type and includes count data on
1712 different livestock species. Each farm is given a unique FOL identifier based on its locality and some
1713 have been historically matched to AgriBase identifiers using geographic proximity. Updating
1714 information on the livestock species on the farms and information on the farm occupiers relies on
1715 public updating via web interface. Additionally, the FOL data management centre reconciles changes
1716 in boundaries and ownership of approximately 20% of landowner details per year, resulting in a five-
1717 year average period to review the whole FOL dataset. Land in New Zealand is estimated to change
1718 hands at a rate of about 15,000 properties per year (~6% per annum), and changes in ownership
1719 identified by New Zealand Post and OSPRI updates are notified to the data management centre and
1720 updated. FOL retains no historical records when change of ownership information occurs.

1721 **AgriBase**

1722 AgriBase was developed in 1993, and was initially focussed on commercial properties susceptible to
1723 foot-and-mouth disease (Sanson, 1993). Each farm is given a unique AgriBase identifier with records
1724 for production data, geospatial polygons representing land parcels, occupier information, property
1725 information based on farm type, and farm manager contact details. It includes the presence of a
1726 wide range of animal species, including horses and poultry, as well as plant crop types. Data are
1727 updated via a website accessed by landowners and surveys undertaken by a commercial company,
1728 AsureQuality. Data extracts from AgriBase are available under license from AsureQuality at cost, and
1729 are supplied with a data quality statement.

1730 **National Animal Information and Tracing System**

1731 This system records the movements of individual cattle and deer around New Zealand and
1732 commenced operation in July 2011. The requirement for farmers to register cattle and deer with
1733 NAIT and report movements is compulsory, and legislated under the National Animal Identification
1734 and Tracing Act, 2012. The major unit of registration is a NAIT number, which is applied for by a
1735 person in charge of animals (PICA). The person in charge of animals may register their NAIT number
1736 as being associated with zero or more FOL records, resulting in a many-to-many relationship
1737 between the FOL and NAIT databases. Individual animals are tagged with radio frequency
1738 identification ear tags, and their movements must be recorded via a web interface by the source and
1739 destination person in charge of animals. This double-entry system thus provides a data validation
1740 mechanism. Each NAIT number is associated with a production type consisting of any combination of
1741 dairy, beef, venison, velvet and trophy. A person in charge of animals who manages animals on
1742 farms within 20 km of each other for the purposes of the same business may register all these FOL
1743 properties to a single NAIT number, and does not have to register movements between these farms.
1744 Conversely, if the farms are >20 km apart, then a person in charge of animals must register multiple
1745 NAIT numbers.

1746 **Data analyses**

1747 Snapshots of the dataset in FOL, AgriBase, and NAIT were obtained on 16 September 2014, 1
1748 September 2014 and 30 April 2014, respectively and loaded into a geographic information system
1749 (GIS)-enabled relational database (PostgreSQL 8.4 with PostGIS 2.2.0 extensions). This provided a
1750 convenient platform for interfacing with statistical software, R version 3.0.1 (R Core Team, 2014)
1751 enabling us to process large amounts of geospatial data.

1752 The overall spatial distribution of farms reported in both AgriBase and FOL was first calculated using
1753 two dimensional kernel density estimation of centroids of farms reporting presence of animals

1754 across New Zealand (Diggle, 1985). To detect differences in spatial coverage of AgriBase and FOL, we
1755 split each database into subsets of farms reporting the presence of cattle, pigs, deer, and small
1756 ruminants (sheep and goats). For each subset in each database, a two dimensional kernel density
1757 estimate was calculated. The resulting surfaces for AgriBase were then divided by those for FOL, and
1758 plotted to give maps of the ratio of spatial farm densities for AgriBase versus FOL. For both parts of
1759 this analysis, a fixed kernel bandwidth of 37 km was used, chosen for presentation purposes.

1760 To examine the species composition of each dataset, a frequency table of animal numbers per farm
1761 was calculated. For the subset of farms in FOL that had been matched to AgriBase identifiers, the
1762 agreement of presence or absence of animals by species between the two databases was measured
1763 by percentage agreement as well as Cohen's kappa statistic (Cohen, 1960). Cohen's kappa (K)
1764 measures agreement between two categorical outcomes relative to an expectation under complete
1765 randomness: if both datasets are in complete agreement then $K = 1$ and if there is no agreement
1766 between the two categories other than what would be expected by chance, $K = 0$.

1767 The NAIT data may be characterised as a dynamic network, in which nodes represent NAIT numbers,
1768 and edges represent animal-batch movements or the movements of single animals. These data were
1769 represented by two separate tables. Firstly, an edge list of movements was constructed where each
1770 row represented a single movement with source and destination nodes, the date of movement, and
1771 the number of animals moved in the batch. Secondly, a node list of NAIT numbers was created
1772 where rows represented each NAIT number with the associated production type; dairy, beef, velvet,
1773 venison or trophy (deer used for trophy hunting), or any combination of these, and classification;
1774 farm, lifestyle block, sale yard, show and rodeo grounds, meat processing plants, import or export
1775 transitional facilities, quarantine areas, game estates, safari parks, zoos and other. A further table
1776 was constructed to represent the NAIT-FOL relationship, allowing the edge list source and
1777 destination nodes to be linked to centroids contained in the FOL database and hence movements to
1778 be georeferenced.

1779 Descriptive analysis of NAIT data was undertaken by examining the distributions of the number of
1780 FOL identifiers per NAIT number, the number of NAIT numbers per FOL identifier, and the number of
1781 NAIT numbers associated with each of the classifications and production types. The network density
1782 was calculated as

$$1783 \quad \rho = \frac{100 \times n_c}{N(N-1)}$$

1784 where N is the total population size (the number of unique NAIT numbers), and n_c is the number of
1785 directed pairs of NAIT numbers having ≥ 1 animal movement between them. The apparent level of

1786 compliance with the 20 km limit on the distance between FOL properties registered against a NAIT
 1787 number was assessed by calculating the maximum distance between pairs of FOL identifiers
 1788 associated with each NAIT number, and the proportion of NAIT numbers where this distance
 1789 exceeded 20 km.

1790 To examine the agreement between NAIT production type and presence or absence of deer, and
 1791 dairy and beef cattle in both FOL and AgriBase, the NAIT database was flattened so that each record
 1792 represented a unique NAIT/FOL linked pair. The NAIT-listed production type was used to allocate the
 1793 presence or absence of beef, dairy and deer to the FOL blocks associated with a NAIT number. These
 1794 were then compared with the FOL recorded presence and absence of these species, again using
 1795 percentage agreement and Cohen’s kappa statistic. Similarly, those properties that had both an
 1796 AgriBase key generated by MPI and were associated with a NAIT number were examined to
 1797 determine the level of agreement amongst the datasets on the presence of beef cattle, dairy cattle
 1798 and deer.

1799 Results

1800 The FOL data extracted contained records on 220,485 registered premises. Of these, 100,892 (46%)
 1801 were recorded as having ≥ 1 animal of any of the species of interest present. The majority of
 1802 premises were recorded as having only one animal present, as shown in Table 5.1. The extract from
 1803 AgriBase contained records for 134,377 farms, of which 84,423 (63%) were recorded as having ≥ 1
 1804 animal of any species on the property. This included 578 properties that had only avian species
 1805 present, which were removed to leave 83,845 properties to compare with FOL. Numbers of
 1806 properties recorded with animals of the different species in each database are shown in Table 5.1.

1807 **Table 5.1:** Number of premises recorded on the Farms on Line (FOL) and AgriBase databases with the
 1808 reported number of animals per property for different animal species, on 16 September 2014 and 1
 1809 September 2014, respectively.

Database	Animal type	Number of animals					None ^a	Max ^b
		1	2–10	11–100	101–1,000	>1,000		
FOL								
	Beef	70,007	253	398	271	23	149,533	32,000
	Dairy	23,581	6	43	248	13	196,594	2,900
	Deer	5,359	3	8	43	9	215,063	3,000
	Sheep	44,815	43	56	39	58	175,474	48,000
	Goats	4,584	1	7	0	1	215,892	1,000
	Pigs	7,663	11	3	0	2	212,806	6,000
AgriBase								
	Beef	1,583	21,266	20,766	9,611	415	80,736	14,500

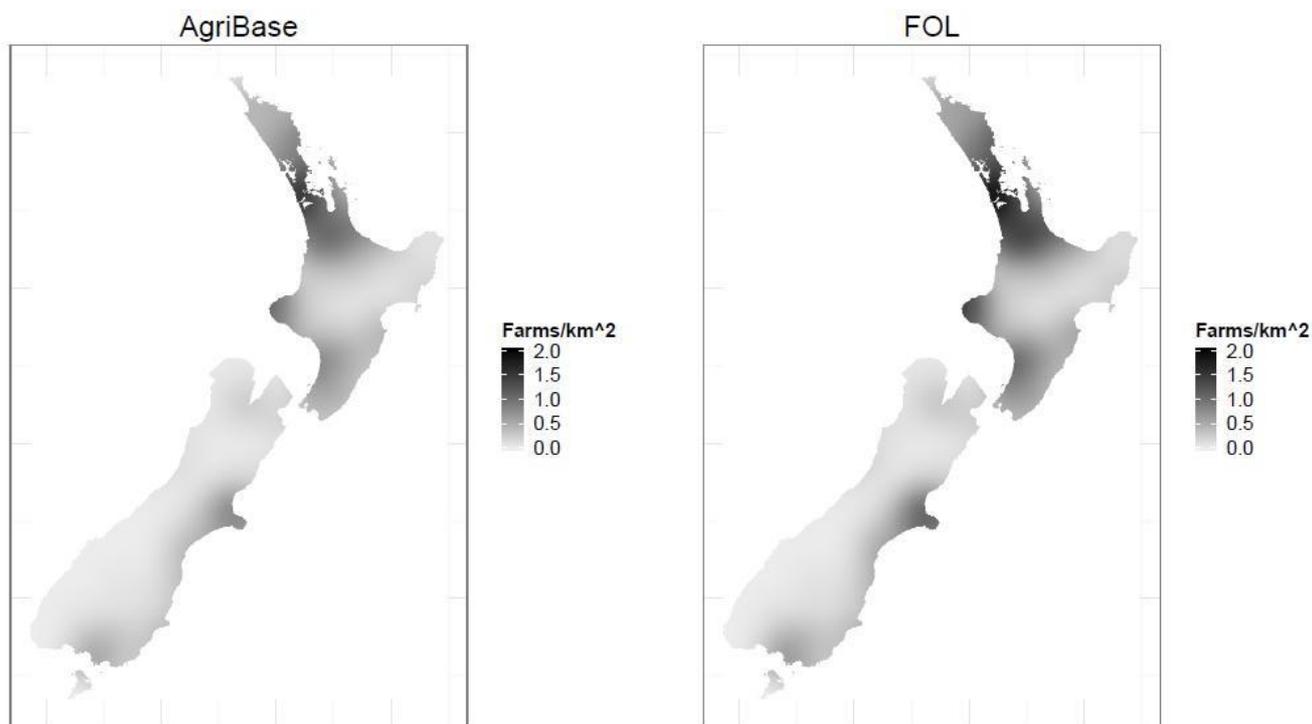
Dairy	262	646	1,564	11,676	709	119,388	10,220
Deer	95	338	1,265	1,920	299	130,460	19,249
Sheep	775	8,637	13,044	6,637	9,298	95,986	115,620
Goats	2,270	2,019	984	346	15	128,743	4,000
Pigs	1,165	2,904	465	94	83	129,666	44,000

1810 ^a Number of premises with no records of the animal species

1811 ^b Maximum number of animals

1812

1813 The overall distribution of the density of farm centroids for FOL and AgriBase records are shown in
1814 Figure 5.2.



1815
1816 **Figure 5.2:** The spatial distribution of properties in New Zealand recorded in the (a) AgriBase and (b)
1817 Farms on Line databases reporting the presence of cattle, sheep, goats, deer and pigs, on 1
1818 September 2014 and 16 September 2014, respectively, calculated using spatial kernel density
1819 estimation with a 37 km bandwidth. The scale bars represent the density of farm centroids per km²
1820 for the different regions.

1821 For both datasets, there were greater farm densities in the Waikato and Auckland regions, and also
1822 within the Christchurch area of the Canterbury region when compared with the rest of New Zealand.
1823 The general distribution of farms appeared the same for both databases, but AgriBase showed a
1824 systematically lower density, consistent with there being fewer farm records compared to FOL. In
1825 FOL, there were 147 centroids that had multiple associated FOL identifiers, and 14 FOL identifiers
1826 had no recorded centroid. In comparison, AgriBase had 285 centroids associated with multiple
1827 AgriBase identifiers, with no records having unrecorded centroids.

1828 The ratio of the spatial densities in AgriBase relative to FOL, for cattle, pigs, deer, and small
1829 ruminants is shown in Figure 5.3. For pigs and deer, the density ratio was reasonably homogeneous
1830 across New Zealand, with AgriBase having a farm density approximately 80% of FOL. However, for

1831 cattle and small ruminants, there was considerable heterogeneity, with AgriBase showing a density
 1832 of cattle farms in Central Otago region that was 20% of FOL, and a density of small ruminant farms in
 1833 the central West Coast area that was twice that of FOL. These observations are in agreement with
 1834 the results presented in Figure 5.3. With the exception of farms reporting goats, the number of
 1835 properties that were recorded with ≥ 1 animal was less in the AgriBase compared to FOL. However,
 1836 AgriBase contained more detail on numbers of animals, whereas FOL generally described only
 1837 presence or absence of species.

1838 Of the 220,819 records present in FOL, 81,626 had associated AgriBase identifiers. These farms
 1839 represented 37% of the FOL records, and 87% of the AgriBase records. For these records, the
 1840 calculated agreement for the presence of different species, expressed as percentage agreement and
 1841 Cohen's kappa are shown in Table 5.2. For both measures, there was a substantial level of
 1842 agreement across all species.

1843 **Table 5.2.** Agreement between the AgriBase, Farms on line (FOL) and NAIT databases in terms of
 1844 recorded presence of animals of different species on properties in the databases, calculated as
 1845 percentage agreement and Cohen's kappa statistic (K)^b.

Species	AgriBase and FOL		FOL and NAIT ^a		AgriBase and NAIT ^a	
	Agreement (%)	K	Agreement (%)	K	Agreement (%)	K
Beef	89.1	0.8	71.2	0.3	70.1	0.3
Dairy	96.0	0.8	87.7	0.7	86.3	0.6
Deer	99.0	0.8	96.4	0.6	96.4	0.6
Sheep	93.0	0.8	NA	NA	NA	NA
Goats	98.0	0.7	NA	NA	NA	NA
Pigs	97.0	0.6	NA	NA	NA	NA

1846 ^a Note that NAIT only contains records for beef and dairy cattle, and deer

1847 ^b Where $K=0$ is no agreement, 0–0.2 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is
 1848 moderate agreement, 0.61–0.80 is substantial agreement and 0.81–1 is almost perfect agreement
 1849 (Landis and Koch, 1977)

1850 NA=not applicable

1851

1852 The NAIT data set contained records for 982,552 individual movement records, from which 61,747
 1853 unique NAIT numbers were identified, along with their premises and production types. Of these
 1854 NAIT numbers, 61,049 (99%) were associated with land parcels classified as farm or lifestyle block;
 1855 3,456 (5.5%) NAIT numbers were assigned no production type, and 5,963 (9.5%) had >1 production
 1856 type to describe their operations. Bearing in mind that each NAIT number may have multiple
 1857 production types, 15,561 (25.0%) NAIT numbers were assigned to dairy; 46,334 (75.0%) to beef; 220

1858 (0.4%) to trophy; 972 (1.5%) to velvet and 2,143 (3%) to venison. Out of a total of 3.8 billion possible
1859 connections between directed pairs of NAIT numbers, only 304,020 had >1 animal movement during
1860 the study interval, giving a network density of 0.008%.

1861 In total, the NAIT extract referenced 67,189 unique FOL identifiers. The number of FOL identifiers
1862 per NAIT number ranged between 0–34, and was highly right skewed with 1,162 (2%) of NAIT
1863 numbers being associated with no FOL identifier, and 51,851 (84%) with only one FOL identifier. The
1864 distribution of the number of NAIT numbers per FOL identifier was similarly right skewed, ranging
1865 from 0–14, with 60,777 (90%) FOL identifiers being associated with a single NAIT number.

1866 Georeferencing showed that 66,982/67,189 (99.7%) FOL identifiers referenced in the NAIT data
1867 were present in the FOL extract. Of the 8,734 NAIT numbers associated with >1 FOL identifier 769
1868 (9%) had FOL centroids >20 km apart, with 183 (2%) >50 km apart. In the course of the analysis,
1869 there were 802 movements that were registered as having the same source and destination NAIT
1870 number.

1871 There were 75,699 unique NAIT/FOL combinations with associated NAIT production type. All of
1872 these records were matched to records in FOL, and 49,025 were matched to records in AgriBase. The
1873 percentage agreement and unweighted Cohen’s kappa statistic presented in Table 5.2 indicate fair
1874 to moderate agreement for these comparisons.

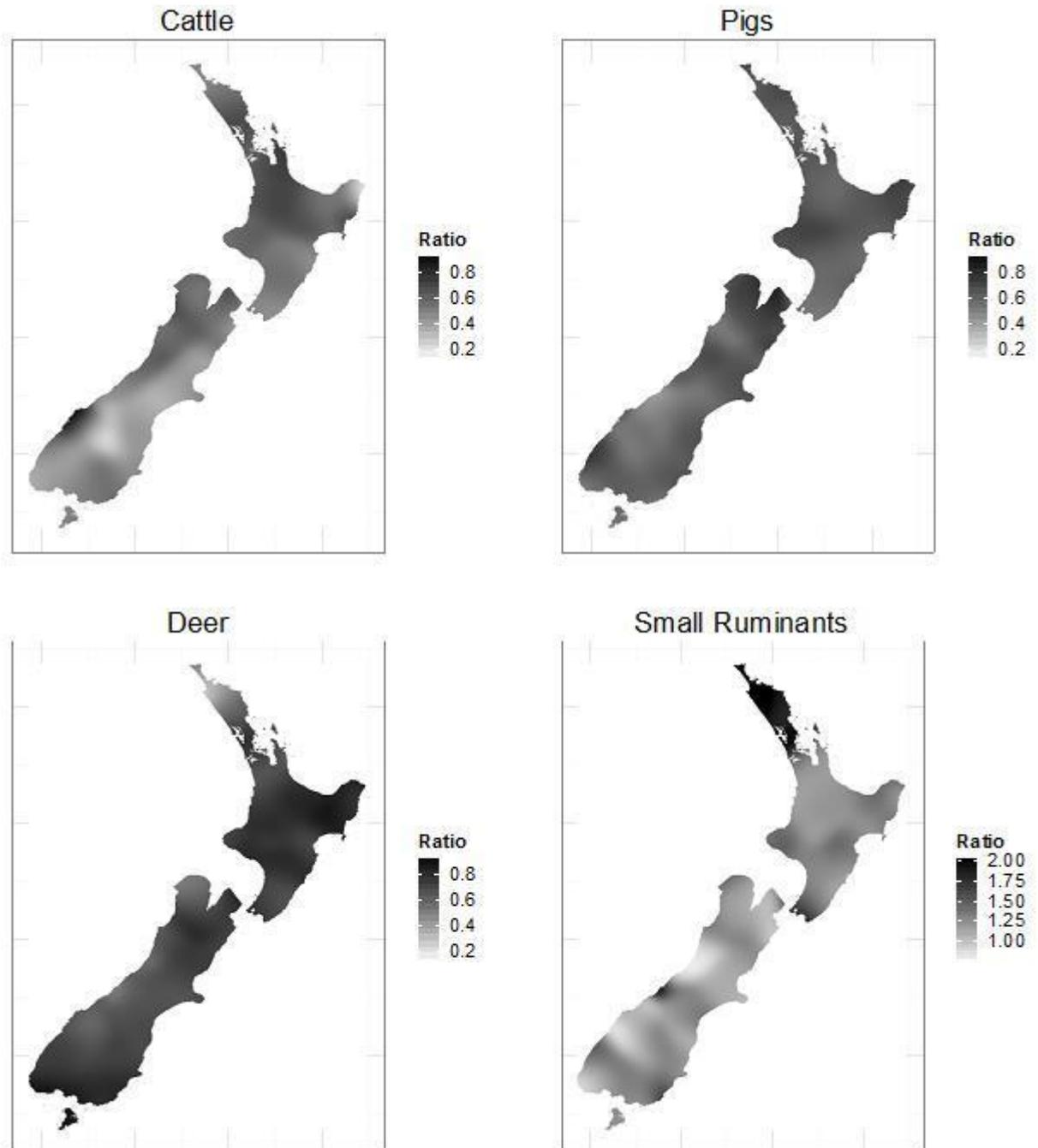
1875 **Discussion**

1876 The presence of multiple datasets of livestock demographic data in New Zealand presents the
1877 opportunity for a rich corpus of information on which to base models of disease spread. In this
1878 chapter, we have taken the three most prominent of these databases, have compared their content,
1879 and investigated their propensity to link together to form a single syndicated data resource.

1880 AgriBase, FOL, and NAIT have each been developed to address a specific biosecurity aim, and joining
1881 and assimilating their data raises the possibility of exploiting synergy in their complementary
1882 information. However, our results show that not only do these databases give different pictures of
1883 the livestock geographic landscape, they also lack a common identifier making linking between
1884 records difficult and fraught with assumptions and uncertainty.

1885 AgriBase and FOL are both aimed primarily at recording geospatial information on individual farms.
1886 The major difference is in the number of records contained in each database, which is attributed to
1887 their initial intended coverage. Whereas FOL is intended to capture all livestock-owning premises in
1888 New Zealand, AgriBase initially recorded premises with foot-and-mouth disease susceptible animals
1889 only, albeit with much greater detail. Development later in the history of AgriBase focussed more on

1890 non-susceptible animals and crops. The difference in national-level livestock spatial densities,
1891 revealed in Figure 5.3, may be explained by the different ways that these two databases are curated.
1892 FOL and AgriBase originate from a common historical data source, but inconsistencies in the way in
1893 which identifiers (owners) are assigned to land parcels, and the differences in management practices
1894 relating to updating of information, has over time led to the evolution of two distinct data lineages.



1895

1896 **Figure 5.3:** *The density of farm centroids per km² recorded in the AgriBase database as a proportion*
1897 *of that in the Farms on Line databases for (a) cattle, (b) pigs, (c) deer, and (d) small ruminants (sheep*
1898 *and goats). Note that the scale for small ruminants is different to the other three species.*

1899 In order to reconcile the data contained in FOL and AgriBase, augmenting the former with detailed
1900 information from the latter, it is necessary to accurately link records between the two databases.
1901 However, the current lack of coherent farm identifiers makes unification of the databases impossible
1902 without introducing untestable assumptions with associated uncertainty. In principle, spatial joins
1903 may be used to link AgriBase and FOL records (Shekhar and Chawla, 2003). For example, using the
1904 geospatial polygon field it may be assumed that an AgriBase record could be assigned to a FOL
1905 record with which it has maximum overlap. However, pilot exploratory analysis showed a number of
1906 AgriBase records having no overlap with FOL records, thus requiring further criteria such as linking
1907 by proximity. Whatever linking strategy is used, the validity of the spatial assumption is difficult to
1908 assess without extensive surveys on the ground. The uncertainty that this engenders, and its impact
1909 on subsequent analysis, is therefore difficult to quantify. Given that FOL originally took data from
1910 AgriBase, an alternative strategy would be to construct a genealogy for each database, tracing
1911 modifications and updates back in time to a common ancestor. Unfortunately, as far as we are
1912 aware, the databases do not contain any timestamp fields dating modifications to individual records.

1913 Though NAIT and FOL share FOL identifiers in a well-defined way, the greatest weakness of NAIT is
1914 that animal movements are only recorded between groups of FOL identifiers. The many-to-many
1915 relationship between NAIT numbers and FOL identifiers renders it impossible to ascertain exactly
1916 between which of the associated FOL entities a movement occurred. We suggest that a pragmatic
1917 approach to this problem is to assume that a movement could have occurred with equal probability
1918 between all combinations of source and destination centroids. Nevertheless, we highly recommend
1919 adjusting NAIT such that the FOL identifiers for the movement are requested at the point of data
1920 entry. Of course, as NAIT does not possess AgriBase identifiers, linking animal movement to the
1921 detailed information contained in AgriBase is not directly possible except through an uncertain FOL-
1922 AgriBase join as previously discussed. We remark that this deficiency is particularly serious in the
1923 case that AgriBase identifiers are used to archive disease case reporting data, undermining the utility
1924 of NAIT for disease tracing purposes (Jewell and Brown, 2015).

1925 The proportion of connected directed pairs of farms in NAIT is low in comparison to a similar study
1926 of Scottish data (Volkova et al., 2010). This may well reflect a difference in the behaviour of farming
1927 industries between New Zealand and Scotland, but it may also point to compliance issues in relation
1928 to NAIT being a recent requirement. As discussed in Jewell and Brown (2015), the low density of this

1929 network exacerbates mis-assignment of animal movements to geographic locations, giving rise to
1930 serious bias in the results of quantitative disease outbreak analyses. The degree to which this bias
1931 improves in subsequent years will in time answer this question. We also note that 9% of NAIT
1932 numbers were found to violate the requirement of having FOL centroids <20 km apart making it
1933 possible for animal movements of >20 km to go unrecorded. There is some scope for error here, as
1934 the farm is a polygon, but a centroid was used in this analysis. The authors recognise that there
1935 could be a small amount of error associated with this method compared with the use of farm
1936 polygons. As animal movements are being tracked, a more in-depth analysis of NAIT data may, in
1937 time, enable an accurate estimate of the number of animals per NAIT number, providing a high-
1938 accuracy source of data to update FOL. Further restrictions on tagging exemptions from July 2015 for
1939 cattle and February 2016 for deer will increase coverage of the NAIT database and therefore
1940 improve its ability to achieve this. We note that further research on NAIT data will undoubtedly yield
1941 relevant findings on movement patterns both spatially and temporally, informing measures of
1942 disease spread risk, as demonstrated in the United Kingdom (Robinson et al., 2007a). We conclude
1943 that New Zealand has the necessary databases for a comprehensive quantitative decision support
1944 system that monitors and informs strategic interventions to support production animal health.
1945 However, despite significant progress in developing these systems over the last 20 years, they do not
1946 yet have the ability to link seamlessly together to take advantage of modern analytical techniques in
1947 outbreak decision support. A critical step towards resolving this compatibility problem will be the
1948 establishment of a national standard for identifying farms, with uptake by all relevant data archiving
1949 facilities. From a quantitative perspective, this would allow an immediate single view of the data,
1950 whilst maintaining current efficiency in the distribution of data collection and curation. We note that
1951 a central archive of all livestock-related data has been attempted in the United Kingdom through the
1952 Rapid Analysis and Detection of Animal-related Risks (RADAR) system. This system assimilates key
1953 livestock information, and has the advantage that it is kept current via updates from separate data
1954 collection and storage systems within government divisions charged with specific surveillance tasks
1955 (Lysons et al., 2007).

1956 We also stress the importance of high quality metadata accompanying database extracts, giving
1957 information on data accuracy (Strong et al., 1997). This is necessary to support scientific validity
1958 where data uncertainty exists, and hence justification for policy decisions made in response to
1959 scientific outputs. Furthermore, we postulate that regular and transparent data review processes
1960 will increase the quality of New Zealand's data archive, and enhance its utility for effective decision
1961 making. With New Zealand's economic focus on primary industry, fast access to accurate
1962 demographic data are a necessary component of outbreak preparedness. Therefore, through our

1963 analysis we hope to raise a second generation discussion on how New Zealand's livestock data can
1964 be better syndicated for rapid disease response and epidemic modelling purposes.

1965

1966

1967

1968 **Deliverable 4: Machine Learning Models**

1969 **Tracey Hollings¹, Andrew Robinson¹, Mary van Andel², Chris Jewell³, Mark Burgman¹**

1970 ¹ *Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Melbourne Australia*
1971 *3010*

1972 ² *Ministry for Primary Industries, Wellington, New Zealand 6140*

1973 ³ *Lancaster University, Lancaster, UK, LA1 4YG*

1974

1975 **This chapter has been submitted for review entitled "Comparison of species distribution models for*
1976 *predicting national-level livestock demographics"*

1977 **Introduction**

1978 Rising demand for animal products is leading to the global expansion and intensification of livestock
1979 industries (Prosser et al., 2011). A concomitant increase in global movement of animals and animal
1980 products via international trade is providing greater chances for the transmission of invasive pests
1981 and diseases into naïve countries (Levine and D'Antonio, 2003), which can have catastrophic
1982 economic, social and animal welfare impacts. For example, the eight-month foot-and-mouth disease
1983 (FMD) outbreak in the United Kingdom in 2001 led to the culling of more than six million animals for
1984 disease control and 'welfare' (Anderson, 2002), and cost the UK more than £7 billion (Thompson et
1985 al., 2002), including substantial restrictions on trade.

1986 Reliable up-to-date spatial distribution and abundance records for livestock are critical for planning
1987 and policy applications in livestock industries, especially for disease risk modelling and emergency
1988 response (Robinson et al., 2007b; Robinson et al., 2014). Availability of a complete census of
1989 vulnerable animals would facilitate a rapid response for the containment and eradication of an
1990 incursion such as FMD, where rapidity is a key factor in substantially reducing the economic,
1991 biological and social impacts of epidemics (Belton, 2004). However, few countries maintain accurate
1992 up-to-date livestock census data, particularly at the resolution and spatial scale required (Robinson
1993 et al., 2007b; Prosser et al., 2011).

1994 Statistical modelling can be used to fill gaps, where data are too expensive and time consuming to
1995 collect. Species distribution models (SDM) predict species distributions by relating species
1996 occurrence or abundance records to environmental, demographic, climatic and satellite derived
1997 predictor variables (Elith et al., 2006). A diverse range of applications and greater availability of
1998 candidate predictor data have resulted in a rapid rise in ecological and conservation applications

1999 (e.g. Elith et al., 2006), and more recently for agricultural (e.g. Van Boeckel et al., 2012; Robinson et
2000 al., 2014; Nicolas et al., 2016) and disease risk modelling (Bogh et al., 2007; Stevens et al., 2013).

2001 To date, models that predict livestock and commercial poultry populations have focused almost
2002 exclusively on stratified regression using such predictor variables as land quality and human
2003 population density (White et al., 2001; Neumann et al., 2009; Prosser et al., 2011; Van Boeckel et al.,
2004 2011; Van Boeckel et al., 2012; Robinson et al., 2014). Reasonable predictive accuracy (correlations
2005 of up to 80% between predicted numbers and direct counts) has been demonstrated at local,
2006 regional (Prosser et al., 2011; Van Boeckel et al., 2011) and global scales (Robinson et al., 2007b) In
2007 most of these studies training data has comprised the disaggregation of coarse spatial resolution
2008 census data into animal densities. The Gridded Livestock of the World project (GLW) estimated
2009 global distributions at 1km² resolution for major livestock species of cattle, goats, buffalos, sheep,
2010 pigs and poultry (Robinson et al., 2014). The GLW data has been applied in animal health research
2011 (Franceschini et al., 2009) including for predicting the risk of avian influenza infection (Gilbert et al.,
2012 2014) and the incidence and distribution of FMD (Sumption et al., 2008). Asian poultry stock has also
2013 been estimated for HPAI epidemiology, which can have serious human health implications (e.g.
2014 Prosser et al., 2011; Van Boeckel et al., 2011; Van Boeckel et al., 2012).

2015 The choice between SDM approaches is important because there is considerable variation in
2016 predictive performance between different approaches (Elith et al., 2006). Studies evaluating SDM
2017 techniques have arisen in response to the large variety of models available, types of data and
2018 research questions, but most comparisons use presence-only or presence/absence data (as opposed
2019 to counts), are based on a small number of species, and/or applied over a limited geographic extent
2020 (Segurado and Araujo, 2004; Elith et al., 2006; Tsoar et al., 2007; Elith and Graham, 2008). None
2021 have been reported for livestock data.

2022 The ecological and environmental requirements of most livestock species are well understood, and
2023 relevant training and predictor data are often readily available, both of which make widespread
2024 predictions feasible. However, models for common, widespread and generalist species generally are
2025 less accurate than environmentally or geographically restricted species (Segurado and Araujo, 2004;
2026 Elith et al., 2006; Tsoar et al., 2007). Therefore the statistical standards used to assess the quality of
2027 ecological models for narrow range species may not be suitable for widespread species.

2028 Machine learning methods have been proposed for livestock estimation due to their potentially
2029 higher predictive performance, and their capacity to incorporate complex interaction effects and
2030 noisy data (Elith et al., 2006; Robinson et al., 2014). They have been implemented recently for this
2031 purpose (Nicolas et al., 2016). Advances in computing capacity, software and statistical innovation

2032 (Guisan and Thuiller, 2005) have made machine learning techniques such as Random Forests (RF),
2033 boosted regression trees (BRT) and k-nearest neighbour (K-NN) models, practical options for large
2034 datasets (Robinson et al., 2014). Methods such as RF and BRT are relatively novel for ecological
2035 applications but they consistently outperform more established methods (Elith et al., 2006; Cutler,
2036 2007). Examining and comparing machine learning techniques and SDM approaches (including
2037 Bayesian techniques in the following chapter) using livestock data is critical to ensure the application
2038 of appropriate models, given their growing popularity for estimating livestock and poultry
2039 populations and the potential value and application of the predictions.

2040 New Zealand is considered to have one of the strongest biosecurity frameworks in the world
2041 (Meyerson and Reaser, 2002). Its geographic isolation and strict biosecurity regulations have
2042 prevented the incursion of many significant agricultural pests and diseases, including FMD.
2043 Preventing the arrival of agricultural diseases is crucial to maintaining access to international
2044 markets, reputation, and protecting the economy and industry. Biosecurity investment by the NZ
2045 government is substantial but the recent and economically damaging incursions of varroa mite, fire
2046 blight and bacterial kiwifruit vines disease (PSA) show that significant agricultural pests still pass the
2047 border undetected. New Zealand has an accurate, up-to-date national-level database of the
2048 geographic locations of farms called Agribase (See previous chapter for an overview of this
2049 database); however counts of livestock on these holdings are not reliably available. Yet, this
2050 information is important for biosecurity and disease preparedness (Emelyanova et al., 2009; Ministry
2051 for Primary Industries, 2012).

2052 In this chapter we develop and test three of the most promising machine learning species
2053 distribution models to estimate national-level farm animal demographic data: boosted regression
2054 trees (BRT), random forests (RF) and K-nearest neighbour (K-NN). The models are built from a
2055 commercial livestock database and environmental and socio-economic predictor data for New
2056 Zealand. The context of the model development is to provide support for decision-making in an
2057 emergency response situation and as initial values for stochastic models that are used for planning.
2058 We assessed their performance using a regional cross-validation with model predictions evaluated
2059 for local government areas to assess how well models can predict to new regions. The models were
2060 assessed at two spatial scales: at the farm level and using quarantine zones whereby models were
2061 tested in emergency disease response situations by evaluating model predictions in a 3km radius
2062 around randomly selected 'outbreak' farms. This 3 km radius reflects the emergency quarantine
2063 zones that may be established immediately following notification of a serious disease such as FMD.
2064 In this chapter the model predictions are evaluated and validated using the same database. The

2065 following chapter uses Bayesian techniques with the same data and spatial stratifications to allow a
2066 comparison of model predictions from the different modelling techniques in Chapter 6.

2067 **Methods**

2068 **Livestock data**

2069 AgriBase™ (AsureQuality) is a national-level commercial livestock database that holds contact data
2070 and geospatial information about agricultural and horticultural properties across NZ. The database
2071 was initially developed for biosecurity and emergency management applications and currently lists
2072 134,377 individual properties in vector polygon data format. Entry in the database is voluntary and
2073 more than 60% of the properties have records for livestock. ‘Lifestyle farms’ on the urban fringe are
2074 the most numerous farming type by a factor of almost four (>60,000), followed by commercial beef
2075 (15,433) and dairy farms (11,840). No data on enterprise activities of any kind are held for 36,815
2076 properties, of which 87% are categorised as ‘lifestyle’ farms; the majority of these are considered to
2077 be missing data rather than true zero’s. This is due to the voluntary nature of the database with no
2078 incentive for property owners to declare their livestock holdings. However, because it was not
2079 possible to distinguish between missing values and values that are really zero, all were removed for
2080 model development and assessment. There were also approximately 6,000 farms with identical
2081 spatial information reflecting cross-leased land where information from multiple polygons has been
2082 merged into a single data point. This prevented the correct extraction of predictor information for
2083 the individual farm polygons and therefore precluded their use in the models. In total 91,662
2084 properties were used for model training and validation.

2085 Many livestock demographic applications require knowledge of the numbers of livestock contained
2086 on a property. We built models with a consistent response variable across properties by calculating
2087 the livestock units (LSU) as grazing equivalents for each property. This value was calculated by
2088 multiplying the total number of animals of a given species by a value that represents their grazing
2089 equivalence to an adult dairy cow (Table 6.1; Eurostat, 2013; The Scottish Government, 2014). In
2090 certain contexts, knowledge of specific livestock species is required, so single species models were
2091 also assessed for cattle (derived from beef and dairy records). It was not possible to differentiate
2092 between free range and intensively farmed animals.

2093

2094

2095 **Table 6.1:** Species equivalence values to an adult dairy cow that were used for analysis (Eurostat,
 2096 2013; The Scottish Government, 2014)

Livestock unit (LSU) coefficients	LSU
Dairy cows	1.00
Beef cows	0.80
Deer	0.80
Horses	0.80
Pigs	0.40
Goats	0.10
Sheep	0.10
Poultry (not used in this study)	0.01

2097

2098 **Predictor variables**

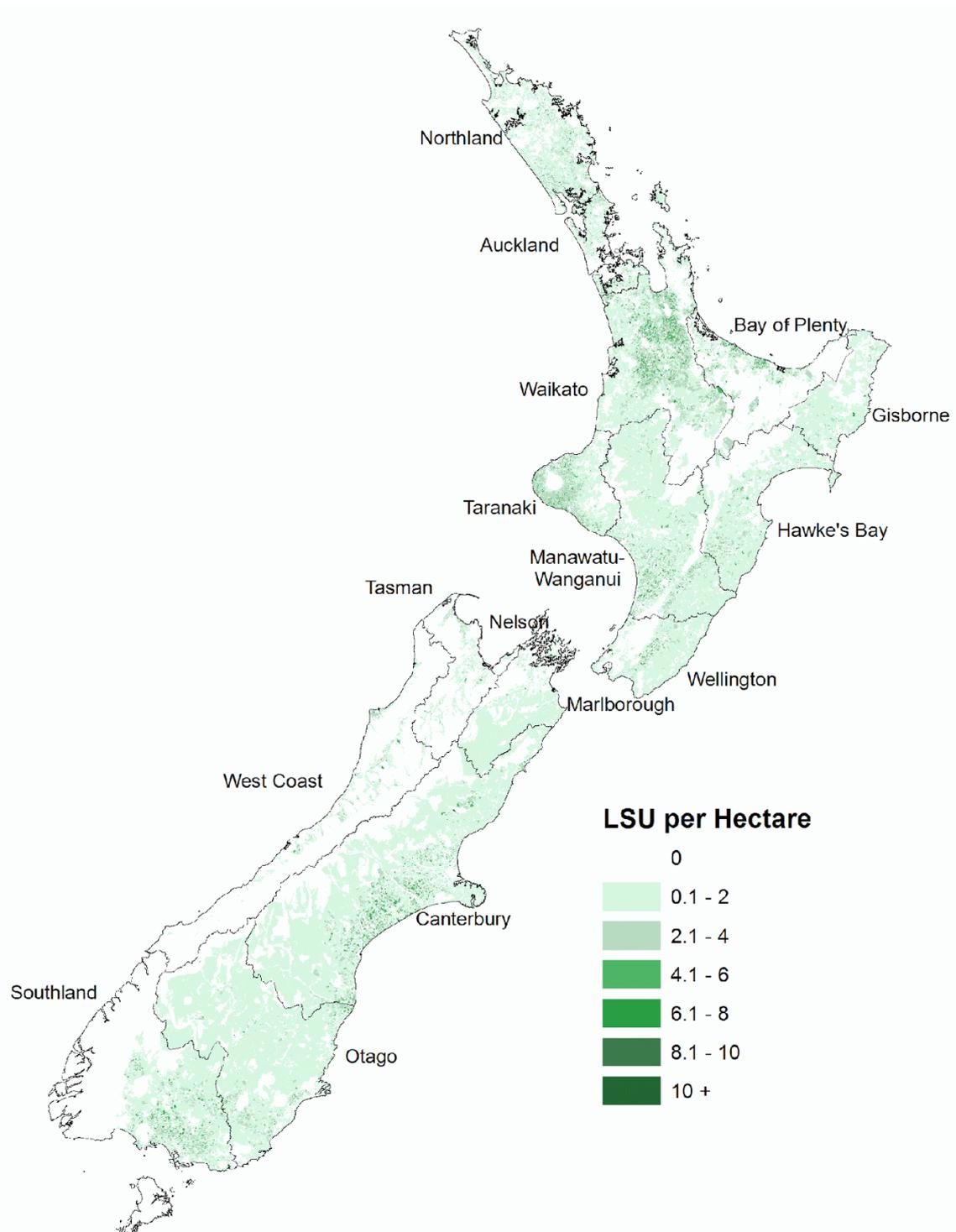
2099 Predictor variables were calculated and/or extracted for individual property polygons from spatial
 2100 data layers using the open-source statistical environment, R (R Core Team, 2014, version 3.1.1;
 2101 libraries: 'rgeos', 'rgdal', 'maptools', 'raster', 'geosphere', and 'sp') (Pebesma and Bivand, 2005;
 2102 Bivand et al., 2014; Bivand and Lewin-Koh, 2014; Bivand and Rundel, 2014; Hijmans, 2014b, a). A
 2103 short description of the data source and calculation for each variable is given in Table 6.2. The 22
 2104 candidate predictor variables that could potentially influence livestock distributions and farming
 2105 types were a combination of environmental (slope, aspect, elevation, pasture quality, distance to
 2106 river), climatic (mean annual rainfall, mean annual temperature) and demographic (social
 2107 deprivation index, distance to major road, distance to urban centre) variables. Environmental and
 2108 climatic factors account for natural limitations and species tolerance thresholds, for example, food
 2109 and water availability, affects of aspect on vegetation and solar radiation etc, and temperature.
 2110 Demographic variables account for human factors associated with livestock production, including
 2111 market access and human population density (Van Boeckel et al., 2011; Robinson et al., 2014).
 2112 Variability in continuous predictor variables is expected across properties, which cannot be
 2113 represented by simply using the variable mean, for example 'hilliness' is not accounted for by the
 2114 mean of the slope. To account for this variability, the within-polygon standard deviations of several
 2115 variables were used as candidate predictors in addition to the within-polygon means. The total
 2116 absolute area of each property was divided into four separate predictor variables: high quality
 2117 pasture, low quality pasture, forest, and 'other' land cover calculated as the remaining area of the
 2118 property polygon, all measured in hectares.

2119 There are regional differences in livestock holdings across NZ, related to environmental and
 2120 demographic conditions. Therefore, where possible the models included a categorical spatial
 2121 predictor variable representing regional councils (Figure 6.1) to improve statistical relationships
 2122 between training level data and predictor variables (Robinson et al., 2014). KNN techniques are
 2123 unable to adequately accommodate categorical predictors, so the KNN models did not include this
 2124 predictor variable.

2125 **Table 6.2:** Summary of the source and calculation of predictor variables used for modelling.

Variable (Abbreviation)	Description	Source of data
Distance to main roads (ROAD_DIST)	Calculated from the centroid of the farm property to the nearest major sealed road from vector line layer	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to urban centres (URBAN_DIST)	Calculated from the centroid of the farm property to the nearest urban centre	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to rivers (RIVER_DIST)	Calculated from the centroid of the farm to the nearest mapped river from vector line layer	Land and Information NZ (LINZ). https://data.linz.govt.nz/
Social deprivation index (DEPRIV)	Extracted from raster layer by the intercept with the farm centroid.	Statistics New Zealand census data. Decile measure of socio-economic status; 1 being lowest and 10 highest.
Area of high quality pasture (HQ_HA)	Overlay of the vector polygon LCDB layer with farm property polygon layer to calculate number of hectares classified as 'high producing exotic grassland'	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 (Hectares) (LQ_HA)	Overlay of the vector polygon LCDB layer with farm property polygon layer to calculate number of hectares classified as 'low producing grassland', 'depleted grassland', 'tall tussock grassland'.	Land Cover Database (LCDB) Version 4. A satellite derived GIS vector polygon layer of land use. https://iris.scinfo.org.nz
Forest cover (HA_FOREST)	The union of endemic and exotic forest cover of New Zealand as ha of each farm	https://koordinates.com/layer/168-nz-native-polygons-topo-150k/ https://koordinates.com/layer/131-nz-exotic-polygons-topo-150k/
Area of 'other' land cover (OTHER_AREA)	The size of the farm was calculated in post-GIS with the st_area command for the property polygon. HQP, LQP and FOR were subtracted from the total area to give this	Agribase™

	variable.	
Mean and standard deviation of annual temperature (TEMP)	The mean and sd of the annual temperature across the farm calculated from raster cells within a farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al., 2002). www.koordinates.com
Mean and standard deviation of mean minimum temperature of the coldest month (TEMP_MIN)	The mean and sd of mean minimum temperature of coldest month across the farm calculated from raster cells within a farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al., 2002). www.koordinates.com
Mean and standard deviation of annual solar radiation (SOLAR)	The mean and sd of the annual solar radiation across the farm calculated from raster cells within a farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al., 2002). www.koordinates.com
Mean and standard deviation of winter solar radiation (SOLAR_WINTER)	The mean and sd of the winter solar radiation across the farm calculated from raster cells within a farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al., 2002). www.koordinates.com
Annual water deficit (WATERDEF)	The mean and sd of the annual water deficit across the farm calculated from raster cells within a farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al., 2002). www.koordinates.com
Slope mean and standard deviation, aspect (SLOPE)	Slope x sin(aspect) and Slope x cosin(aspect) according to the methods in Stage (1976)	Digital elevation model from Land and Information NZ (LINZ). https://data.linz.govt.nz/
Elevation mean and standard deviation (DEM)	Mean and sd of farm elevation calculated from the grid falling within farm polygon	Digital elevation model from Land and Information NZ (LINZ). https://data.linz.govt.nz/
Regional Council (COUNCIL NAME)	The Agribase farm polygon classified into one of the 17 regional council regions of NZ	NZ Regional Councils https://koordinates.com/layer/4240-nz-regional-councils-2012-yearly-pattern/



2127

2128 **Figure 6.1:** Map of Livestock Units (LSU) per hectare across New Zealand from Agribase™. White
 2129 areas represent areas removed such as cross leased land or areas of missing data. Areas of more
 2130 than 10 LSU/cattle per hectare are represented in dark green. LSU represented per hectare is used for
 2131 display purposes only.

2132 **Spatial stratifications**

2133 Models were built and evaluated with training and validation data based on regional stratification,
2134 using randomly selected local government authority areas, to assess the ability of the models to
2135 predict to new regions. NZ mainland has 16 different regional councils, hereafter called regions
2136 (Figure 6.1). Training data comprised a stratified sample of 11 randomly selected regions, six from
2137 the north island (out of nine) and five from the south island (out of seven) and the validation data
2138 comprised the remaining 5 regions. Two hundred independent training and validation datasets were
2139 generated in this way. This analysis also provides information on whether prediction accuracy of the
2140 models differed across regions of NZ.

2141 Model performance was assessed at two spatial scales. The first was a comparison of the
2142 predictions and the actual values for farms in the withheld regions. The second level assessed the
2143 adequacy of the models in emergency disease response situations by validating model predictions in
2144 nominal “quarantine zones” around randomly selected farms. We chose a radius of 3 km to reflect
2145 the emergency response quarantine zones that typically would be established immediately following
2146 notification of a serious disease such as FMD. A grid of 9000m x 9000m cells was overlaid across NZ,
2147 providing approximately 3,000 grid cells. The property intersecting with the centroid of each grid cell
2148 was extracted using ArcGIS (Version 10.2.2). Two hundred of these intersecting properties were
2149 selected at random and then each was used as the hypothetical ‘outbreak farm’, the focal point for a
2150 3km quarantine zone. All farms within a 3km radius of the focal property boundary were extracted
2151 and used as validation data using the spatial libraries ‘rgeos’, ‘rgdal’, and ‘sp’ in R (Pebesma and
2152 Bivand, 2005; Bivand et al., 2014; Bivand and Rundel, 2014, version 3.1.1; R Core Team, 2014). The
2153 200 ‘outbreak farms’ provided a balance between providing enough validation data for modelling
2154 without overlapping zones becoming an issue. The remaining farms outside the 200 hypothetical
2155 quarantine zones were used as training data. Ten iterations of these methods provided ten
2156 independent sets of training and validation data extracted from the database, to give a total of 2000
2157 outbreak farms (200 outbreak farms x 10 independent datasets). Three locations were omitted as
2158 no valid farms were found in their corresponding zones, leaving 1997 for comparison. For each new
2159 iteration the bottom left corner of the grid was moved at random to produce ten distinct grids, and
2160 thereby different sets of grid centroids to intersect with farms. Since sampling of the outbreak farms
2161 was performed spatially by proximity to the farm centroid, this sample is with probability
2162 proportional to size.

2163

2164 **Models Tested**

2165 Three SDM modelling techniques were considered. Various models have parameters which can be
2166 changed by the user, for example the number of trees to be grown. The optimal parameters were
2167 assessed for each model type prior to fitting the final models and are outlined for each model type
2168 below. Each model strategy assumes that the sample is representative of the population, which is a
2169 very common statistical assumption. A formal interpretation of our assessment of the models
2170 requires the assumption that the residuals are uncorrelated; this is unlikely owing to the spatial
2171 nature of the data; however, such assumptions are again common and we do not think that it would
2172 substantially undermine our conclusions.

2173 *Random Forests (RF)*: RF is a bagging algorithm that creates decision trees by repeatedly selecting
2174 bootstrap samples from a training set, fitting trees to each replicate. Bagging randomly generates a
2175 set of data from the original with replacement. A random sub-sample of the predictors is used to
2176 split each node in a tree (Breiman, 2001). The given number of trees is averaged to obtain the
2177 estimates. Parameters can be changed in the model, including the number of trees to be grown and
2178 the number of randomly selected variables to be used as candidates at each split. RFs were built
2179 using the 'randomForest' package in R (Liaw and Wiener, 2002; R Core Team, 2014, version 3.1.1)
2180 with the number of trees set to 2000. For the RF models, each response variable was transformed
2181 using a square root, based on statistical practice (the metrics are counts, for which the square root is
2182 a common variance-stabilizing transformation) and empirical inspection of residual plots.

2183 *Boosted Regression Trees (BRT)*: BRT's combine two algorithms, namely regression trees and
2184 boosting, to build and fit many models and improve predictions. The algorithm builds a large
2185 number of simple decision trees adaptively, and the final BRT model can be described as an additive
2186 regression model (Elith et al., 2008). Several parameters can be varied in the model: a user-defined
2187 bagging fraction introduces stochasticity into the model and defines the proportion of data drawn at
2188 random from the original data at each step; the learning rate varies the contribution of each tree
2189 added to the model and the tree complexity defines the number of nodes for each tree (Elith et al.,
2190 2008). Predictive performance was assessed under different bagging conditions, learning rate and
2191 tree complexity and the parameters were chosen based on those which achieved the minimum
2192 predictive error. Parameters in the final BRT models had a bagging factor of 0.5, learning rate of
2193 0.005 and tree complexity of 10. Boosted regression trees (BRT) were implemented using the
2194 'dismo' and 'gbm' packages in R (Ridgeway, 2013; Hijmans et al., 2014). For the BRT, the response
2195 variable was assumed to have a conditional Poisson response and log link function.

2196 *K-Nearest Neighbour (K-NN)*: K-NN is a distance-weighted nearest neighbour prediction algorithm in
2197 which training sites closest in predictor distance to the new point of interest are used to predict its

2198 value. Closer training samples can be weighted higher than those further away. The output of the k-
2199 NN for regression is the average value of the k nearest neighbours, where the k nearest properties
2200 were used to predict the LSU/cattle of the target properties. We defined the number of samples (k)
2201 as 5, and the distance metric used was Euclidean distance. K-NN models were fit using the ‘class’
2202 package in R (Venables and Ripley, 2002). No transformation was needed for the K-NN models
2203 because they make no assumption about the conditional distribution of the response variable.

2204 **Model development and evaluation**

2205 The predictive accuracies of Boosted regression trees (BRT), Random forests (RF) and K-nearest
2206 neighbour (K-NN) were evaluated and compared for national-level livestock estimation in NZ. Two
2207 response variables were separately modelled to test multi-species aggregates and single species
2208 models: LSU rounded up to the nearest whole number to give counts of LSU, and cattle as a single
2209 species model case study, derived from beef and dairy records. For each response variable, 200
2210 replicate BRT, RF and K-NN models were run, corresponding to 200 independent training and
2211 validation datasets. The models were set up with the same 22 predictor variables (Table 6.2). The
2212 models were assessed for how well they predicted livestock on individual farms and the aggregated
2213 farms within quarantine zones.

2214 The model fits were determined primarily by how well they predicted LSU in the withheld samples. A
2215 goodness of fit measure between the withheld validation data and predicted values was used to
2216 compare the BRT, RF and K-NN models: the discrepancy quantified using the Root Mean Square
2217 Prediction Error (RMSPE).

$$RMPSE_i = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

2218 where the quantity is only computed for those farms in the regions that were excluded from the
2219 replicate sample, and \hat{y}_i is computed by predicting the farm (or summing the predicted farms to the
2220 quarantine zone) lsu or cattle.

2221 **Results**

2222 **Methods comparison**

2223 Boosted regression trees were the best performing models overall when assessed on RMSPE.
2224 However the prediction accuracy of random forests was higher for many regions (Table 6.3 and
2225 Supplementary Tables). This was the case for farm-level and quarantine zone predictions for both

2226 LSU and cattle (Table 6.3). However when assessing the RMSPE by farm size, overall RF predictions
 2227 were substantially better than BRTs for all but very large farms (> 250 LSU) (Table 6.4). Random
 2228 forests predicted better for farms with zero values and farms generally which generally fall into the
 2229 lifestyle farm category with less than 15 LSU. These differences in goodness of fit measures between
 2230 RF and BRTs mean the choice of model should be based on context of model development and
 2231 requirements of the model predictions. Both BRT and RF performed significantly better than K-NN
 2232 across all spatial stratifications and response variables (Table 6.3, Supplementary Tables). No single
 2233 K-NN model performed better than BRT and RF with any goodness of fit measure, with the same
 2234 predictor and response data.

2235 There was contrasting model performance with respect to the predicted values. K-NN predictions
 2236 included zero values, but overall predictions were poor. Neither RF nor BRTs predicted zero values at
 2237 all, and BRT had poor performance relative to RF when predicting LSU values less than 15, which is
 2238 common for over 90% of lifestyle farms (Table 6.4). Random forests performed the best for all but
 2239 very large farms. In contrast, for large LSU values BRT provided much better predictions than RF.

2240 Total number of farms was 91662. Number of farms per quarantine zone was 64 on average, and the
 2241 average by region ranged from 25.5 to 101. The mean number of zones per region was 125, and the
 2242 range was 2 to 445.

2243 **Table 6.3:** *The prediction root mean squared error (RMSPE) for LSU and cattle using the withheld*
 2244 *results for the regional spatial stratification for both individual farms within regions and to 3km*
 2245 *radius quarantine zones. Table also shows the mean count of LSU and cattle per farm and the*
 2246 *standard deviation for all farms in brackets. Results for Quarantine zone predictions are in 1,000s of*
 2247 *LSU and cattle. Full results for all models are shown in the Supplementary Materials.*

Response variable	Spatial stratification	Mean count per farm (SD)	RMSPE		
			RF	BRT	KNN
LSU	Farm	131.5 (342.1)	215.6 (45.4)	216.4 (48.2)	380.2 (57.3)
	Zone ('000)	12.7 (19.6)	5.4 (1.0)	4.8 (0.8)	9.0 (1.2)
Cattle	Farm	89.9 (237.2)	172.9 (31.8)	171.4 (31.6)	275.3 (32.4)
	Zone ('000)	7.8 (14.7)	6.4 (1.3)	3.5 (0.6)	5.2 (1.4)

2248

2249 **Table 6.4:** Comparison of BRT, RF, and KNN by the RMSPE for LSU across different size farms.

Farm Size (number of LSU)	Number of farms	RF	BRT	KNN
No animals	16650	61.5	74.4	292.0
Small (1- 15)	34582	31.0	37.3	111.6
Medium (16-150)	18623	82.3	99.8	252.3
Large (151-250)	5618	134.9	152.7	312.8
Very large (>250)	15919	505.4	498.5	780.5

2250

2251 **Region cross-validation with Random Forests**

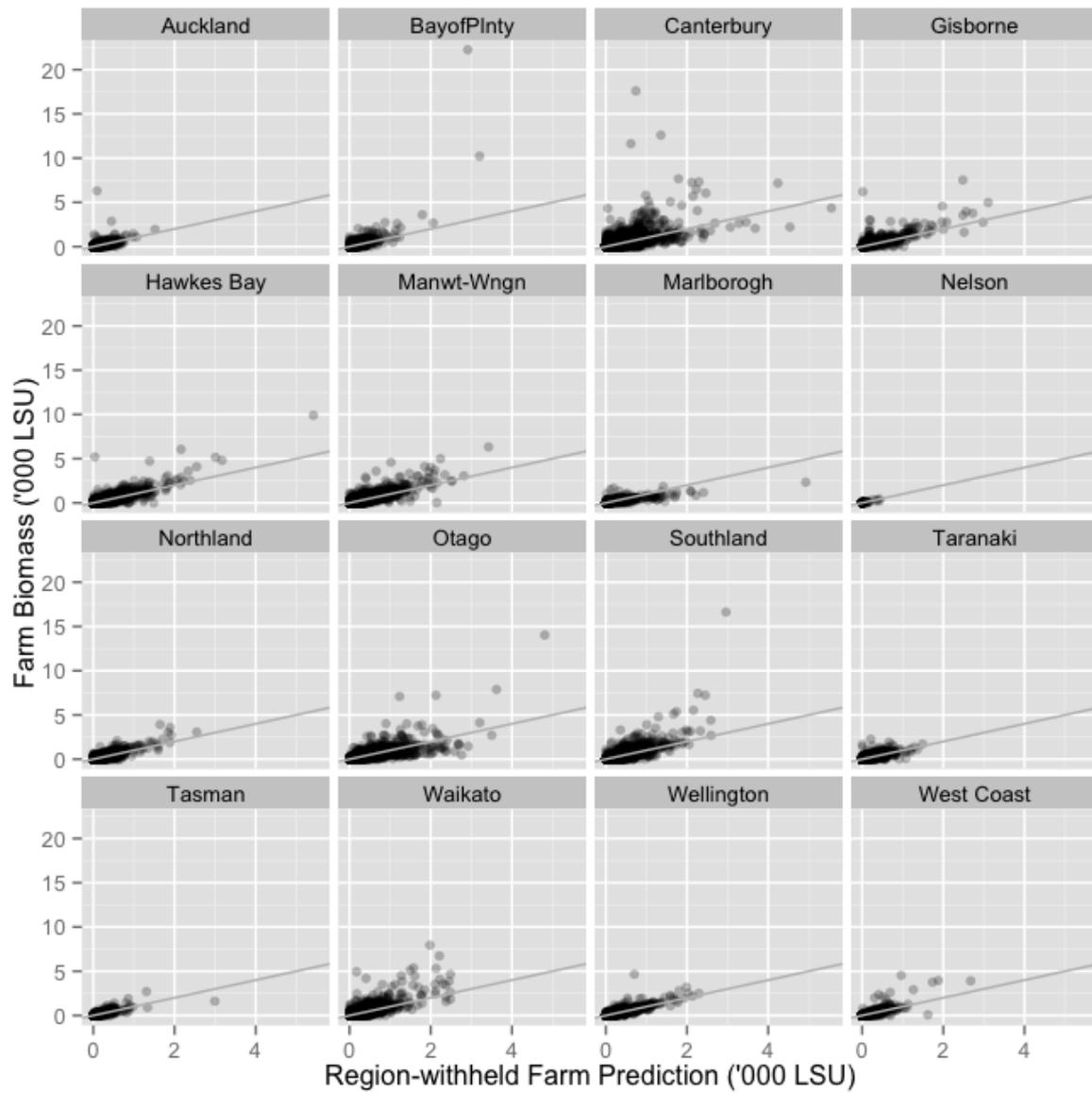
2252 ***Farm-level predictions***

2253 Figure 6.2a provides a scatterplot of the observed against the fitted farm-level LSU values for the
 2254 region-withheld RF model, by region. The results for all regions tend to cluster closely around the 1:1
 2255 line, which confirms the model quality. The predictions also show evidence of fanning, meaning that
 2256 the natural variation increases with the size of the response variable. The fanning was
 2257 accommodated in the model by means of a square root transformation of the response variable, as
 2258 noted in the Methods. The scatterplots of cattle results are not shown, as they are a subset of the
 2259 LSU variable and relationships are similar to LSU across all regions simulations.

2260 ***Quarantine zone predictions***

2261 Table 6.3, Supplementary Table 6.5 and Figure 6.2b summarise the results for the quarantine level
 2262 predictions for the regional stratification model. We summed the predictions for every farm that had
 2263 any area within a 3 km radius of the centroid of the subject farm, and applied the same statistics to
 2264 these summed quarantine-level predictions and observations. Again, the results cluster closely
 2265 around the 1:1 line, and as expected there is much less variation around the line than in farm-level
 2266 predictions.

2267 ***a) Farm-level LSU***

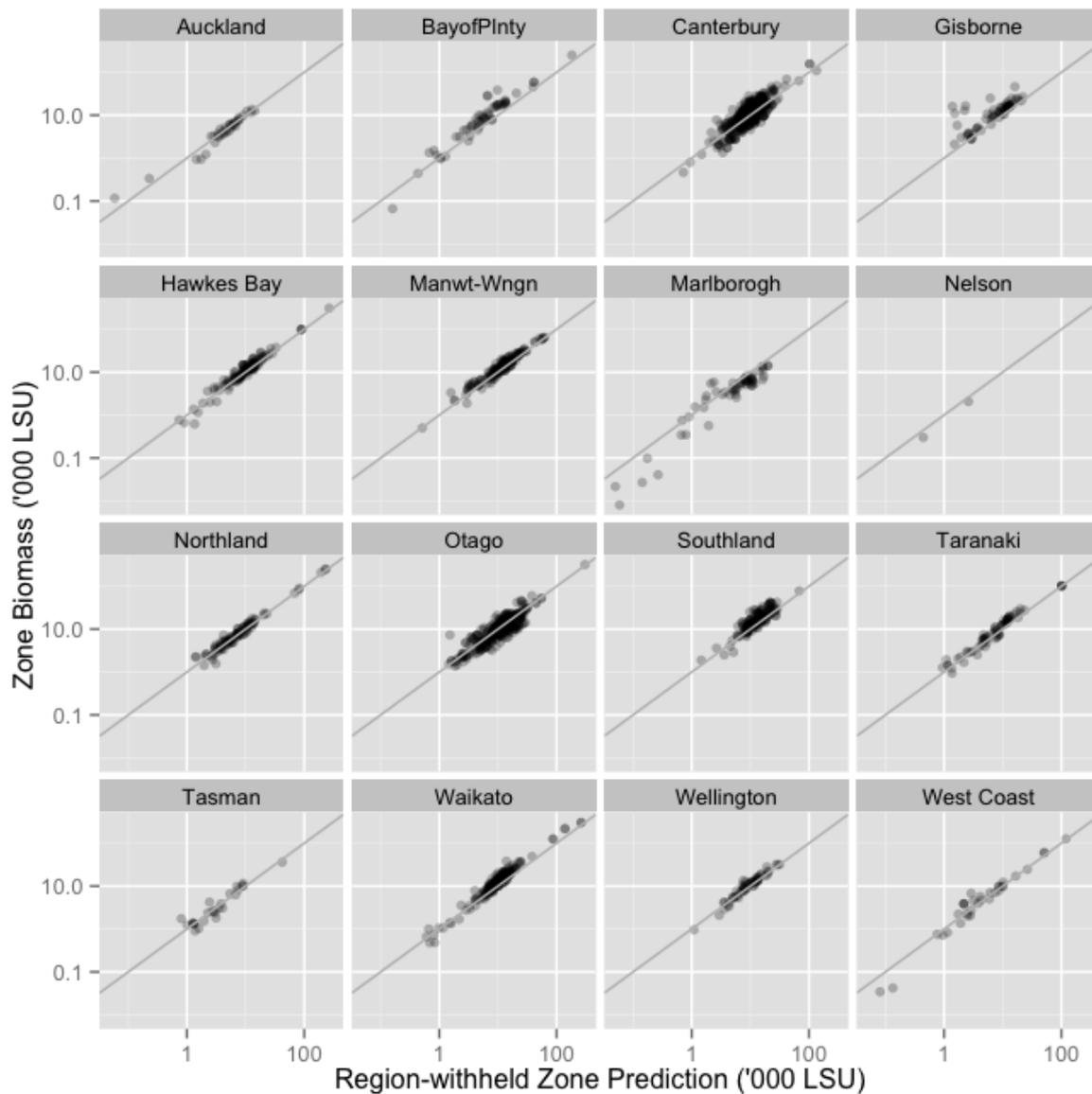


2268

2269

2270

b) Total quarantine zone LSU



2271

2272 **Figure 6.2:** Scatterplot of the observed against the predicted a) farm-level LSU and b) total
2273 quarantine zone LSU for the regional spatial stratification, plotted by region. Total quarantine zone
2274 LSU was calculated by summing the observed and predicted values for all farms within the 3km
2275 quarantine zone. Cattle results are not shown as plots they are a subset of LSU.

2276

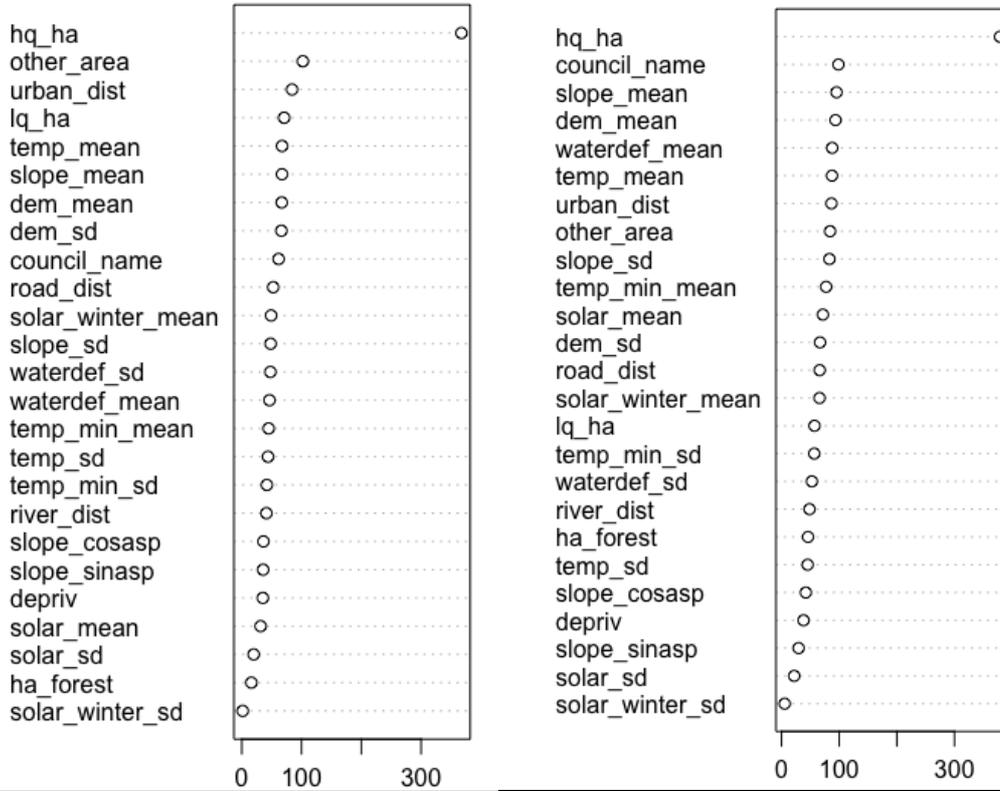
2277 **Variable importance**

2278 We determined the most important variables for predicting both LSU and cattle by running the
2279 random forest model on the full dataset with all covariates. Comparing separate models for each of
2280 the two response variables can identify differences in variable importance for multi species and
2281 single species models.

2282 Figure 6.4 shows the relative influence of predictor variables for the RF models for both LSU and
2283 cattle response variables. The x axis represents the mean decrease in accuracy, as measured by an
2284 increase in mean square error (MSE), if the variable were randomly permuted. High quality pasture
2285 was overwhelmingly the most important predictor for both LSU (367%) and cattle (378%) and had
2286 substantially more influence than the second highest ranked predictor variable. For LSU the size of
2287 the farm, accounted for by the combination of area, high quality pasture, low quality pasture and
2288 the variable 'other area' was relatively important, as was the distance to an urban centre. For cattle,
2289 the region, as well as variables accounting for the slope and altitude of the farm were the most
2290 highly ranked variables after high quality pasture. Variables accounting for aspect (namely, slope
2291 cosine and slope sine), deprivation index and the two solar standard deviation predictor variables
2292 were not important for either response but still had some impact on model performance, given by
2293 positive relative importance measures.

2294

LSU



Increase in MSE (%) if variable randomly permuted

Figure 6.4: Relative influence of predictor variables from Random Forest models for LSU and predictions if the variable of interest were randomly permuted.

2296 Discussion

2297 Comparative species modelling exercises illuminate the strengths and weaknesses of different
2298 techniques (Guisan and Zimmermann, 2000) but such studies have not previously been undertaken
2299 for livestock data, and machine learning techniques have been applied to livestock data only very
2300 recently (e.g. Nicolas et al., 2016). The three machine learning techniques assessed in this study, RF,
2301 BRT and K-NN, produced vastly different results even with identical predictor and response
2302 variables, as seen in ecological studies (e.g. Segurado and Araujo, 2004; Elith et al., 2006).

2303 Boosted regression trees were the best performing technique overall for national-level livestock
2304 estimation, but RF's performed equally well or better in many regions and outbreak simulations.
2305 Random forests performed best for all but very large, typically commercial farms, which make them
2306 more valuable in disease response modelling and national livestock estimation, where data
2307 deficiencies tend to be clustered around smaller non-commercial and peri-urban farms. K-NN
2308 models performed poorly relative to the other models across all simulations. Other comparative
2309 studies have shown similar results to ours even when comparing up to 10 supervised learning
2310 techniques on non-ecological binary data (Caruana and Niculescu-Mizil, 2006).

2311 Discrepancies exist between the goodness of fit measure for RF and BRTs for some simulations and
2312 regions which make it difficult to determine a single best model. RMSPE is an indication of accuracy,
2313 i.e. how close the predicted values are to observed values (Robinson et al., 2014) and is heavily
2314 penalised for large departures from observed values. RF fit poorly for farms with very large counts in
2315 contrast to BRT and due to the nature of RMSPE, which heavily penalises large departures from
2316 observed values, the performance of RF based on RMSPE may have been disproportionately
2317 influenced relative to BRT.

2318 No universally best learning algorithm exists and the choice of method will depend on the data and
2319 the objectives of the study (Segurado and Araujo, 2004). All will perform poorly in some contexts,
2320 and significant variability exists across problems and assessment criteria (Caruana and Niculescu-
2321 Mizil, 2006). Model performance can also be affected by selection of a different spatial scale or
2322 resolution (Segurado and Araujo, 2004). Thus, both for this exercise and more generally, the choice
2323 of technique should depend on the application of the predictions. For NZ, the context of the spatial
2324 model development is to provide support for decision-making in an emergency response situation
2325 and in initialising planning models, and therefore RF's are recommended given their better
2326 performance for farm sizes of most interest. For areas consisting largely of commercial grazing (e.g.
2327 parts of the south island) BRT's could be considered. Stochasticity of populations at a local level
2328 means that model predictions should be used with caution for disease control decisions during an

2329 outbreak. Epidemic responses could be severely impacted by missing a piggery in the outbreak zone.
2330 The predictions however are valuable for regional and national responses to diseases, and planning
2331 and simulation modelling.

2332 Generally, SDM model comparison studies have operational shortcomings; often a single species has
2333 been used, they have been assessed over relatively small spatial scales, or their predictive
2334 performance has been assessed on the same data the models were built on, making the
2335 interpretation of results more difficult (Elith et al., 2006). Despite the model predictions not being
2336 assessed against a ground truthed validation dataset, this study does not suffer from many common
2337 SDM comparison studies issues and in contrast, uses abundance data successfully to predict
2338 national-level population numbers within property boundaries.

2339 Widespread species, such as livestock, generally have complex distributions and often do not
2340 respond predictably to environmental gradients. Such species are expected to be better modelled
2341 with non-parametric models such as RF, BRTs and K-NN (Segurado and Araujo, 2004). Random
2342 forests (Cutler et al 2007) and BRTs (Elith et al., 2006) are relatively novel in ecology though they
2343 have been widespread in other disciplines for some time. These machine learning techniques require
2344 greater computing capacity than classical tools such as generalized linear models, but are generally
2345 considered to have great predictive accuracy (Elith et al., 2006; Cutler, 2007; Robinson et al., 2014).
2346 Interaction terms are often omitted from traditional SDM techniques because they greatly increase
2347 the number of model parameters (Guisan and Thuiller, 2005), but RF and BRT can easily fit complex
2348 relationships and interactions without the need to explicitly specify them, they do not overfit and
2349 are stable with noisy data (Elith et al., 2006; Cutler, 2007; Elith et al., 2008). BRT's have successfully
2350 been used for disease and disease risk modelling (Bhatt et al., 2013; Gilbert et al., 2014) and RF's for
2351 the presence of rare and invasive plant species (Cutler, 2007). The capability of machine learning
2352 techniques, and RF in particular, for population mapping is being realised and many organisations
2353 are beginning to investigate these techniques including FAO and the WorldPop consortium
2354 (Robinson et al., 2014; Stevens et al., 2015; Nicolas et al., 2016).

2355 RFs have some advantages over other machine learning techniques. They can compute the
2356 contribution of each predictor through post-hoc analysis of variable importance measures, even
2357 where interactions exist (Cutler, 2007; Stevens et al., 2015), which has previously been cited as a
2358 challenge facing SDMs (Araujo and Guisan, 2006). RF's also have fewer user-defined parameters
2359 than BRT's, which can be of significant benefit when using automated fitting procedures (Stevens et
2360 al., 2015). The final advantage of RF over KNN and BRT is technological: at present, it is relatively
2361 straightforward although time consuming to produce distributions for the point predictions, using
2362 quantile RF (Meinshausen, 2006). These distributions can then be used to represent and propagate

2363 the statistical uncertainty of the prediction through instances of model simulation, although care
2364 must be taken to consider the potential correlation of prediction errors.

2365 For livestock predictions, high quality pasture was the most important predictor for all model
2366 simulations, and by an important margin. Farm slope, elevation and hilliness are also important and
2367 reflect what livestock species can be kept in certain areas, e.g. high country farms, renowned for
2368 steep terrain and winter snow, are often confined to sheep farming. Incorporating socio-economic
2369 and anthropogenic variables can improve model performance (Robinson et al., 2014) and this was
2370 demonstrated with the high relative importance of distance to urban centre and distance to road.
2371 This is likely to be the influence of lifestyle farms on the urban fringe. Similarly, human population
2372 density was a key predictor in determining livestock distributions in the GLW regression models
2373 (Robinson et al 2007), and large population centres and roads were both important in explaining
2374 cattle farm locations in Australia (Emelyanova et al., 2009). In our study, deprivation index, a
2375 measure of socio-economic status, was not statistically important for the LSU or cattle. Our results
2376 suggest that RF can be used successfully to estimate peri-urban livestock populations, despite other
2377 studies citing the potentially confounding issues of multifunctional land-use and human settlement
2378 (Neumann et al., 2009). In other studies, areas of lowest animal density on the periphery have been
2379 more difficult to accurately model with higher coefficients of variation and levels of uncertainty
2380 relative to higher density areas (Van Boeckel et al., 2011).

2381 Treating livestock species separately can improve epidemiological applications due to different
2382 disease susceptibility (Prosser et al., 2011). Infected FMD cattle become infected easily and produce
2383 large quantities of virus relative to sheep, which can have limited pathology and act as silent
2384 spreaders (Davies, 2002; Volkova et al., 2011). Pigs can also shed large volumes of virus and
2385 therefore can transmit disease easily. Therefore different farm types can play varying roles in FMD
2386 transmission depending on context (Ferguson et al., 2001; Keeling et al., 2001; Kao, 2002). Cattle
2387 were used in this study as a case study to compare single species vs multispecies model predictions,
2388 it may be worthwhile investigating the merits of further disaggregation into dairy and beef cattle.

2389 While it is possible to disaggregate LSU predictions to species level from observed data, running
2390 single species models can provide more information about the species' distribution and important
2391 variables. For example, for cattle, the regional identifier variable was very important, indicating that
2392 regional variations influence the distribution and abundance of livestock species. In NZ this is
2393 particularly the case for dairy, which is usually concentrated in specific areas. This outcome also
2394 indicates that spatial stratification is important. Different model performance based on spatial
2395 stratification has been observed in other studies but with much larger scale differentials. Similar
2396 work by Prosser et al (2011), estimating poultry populations, found regionally stratified models

2397 performed better than country wide models on the basis that the factors driving livestock densities
2398 will be similar in comparable ecological and administrative areas (Robinson et al., 2014).

2399 The spatial sampling approach used to identify 'outbreak' farms for establishing quarantine zones
2400 selected the farms with probability proportional to their area. This would not be an issue in disease
2401 response modelling as the outbreak farm would be known, rather than randomly selected. Region is
2402 a more unbiased measure for choosing predictor data but also has biases related to environmental,
2403 geographic and other factors. For example, certain regions in the south island may comprise mainly
2404 large-scale sheep farms due to the hilly terrain and low productivity land, and these types of farms
2405 may not be adequately represented if this region is removed for model validation.

2406 Unbiased data are rare for large numbers of species across extensive geographical areas (Araujo and
2407 Guisan, 2006) and low density areas can increase the uncertainty of predictions (Prosser et al., 2011;
2408 Van Boeckel et al., 2011). This was true for the AgriBase database used in this study. First, the
2409 database is voluntary which makes differentiating missing data from true zero's all but impossible.
2410 The majority of the 'missing' data was attributed to small peri-urban lifestyle farms. However, the
2411 removal of this data was unlikely to have had significant effects on model predictions because one
2412 third of the remaining data was classified as lifestyle farms, suggesting they were well represented in
2413 the data including with respect to location and variability of livestock holdings Second, discrepancies
2414 existed for some polygons, which were represented by merging of multiple farm ids and/or
2415 discontinuous polygons. This is done in the database as a way of representing cross leased land. All
2416 were removed due to the substantial investment required to investigate and separate each group of
2417 polygons relative to the likely cost to model accuracy. This data represented only a small fraction of
2418 the database (<5%) and was not expected to introduce any significant biases Additional errors were
2419 expected to arise from seasonal influences on animal movements and breeding, economic factors,
2420 droughts etc which can all affect the temporal and spatial representativeness of census data (Kao,
2421 2002). None-the-less the model predictions had high accuracy, but true validation would require on
2422 the ground farm surveys (Robinson et al., 2014).

2423 Livestock estimation is most commonly represented as numbers per square area (e.g. Neumann et
2424 al., 2009; Prosser et al., 2011; Van Boeckel et al., 2011; Robinson et al., 2014). However, this
2425 representation is problematic for estimating numbers in individual grid cells because livestock
2426 populations are spatially and temporally dynamic (Robinson et al., 2014). This study estimated
2427 livestock for each property, which is arguably a more valuable measure by allowing incorporation of
2428 farm-level attributes, with greater utility for disease risk planning and policy applications. The
2429 importance of having this level of information was demonstrated during the 2001 FMD outbreak in
2430 the UK, whereby decision makers relied heavily on models built on livestock census and location

2431 data (Ferguson et al., 2001) for two significant reasons. First, farm location data informed disease
2432 containment and eradication strategies (e.g. culling and vaccination) which relied on spatial
2433 relationships between susceptible host populations (Ferguson et al., 2001; Bruhn et al., 2012);
2434 second, farm-level attributes including size, livestock numbers, species mix and number of land
2435 parcels, are all significant risk factors in FMD transmission (Ferguson et al., 2001; Keeling et al.,
2436 2001). This was the first time that mathematical modelling had been used for management of an
2437 outbreak (Haydon et al., 2004).

2438 The quantity and quality of livestock census data available for the 2001 FMD UK outbreak was
2439 unprecedented (Kao, 2002), but despite the acknowledged value, few countries maintain
2440 comprehensive agricultural census data and/or geospatial databases (Bruhn et al., 2012). The cost of
2441 collecting and maintaining databases can be prohibitive, particularly on large- or national-scales,
2442 along with other issues including privacy concerns. The application of SDMs for livestock estimation
2443 could potentially be extrapolated to other countries where little or no data are available provided
2444 reasonable quality, high resolution data on environmental conditions and farming systems are
2445 available and comparable. Good predictions have been obtained for countries where data are only
2446 available at low resolution (Van Boeckel et al., 2011) or comprehensive training data are scarce
2447 (Emelyanova et al., 2009). The machine learning techniques presented here may be better at
2448 predicting to new or under-sampled areas as they have greater ability to work with noisy data and
2449 incorporate complex interactions. A problem facing SDM techniques in the future for livestock
2450 applications is the intensification of farming practises which causes a disassociation between
2451 livestock and their environments, making it difficult to predict populations based on environmental
2452 attributes (Neumann et al., 2009; Robinson et al., 2014). This is already the case for chickens and
2453 pigs in many countries (Robinson et al., 2014).

2454 Animal biosecurity threats will increase with the demand for livestock products, which is projected
2455 to double in the next two decades (Cartín-Rojas, 2012). Several recent high profile biosecurity
2456 outbreaks highlight the value of up-to-date census data. For managers to implement the most
2457 effective disease control strategies they need sufficient information to make informed decisions
2458 (Meyerson and Reaser, 2002) and maintaining or establishing activities for continual preparedness
2459 enables a quick response (Waage and Mumford, 2008). The SDM approaches in this study are
2460 internationally applicable for livestock population estimation and have broad applications in disease
2461 research, biosecurity as well as in policy and planning. In the following chapter, Bayesian techniques
2462 are used with the same data stratifications and predictor data to allow a comparison of the
2463 techniques in Chapter 6 to determine the most effective modelling technique for the NZ situation.

2464

2465

Supplementary tables

2466 **Table 6.5:** *The LSU prediction root mean squared error (RMSPE) using the withheld results for the*
 2467 *regional spatial stratification for both individual farms within regions and to 3km quarantine zones.*
 2468 *Table also shows the mean number of LSU per farm in the region, the standard deviation for all farms*
 2469 *in brackets, and the total number of farms in each region. Results for Quarantine zone predictions*
 2470 *are in 1,000s of LSU.*

Region	Mean LSU per farm (SD)	Number of quarantine zones per region (10 grids)	Total number of farms	RMSPE					
				Farm-level RF	Farm-level BRT	Farm-level KNN	Quarantine zone RF	Quarantine zone BRT	Quarantine zone KNN
Overall	132.7 (341.5)	1997	91662	215.6	216.4	380.2	5.4	4.8	9.0
Auckland	29.7 (112.8)	36	9815	84.5	84.8	122.9	0.7	1.0	1.4
Bay of Plenty	76.7 (372.0)	59	5880	287.6	262.8	394.8	11.6	6.9	10.0
Canterbury	165.1 (462.7)	455	12281	348.8	358.3	526.9	6.5	8.7	24.7
Gisborne	244.8 (556.5)	65	1410	353.5	364.9	638.5	4.1	3.7	4.8
Hawkes Bay	170.0 (419.8)	126	3972	220.3	212.7	426.3	7.1	2.9	10.6
Manwt-Wngn	168.1 (329.2)	180	8759	168.0	159.5	387.3	2.2	1.6	4.4
Marlborough	68.3 (184.1)	60	1900	153.7	186.4	411.6	3.3	4.2	5.1
Nelson	20.7 (59.8)	2	195	31.9	36.4	439.7	0.3	0.7	8.3
Northland	87.1 (200.5)	100	8385	96.7	109.2	255.2	2.7	5.1	5.5
Otago	202.6 (460.9)	382	5038	295.7	325.7	480.9	3.4	3.9	5.8
Southland	259.1 (500.4)	160	4859	338.0	315.8	487.3	6.3	4.6	9.9
Taranaki	134.2 (188.1)	70	5294	109.5	107.3	272.0	2.3	2.4	6.2
Tasman	45.5 (144.5)	26	2637	78.4	74.3	293.7	1.3	1.6	6.8
Waikato	151.7 (292.5)	183	15473	172.9	175.6	281.7	10.6	5.6	12.7
Wellington	92.9 (256.5)	59	4018	116.9	117.5	323.50	1.7	1.3	3.6
West Coast	116.2 (321.1)	34	1746	194.3	191.6	377.2	2.4	6.4	49.5

2471

2472

2473

2474 **Table 6.6:** *The cattle prediction root mean squared error (RMSPE) using the withheld results for the*
 2475 *regional spatial stratification for both individual farms within regions and to 3km quarantine zones.*
 2476 *Table also shows the mean number of cattle per farm in the region, the standard deviation for all*
 2477 *farms in brackets, and the total number of farms in each region. Results for Quarantine zone*
 2478 *predictions are in 1,000s of cattle.*

Region	Mean cattle per farm (SD)	Zones per region	Farms			RMSPE			
			Farm-level RF	Farm-level BRT	Farm-level KNN	Quarantine zone RF	Quarantine zone BRT	Quarantine zone KNN	
Overall	94.4 (243.1)	1997	91662	172.9	171.4	275.3	6.4	3.5	5.2
Auckland	27.7 (113.3)	36	9815	92.1	92.7	122.5	0.7	0.8	1.5
Bay of Plenty	66.1 (197.5)	59	5880	125.6	116.5	208.7	9.8	5.9	8.6
Canterbury	99.6 (345.6)	455	12281	290.4	285.0	377.8	4.0	6.0	16.6
Gisborne	162.3 (407.7)	65	1410	312.7	305.0	427.5	3.3	3.2	3.7
Hawkes Bay	104.0 (275.8)	126	3972	191.5	209.5	315.1	8.8	4.2	4.4
Manwt-Wngn	104.0 (204.1)	180	8759	132.4	126.4	257.8	2.0	1.4	2.8
Marlborough	37.8 (113.3)	60	1900	94.5	122.7	303.1	2.0	2.9	3.7
Nelson	13.6 (45.4)	2	195	30.6	34.5	168.9	0.1	0.4	4.0
Northland	91.3 (203.6)	100	8385	119.4	111.0	231.3	8.0	2.1	9.0
Otago	88.9 (248.4)	382	5038	200.5	203.4	324.7	2.8	2.6	3.1
Southland	132.2 (311.5)	160	4859	257.2	255.5	350.2	2.9	2.8	3.4
Taranaki	124.3 (171.2)	70	5294	109.5	104.1	212.6	3.6	2.5	3.6
Tasman	35.4 (130.7)	26	2637	90.2	86.1	220.8	2.4	2.2	4.9
Waikato	138.4 (256.1)	183	15473	177.4	164.2	252.4	8.9	3.8	7.7
Wellington	54.3 (151.9)	59	4018	89.4	92.5	207.2	1.4	0.8	2.5
West Coast	103.3 (253.3)	34	1746	180.1	200.8	306.8	9.9	2.5	26.8

2479
2480

2481 **Deliverable 5: Report on Bayesian Zero Inflated Poisson**
2482 **Models to predict animal counts – Predicting the spatial**
2483 **distribution of domestic animal populations from**
2484 **incomplete demographic data**

2485 **Mary van Andel¹, Chris Jewell³, Tracey Hollings², Andrew Robinson², Mark Burgman², Paul**
2486 **Bingham¹, Joanna McKenzie⁴, Tim Carpenter⁴**

2487 ¹ *Ministry for Primary Industries, Wellington, New Zealand 6140*

2488 ² *Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Melbourne Australia*
2489 *3010*

2490 ³ *Lancaster University, Lancaster, UK, LA1 4YG*

2491 ⁴ *IVABS, Massey University, Palmerston North, NZ*

2492

2493 **This chapter will be submitted for publication entitled “Predicting the spatial distribution of*
2494 *domestic animal populations from incomplete demographic data”*

2495

2496 **Introduction**

2497 An exotic disease outbreak can be economically devastating for the primary production industries and related
2498 sectors in nations that have substantial primary sector export economies (Blake et al., 2003; Carpenter et al.,
2499 2011). The preparation for and rapid control of outbreaks of exotic animal disease are high priorities for the
2500 Competent Authorities of these countries.

2501 Information on the size and distribution, that is, the demography, of the farm animal populations at risk, and
2502 their proximity to sources of infection are required to adequately prepare for and respond to these outbreaks
2503 (Woolhouse, 2003). Important aspects of planning for exotic disease outbreak management include: defining a
2504 strategy to reduce disease spread, vaccine bank resourcing and identification of the locations of burial or pyre
2505 burning sites for disposal of animal carcasses. Each of these activities relies on knowledge about the farm
2506 animal demography of the area in question (Tomassen, de Koeijer et al. 2002). Impacts of insufficient planning
2507 for disease outbreak management are numerous and may extend many years after the end of an outbreak as

2508 economic losses, ground water contamination and air pollution from carcase disposal sites (Gwyther et al.,
2509 2011; Joung et al., 2013).

2510 Size and distribution of the farm animal populations at risk also fundamentally influence how disease will
2511 spread (Buhnerkempe et al., 2014; Porphyre et al., 2013a). Estimating the risk of spread of infection of
2512 contagious livestock diseases from known infected foci is important for prioritising interventions in the face of
2513 finite disease control resources (Bessell et al., 2010; Carpenter et al., 2004; Haydon et al., 2003; Jewell and
2514 Brown, 2014; Keeling et al., 2001; Thrusfield et al., 2005). Farm animal demographic information facilitates a
2515 more prompt identification of sites that are at high risk of becoming infected and/or at risk of spreading
2516 infection and guides decisions on control strategies early in an outbreak when epidemic information is scarce
2517 (Bessell et al., 2010; Hugh-Jones, 1972; Mansley, 2004; Sarandopoulos, 2015; Tomassen et al., 2002) . This
2518 prioritisation is critical because delays in identification of infected premises (IP) can have a major effect on
2519 epidemic impacts (Carpenter et al., 2011).

2520 Disease spread simulation models, which use animal demographic information, are useful for planning and
2521 preparing for exotic animal disease outbreaks. Models of disease transmission in farm animal populations
2522 typically assume density-dependent transmission, for which the density and distribution of farms and
2523 populations on the farms at risk are critical components of the effective reproduction number, R_e (Dohoo et
2524 al., 2009, Porphyre et al., 2013). In the model, estimates of R_e indicate whether an outbreak is under control
2525 and can inform policy adjustments in the face of an outbreak (Ferguson et al., 2001; Paine et al., 2010).

2526 Some models require only the presence or absence of a susceptible animal population for each geographic
2527 unit. An example of this would be the disease spread model InterSpread Plus (ISP) which does not use farm-
2528 level animal counts in its disease spread simulations (Stevenson et al., 2013). Instead, it uses a farm type
2529 classification, which is based on the type of animals on the farm, with an average probability of transmission
2530 distribution for each farm type generated from average farm populations on each farm type. Farm-level
2531 animal counts are however required when calculations of economic impact and numbers of animals culled or
2532 vaccinated are extrapolated from ISP model outputs. Other more detailed disease process models which
2533 include within herd spread are used to explore the impacts of control decisions at the farm level (Thomas W
2534 Bates et al., 2003; Bradhurst et al., 2015; Halasa et al., 2013; Porphyre et al., 2013a).

2535 Competent Authorities in a number of countries maintain national animal registries of farm and livestock
2536 population distribution (Paiba et al., 2007). The maintenance of accurate data within central animal registries
2537 is an ongoing challenge; particular areas of difficulty include: recording of births and deaths, real time animal
2538 movements, and mismatches and inconsistencies between different sources of data in the repository (Jewell
2539 et al., 2015; Savill et al., 2007). Previous studies have addressed a lack of animal population data by simulating
2540 livestock populations within a country (Burdett et al., 2014) or globally (Robinson et al., 2014). While simulated
2541 livestock population data are available on a global scale at 1 km² resolution from the Food and Agriculture
2542 Organization's Gridded Livestock of the World Project (Robinson et al., 2014), this is not sufficiently accurate
2543 at the farm level to plan disease control operations, which require knowledge of the numbers and species
2544 present on each IP (Mansley, 2004). An alternative approach, which is required in some countries, where
2545 details of farm locations and their animal populations are not maintained, is to model the location of farm
2546 sites and farm types based on environmental variables (Burdett et al., 2015).

2547 New Zealand has multiple sources of animal demographic information (Jewell et al., 2015), which are
2548 maintained by the government and private industry bodies. AgriBase is owned by AsureQuality Limited, a New
2549 Zealand state-owned enterprise that provides operational services to the Ministry for Primary Industries (MPI).
2550 AgriBase contains spatially referenced farm boundary data and species-specific counts of animal populations
2551 for each farm. MPI maintains a spatially referenced properties database FarmsOnLine (FOL) for biosecurity
2552 purposes but livestock counts are not available in this data source. Both AgriBase and FOL hold information on
2553 properties used for all farming activities including those used for recreational farming or "lifestyle" farming as
2554 it is known in New Zealand. Given the costs and challenges of maintaining an accurate database of farm-level
2555 animal populations, a process of estimating livestock counts would be very useful to populate and/or update
2556 animal population data for farms. We also wanted to investigate whether estimates generated from one
2557 database (AgriBase) would successfully provide the data needed to predict the animal counts in another
2558 database that does not contain animal counts(FOL), for which data matching between the two databases is not
2559 possible (Jewell et al., 2015).

2560 The objective of this paper is to present a methodology for predicting counts of livestock units (LSU) and of
2561 cattle present on farms in New Zealand using environmental and socio-economic data. The balance of the
2562 paper is constructed as follows. In the next section we describe the data sets, data handling, and modelling

2563 strategy, and the means by which we assessed and compared the models. We then provide a detailed
2564 description of the results, Discussion, and our Conclusion. The Appendix provides a detailed description of
2565 statistical summaries of the data.

2566 **Table 7.1:** *Species equivalence values used in the models constructed to predict counts of livestock*
2567 *units (LSUs) on rural properties in New Zealand. These values are calculated following the European*
2568 *Union definition of individual species requirements (Eurostat, 2013)*

Species	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

2569 **Materials and Methods**

2570 **Databases and study variables**

2571 *Databases and study variables:*

2572 The model was built using farm boundary and animal population data from AgriBase (Sanson and Pearson,
2573 1997). Some farm boundaries in the dataset overlapped one another as they are farmed under cross lease
2574 agreements. These 6586 (4.5% of the total available farm records) overlapping farms were deleted from the
2575 dataset to avoid bias in predicting animal counts. In this case both of the overlapping farms were deleted. We
2576 deleted a further 36815 properties (27.02% of the total available farm records) which had no information on
2577 production type of any kind (agriculture or horticulture based) because we assumed that these represent
2578 missing data rather than the true absence of the animal species of interest. The remaining 91601
2579 farms(68.50%) were used for model building and validation.

2580 The two outcome variables of interest were farm level counts of LSUs and of cattle; both were drawn from the
2581 AgriBase dataset (Sanson and Pearson, 1997). LSUs were calculated for sheep, goats, cattle, pigs, deer and
2582 horses following the European Union definition of individual species requirements in relation to a dairy cow
2583 (Anonymous, 2013) shown in Table 7. 1. Poultry were not included in the model.

2584 Publicly available land cover, environmental and socio-economic data held by Land Information New Zealand
2585 (LINZ), Statistics New Zealand, Land Environments of New Zealand (LENZ) and Land Resources Information
2586 New Zealand (LRINZ) were used to generate candidate predictor variables (Table 7.2). The New Zealand
2587 Transverse Mercator (NZTM) projection was used for geographic information projection in this analysis.
2588 Predictor variables were loaded into a geographic information system (GIS) relational database (PostgreSQL
2589 8.4 with PostGIS 2.2.0 extensions). This provided a convenient platform for interfacing with statistical software
2590 (Rv3.0.1, Foundation for Statistical Computing, Vienna, Austria), enabling large amounts of geospatial data to
2591 be processed and extracted for each farm level. Predictor variables included in the models, where they were
2592 sourced and how they were derived are presented in Table 7.2.

2593 A Spearman correlation matrix was used to identify significant pairwise collinearity amongst the predictor
2594 variables, defined as a correlation of $|r| > 0.7$. Collinearity was identified between the following pairs of
2595 variables: mean *annual temperature* and mean *minimum temperature*, mean *solar exposure* and minimum
2596 *solar exposure*, and *slope* and *aspect*. Collinearity between the *solar exposure* and *temperature* variables was
2597 addressed by creating new variables using the difference (T) between the mean *annual temperature* (MAT)
2598 and mean *minimum temperature* (MMT) for each farm (*i*) so that:

$$T_i = MAT_i - MMT_i$$

2599 and the difference (S) between mean *annual solar exposure* (MASE) and *minimum solar exposure* (MMSE) for
2600 each farm (*i*) so that:

$$S_i = MASE_i - MMSE_i$$

2601 Mean *annual solar exposure* and mean *annual temperature* were retained in the models along with the
2602 bivariate transformations T_i and S_i .

2603 *Aspect* defines the orientation of the land surface and is measured as a compass bearing with 0 degrees
2604 representing a surface facing true north. Given this system of measurement, a simple average or standard
2605 deviation is not a meaningful summary statistic and it is necessary to further transform *slope* and *aspect*. The
2606 method proposed by Stage (1976) was used to integrate the effects of slope and aspect so that for each farm
2607 the slope and aspect were summarised across the farm polygon by including the product of slope and the
2608 cosine function of aspect and the product of slope and the sine function of aspect (Stage, 1976). Mean and

2609 standard deviation of elevation for the farm were calculated by averaging each raster cell of a surface
2610 generated by a digital elevation model contained within the farm polygon to take into account the effect of
2611 elevation (Stage and Salas, 2007).

2612 An interaction term was included to represent the relationship between *deprivation index* and *distance from*
2613 *urban centres* to capture the population of lifestyle farms. Statistics New Zealand notes that this population of
2614 lifestyle farms in rural areas with high urban influence has the highest median and mean income (and
2615 household expenditure) of any demographic in New Zealand (Pink, 2011). Land held by this population
2616 demographic, which commutes to city centres for work, is thus unlikely to be used in the same way as land
2617 further from cities, where agriculture is the principal source of income. The log of the area in hectares of each
2618 farm covered by *high quality pasture*, *low quality pasture*, *native forest* and *endemic forest* and the
2619 complement of the farm area were used in the final model. Each of these was offset by 0.01 to make the log
2620 transformation possible for farms with no coverage of some of the categories. The log transformation of these
2621 variables was used in the modelling to promote for numerical stability.

2622 *Statistical model framework*

2623 A zero-inflated Poisson (ZIP) regression model was used to generate the predicted counts of LSUs and of cattle
2624 for farm properties. A ZIP model was chosen because many rural properties are not engaged in livestock
2625 production; AgriBase records 37% of properties as having no animals present (Jewell et al., 2015).
2626 Consequently, the ZIP model fit the data significantly better than did a comparable Poisson model.

2627 The model was fitted to farms using a Bayesian approach that calculated a distribution of the predicted
2628 number of cattle or LSUs within a farm polygon on the land surface in New Zealand. The outcome variable for
2629 an individual farm (either number of cattle or LSUs) was modelled as a zero-inflated Poisson (ZIP) random
2630 variable, conditional on the values of the predictor variables for that farm (Mullahy, 1986). A two-stage
2631 process was thus used, in which (i) a Bernoulli random variable first represented the probability of animals
2632 being present on a farm; (ii) then, assuming that animals are present, the number of animals was assumed to
2633 follow a Poisson distribution with mean related via a log link function to the predictor variables. Both the
2634 Bernoulli and the Poisson stages involved modelling using the suite of predictor variables.

2635

2636 The advantage of using Bayesian inference to estimate the coefficients of the ZIP model is that a predictive
 2637 distribution of the number of animals on a farm may be calculated, without resorting either to asymptotic
 2638 assumptions or complex bootstrapping procedures (West, 2014). This provides a precise measure of
 2639 uncertainty surrounding predicted animal numbers for each farm. Samples from the predictive distribution
 2640 may then be used as input for Monte Carlo disease simulation models, such as the Davis Animal Disease
 2641 Spread model (DADs) and the Australian Animal Disease Spread model (AADIS), so that uncertainty in the
 2642 population at risk of infection is reflected in the simulated outputs (Bates et al., 2003; Bradhurst et al., 2015).

2643 In our approach, \mathbf{y} denotes the vector of outcomes (cattle or LSU counts) for each of $i = 1, \dots, n$ farms and \mathbf{X}
 2644 the $n \times 18$ design matrix of predictor variables shown in Table 7.2. Under the ZIP model, it is assumed that

2645

$$y_i = \begin{cases} 0 & \text{with probability } p_i \\ z_i & \text{with probability } 1 - p_i \end{cases}$$

2646 where

$$z_i \sim \text{Poisson}(\lambda_i)$$

2647

2648 Logit and log-linear models are then specified for p_i and λ_i respectively, such that

$$\log\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x}_i^T \boldsymbol{\gamma}$$

2649 and

$$\log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

2650 where \mathbf{x}_i represents the i th row of the design matrix of predictor variables \mathbf{X} (Table 7.2), and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are
 2651 column vectors of coefficients. In a ZIP model, a portion of the zero predictions arise from the Bernoulli
 2652 component with the remainder from the Poisson component (Lambert, 1992).

2653 Independent Normal (0, 10) prior distributions were chosen for each component of the coefficient vectors $\boldsymbol{\gamma}$
 2654 and $\boldsymbol{\beta}$. The model was fitted to the data (\mathbf{y}, \mathbf{X}) using a Markov Chain Monte Carlo (MCMC) algorithm, from
 2655 which the joint posterior distribution $\pi(\boldsymbol{\gamma}, \boldsymbol{\beta} | \mathbf{y}, \mathbf{X})$ was estimated numerically. This normal prior was selected

2656 as the coefficient values were expected to fall between -20 and 20 – the interval encompassed by 95% of this
2657 distribution.

2658 The MCMC was coded in R using a single-site adaptive Metropolis-Hastings step for each coefficient (Jewell,
2659 2016a).

2660 A total of 50 000 iterations of each model was performed and the first 20 000 runs were discarded as burn-in
2661 leaving 30 000 iterations which were thinned systematically every 30 iterations to 1000 observations to be
2662 used in all further analyses.

2663 *Predictive performance:*

2664 For prediction purposes, the Bayesian predictive distribution for the outcome variable y^* of the target data X^*
2665 (i.e. the values of the predictor variables calculated for each farm polygon in the target dataset) is estimated
2666 numerically by simulating from the ZIP model specified above, using draws of the model coefficients γ and β
2667 as represented in the MCMC output.

2668 The rank probability score (RPS), which compares a predictive distribution against an observed value, was used
2669 to assess the performance of the Bayesian prediction using a variety of scales of validation data sets (Brier,
2670 1950). The RPS is defined as

$$RPS(\Pi, y) = \frac{1}{K-1} \sum_{k=0}^K (\Pi(k) - \mathbf{1}[k \geq y])^2$$

2671 where Π is the cumulative mass function for a discrete probability distribution representing a prediction, y is
2672 the observed value, and K is the largest value of outcome variable, predicted or observed, for any farm in the
2673 dataset. Small values of the RPS indicate a good prediction, taking the value 0 if the predictive distribution
2674 takes point mass on the observed value (i.e. a “perfect” prediction with no bias or uncertainty). In the case
2675 where the posterior prediction is a point estimate, the RPS reduces to the mean absolute error.

2676 The predictive performance and internal model validation were evaluated in five ways (i to v). The first two
2677 were: i) an examination of spatial autocorrelation of RPS in two large regions (figure 1), and ii) the differences
2678 in model performance amongst farm types. In these analyses the farm is the unit of interest. The third and
2679 fourth methods were performed at iii) small (3km radial zone) and iv) larger (New Zealand administrative
2680 regions) spatial scales and calculated area level RPS values. In these analyses the area (either the region or the

2681 3km zone) is the primary unit of interest. The final method v) used Receiver Operator Characteristic (ROC)
2682 curves to examine the performance of the Bernoulli component of the model.

2683 We chose 3km radial zones as a unit of interest to examine model performance because the outcomes (farm
2684 level counts of cattle or LSUs) of these models are primarily required for biosecurity purposes. In the event of a
2685 suspected or actual outbreak of FMD, the population in the 3km radial zone around an infected farm is at high
2686 risk of local spread (Thrusfield et al., 2005). This means that accurate data on susceptible animal populations
2687 on farms within 3km of a farm with a suspected outbreak of FMD should be available as early as possible in the
2688 investigation and risk assessment process.

2689 i) Spatial autocorrelation of the RPS (as an analogue of a frequentist model residual) for each farm
2690 in New Zealand was examined using variograms (Diggle and Ribeiro, 1998). Given the size of the
2691 dataset a parallel implementation of the variogram function was written as an R package, making
2692 use of Graphics Processing Unit-accelerated computation (Jewell, 2016b). The variogram was
2693 plotted, and visually examined for evidence of spatial autocorrelation. This was done for an area
2694 in the North Island that comprised the Waikato Region, the Bay of Plenty Region and the
2695 Auckland Region and for an area in the South Island comprising the Southland Region, the Otago
2696 Region and the Canterbury Region (figure 1). The analysis was limited to 350km, being
2697 approximately half the maximum distance between farms in either selected area, with a step size
2698 of 10km for both the LSU and the cattle models. Bootstrapping was used to calculate a null
2699 distribution for the variogram, under the assumption of no spatial correlation (Diggle and Riberio,
2700 1998). The 5th and 95th percentiles of the variogram null distribution are shown by the dotted
2701 lines in Figure 7.2.

2702 ii) To assess how farm type affected model fit, the RPS for farms classified as dairy farms, pastoral
2703 livestock farms, pig breeding and pig finishing establishments and properties used for grazing
2704 young stock and dairy cattle not currently in production was compared. Results of this
2705 comparison for both the cattle and the LSU models were visually examined as boxplots and were
2706 statistically tested with the Kruskal-Wallis test for homogeneity amongst independent non-
2707 parametric groups (R-Core Team, 2013).

2708 The models for the area level validation were built from subsets of training farms with target or validation
2709 farms selected spatially either by 3km radial zones or by New Zealand administrative region. While the primary
2710 unit of interest for these analyses is RPS at the level of the specified area, mean farm level RPS values are also
2711 presented.

2712 iii) To assess model performance at the scale of 3km radial zones, 10 sets of 200 3km zones were
2713 selected by randomly selecting 200 cells from a grid of 9000m x 9000m cells which was overlaid
2714 across a map of New Zealand. Each zone was generated by creating a 3 km buffer around the
2715 farm polygon that intersected the centroid of each randomly selected grid cell. We chose to use
2716 200 3km radial zones following experimentation which found a good trade-off between having
2717 enough 3km radial zones for each model instance to provide sufficient information on the quality
2718 of fit, but not having so many that overlap between farms within adjacent zones became a
2719 substantial issue. All farms that intersected with this 3km buffer zone were extracted using
2720 spatial libraries (Bivand et al., 2014a, 2014b; Pebesma et al., 2014) in R (R-CoreTeam, 2013). The
2721 training data comprised all properties not selected by this method. Models were validated by
2722 comparing the observed and predicted counts of LSUs or cattle in zones in the validation data set.
2723 The posterior prediction for each farm within each zone was aggregated and regarded as the
2724 posterior prediction for the zone. The RPS was calculated by comparing the predicted zone-level
2725 counts against the observed counts aggregated for the farms in each zone. The mean RPS at the
2726 zone level was then the sum of all zone-level RPSs divided by the number of zones (200). This
2727 process of model building and validation for the 3km areas was repeated for 10 independent sets
2728 of randomly selected training and target zones. The results were summarised across the 10
2729 model fits by calculating the mean and standard deviation for RPSs at the level of the 3km zones.

2730

2731 iv) To assess model performance at the scale of regions, training and target datasets were generated
2732 by stratifying first by island (North Island and South Island), and then randomly selecting 6 of 9
2733 regions from the North Island and 5 of 7 regions from the South Island to act as training data. The
2734 remaining 3 North Island regions and the 2 South Island regions were target data sets for
2735 validating each of the LSU and cattle model fits. Region was removed as a fixed effect from the
2736 model. Fifty independent training and target datasets were constructed and both the cattle and

2737 LSU models were set to run 50 000 iterations on each dataset. This analysis aimed to identify if
2738 predicted LSU or cattle counts were more reliable in particular regions. The posterior prediction
2739 for each farm within each region was aggregated and regarded as the posterior prediction for
2740 each region. The RPS was calculated by comparing the predicted counts against the observed
2741 counts aggregated for the farms in each region. The mean RPS at the region level was the sum of
2742 all region-level RPSs divided by the number of regions.

2743
2744

2745 v) As a final predictive performance step, the performance of the Bernoulli part of the model (which
2746 predicts presence or absence of LSUs and cattle on farms) was examined using ROC methodology
2747 (Zweig and Campbell, 1993). The conventional use of a ROC curve establishes the test
2748 characteristics (sensitivity and specificity) of a diagnostic test. Sensitivity is defined as $(T+ | D+)$ –
2749 the probability of a diseased animal (D+) returning a positive test (T+) and specificity is $(T- | D-)$ –
2750 the probability of a non-diseased animal (D-) returning a negative test result (T-). The ability of
2751 the model to predict the absence of animals was regarded as T+ (or Pred Zero +), the absence of
2752 animals in the data as D+ (or Data Zero +), the prediction of animals present as T- (Pred Zero -)
2753 and the presence of animals on the property in the data as D- (Data Zero -). This can be
2754 summarised as follows:

2755
$$\text{Sensitivity} = (\text{Pred Zero +} | \text{Data Zero+})$$

2756
$$\text{Specificity} = (\text{Pred Zero -} | \text{Data Zero-}) = (\text{Pred Animals} | \text{Data Animals})$$

2757 The x-axis of the ROC curve showed $1 - (\text{Pred Animals} | \text{Data Animals})$. The y-axis represented $(\text{Pred}$
2758 $\text{no animals} | \text{Data no animals})$. For each farm, the posterior predictive probability of absence of
2759 animals was calculated. A threshold was then imposed on this probability as the Bernoulli
2760 prediction of whether a farm had animals present or not. A ROC curve was constructed to
2761 compare this prediction against the real data under different values of the threshold. This
2762 process was completed for both the LSU model and the cattle model using the whole dataset as a
2763 training dataset in each instance.

Table 7.2: Predictor variables included in Bayesian ZIP model predicting counts of cattle and counts of livestock units (LSU) relative to dairy cattle on New Zealand farms. * Predictor variables included in the final models.

Predictor Variable	Description (units)	Reference
Distance to sealed roads*	Distance from the centroid of the farm property to the nearest major sealed road (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to urban centres *	Distance from the centroid of the farm property to the nearest urban centre (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to rivers *	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 *	Value at the farm centroid (categorical variable between 1 and 10 with 1 being the greatest degree of deprivation)	Statistics New Zealand 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) *	Number of hectares 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)*	Number of hectares 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 daily 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 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*	Created by finding the difference between the mean solar radiation and the mean winter solar radiation for each farm	
Mean water deficit*	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*	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Cosine of Slope*	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Standard deviation of slope*	Standard deviation of slope for the farm	
Mean and standard deviation of elevation*	Mean and standard deviation of farm elevation for the farm calculated by averaging each raster cell of a raster 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*	One of the 17 New Zealand Regional	https://koordinates.com/layer/4240-

	councils that the farm centroid falls within (categorical variable)	nz-regional-councils-2012-yearly-pattern/
Size of farm	farm area (ha)	Agribase™ for the farm boundaries and post gis
Forest cover (log of)*	The union of endemic and exotic forest cover for each farm (ha)	https://koordinates.com/layer/168-nz-native-polygons-topo-150k/ & https://koordinates.com/layer/131-nz-exotic-polygons-topo-150k/
Remaining land cover (log of)*	Remaining size of the farm once high quality pasture, low quality pasture and forest cover had been taken from the total (ha)	

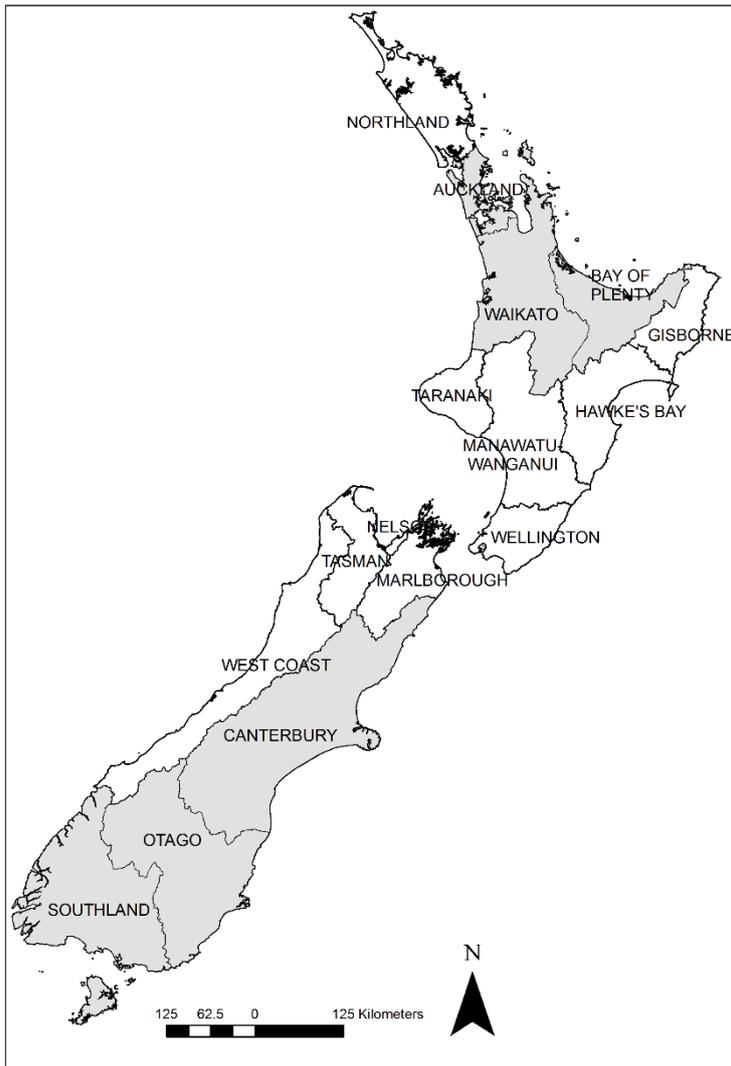


Figure 7.1: Map showing the 16 regions and the two main islands of New Zealand. Grey shaded areas represent the areas used to compute variograms in the North Island and South Island when the spatial autocorrelation of model fit (shown in this case by the rank probability score) is presented.

Results

Predictor variables

Covariate values for the Bernoulli and Poisson components of the ZIP models and their associated odds ratios and risk ratios are presented in the appendix. 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 most important predictor variable for counts of LSUs and cattle (the Poisson part of the models).

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 7.1). This correlation was apparent up to about 18km in both the North and South Islands 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 in both the LSU and cattle models indicating a higher level of spatial correlation of model error (figure 7.4).
- ii) Significant differences in mean RPS values were found for groups of farms stratified by the farm class recorded in AgriBase (dairy farms, pastoral livestock farms, pig breeder and finisher farms, grazing and dry cow 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 class 'dry stock' which included contract grazing and dry cow properties were predicted with the least accuracy (figure 7.3).
- 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 7.3).
- 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 7.3). 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 7.2).

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 (sensitivity) (figure 5). The specificity of this estimate was poor (40%) with many false positive 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 (figure 7.5).

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 7.3: Mean and standard deviation of rank probability scores (RPS) for a model to predict livestock units (LSUs) and cattle on New Zealand farms. RPS are reported for 2 models where training data were withheld first at a 3km zone level and subsequently at a regional level. Results are then examined at the model level and at the area level for each of these instances.

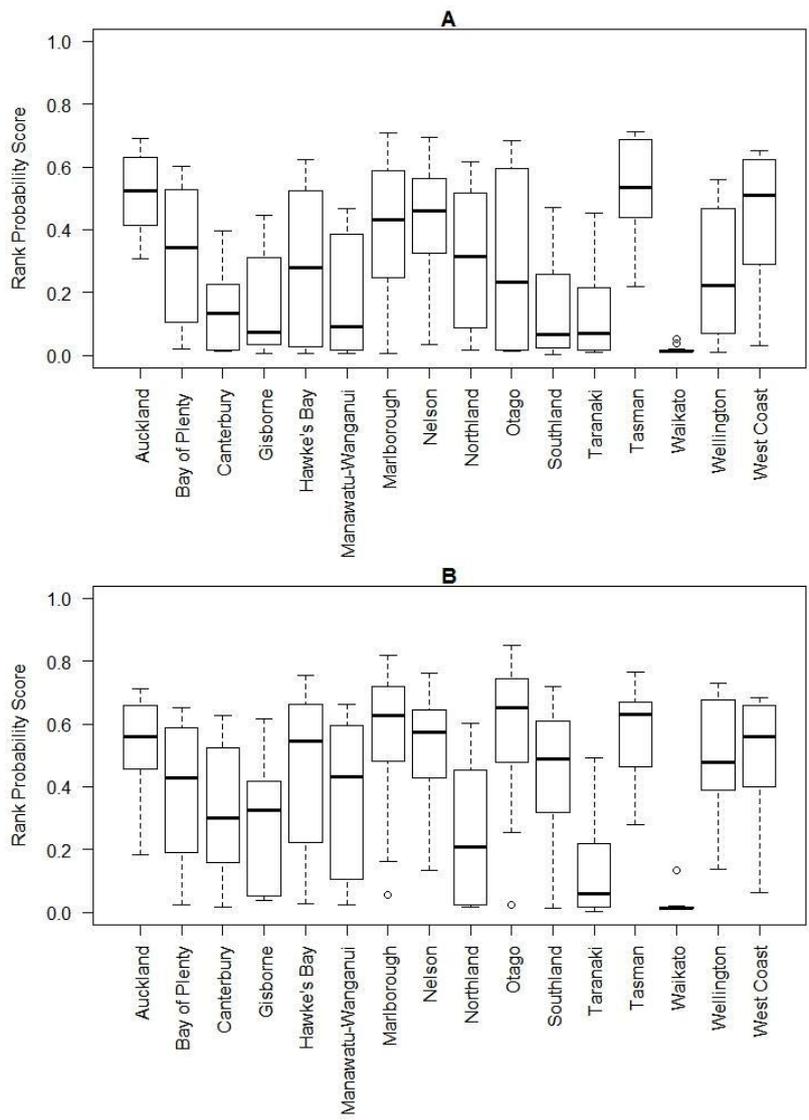


Figure 7.2: Box plots of the Rank Probability Scores (RPS) from 50 resampled fits of our model where region was removed from A, the LSU model and B the cattle model. Data were divided into a training (6 regions from the North Island and 5 from the South Island) and target dataset (the remaining 5 regions) by a stratified random sample of for each of the 50 resampled fits. In each instance RPS was calculated at the region level. All 50 fits were then aggregated and the RPS for each region averaged over all 50 fits to produce the boxplots which show the mean RPS for each region.

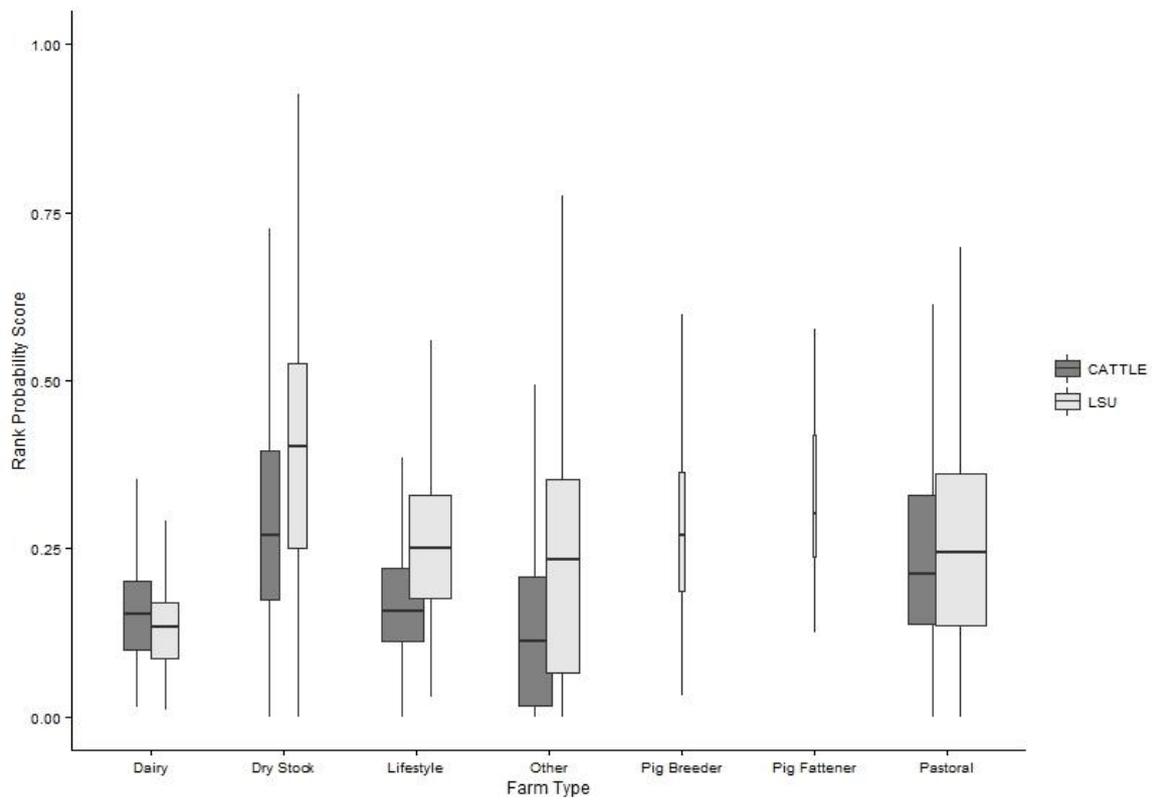


Figure 7.3: Boxplot of rank probability scores (RPS) for each of the 7 farm classes in AgriBase for models predicting livestock units (A) and cattle (B) on rural properties in New Zealand. This figure shows that in both models predictions for grazing and dry cow properties are the least accurately represented. Box widths are proportional to the square root of the number of observations present in each group.

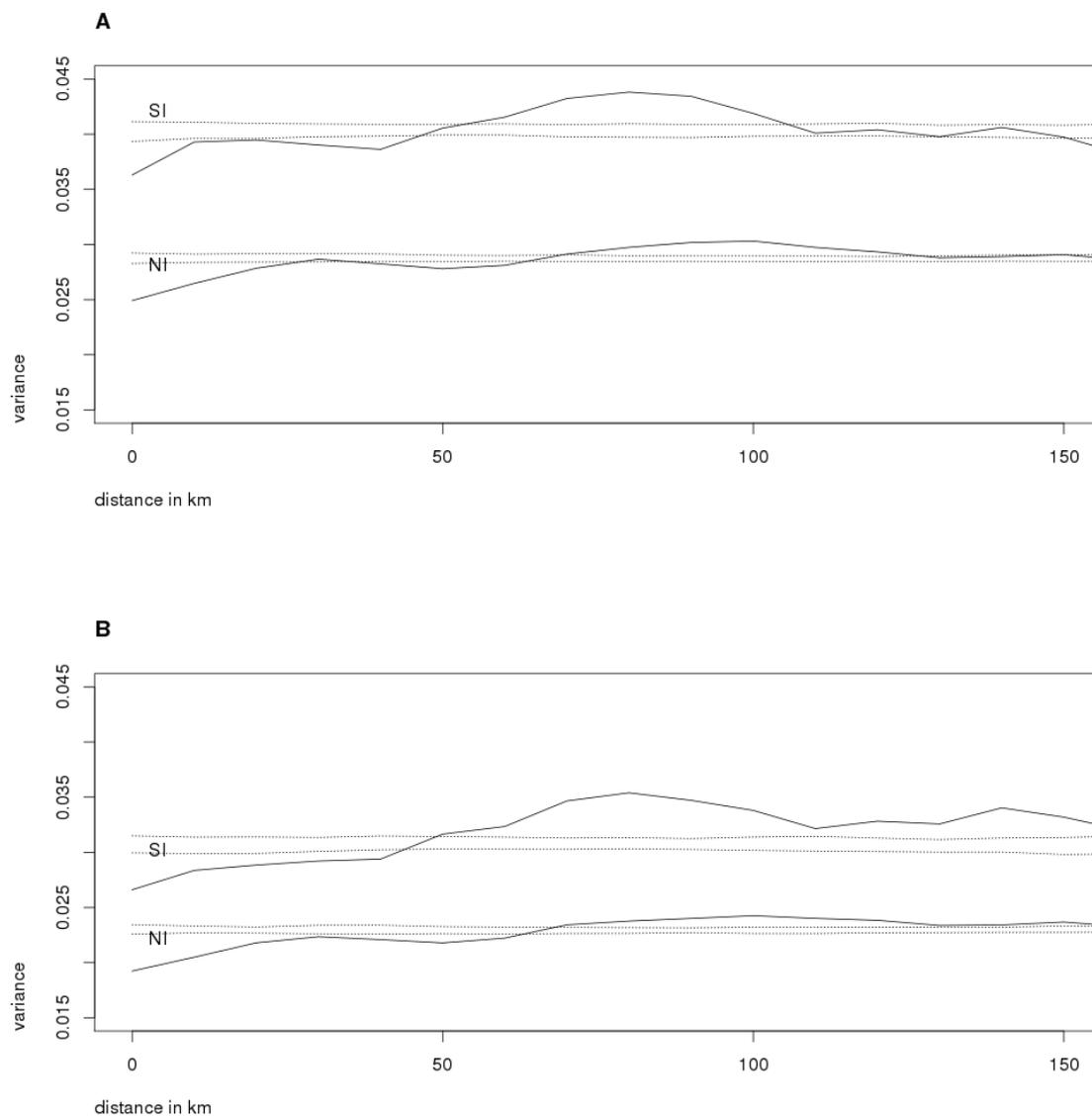


Figure 7.4: Variograms for the LSU model (A) and the cattle model (B) showing the spatial covariance in the farm-level Rank Probability Score (RPS) in an area chosen in the North Island (NI) and the South Island (SI). The dotted lines represent the 5% and 95% of the prediction interval estimated by calculating a variogram for 100 random permutations of the real RPS values for each region.

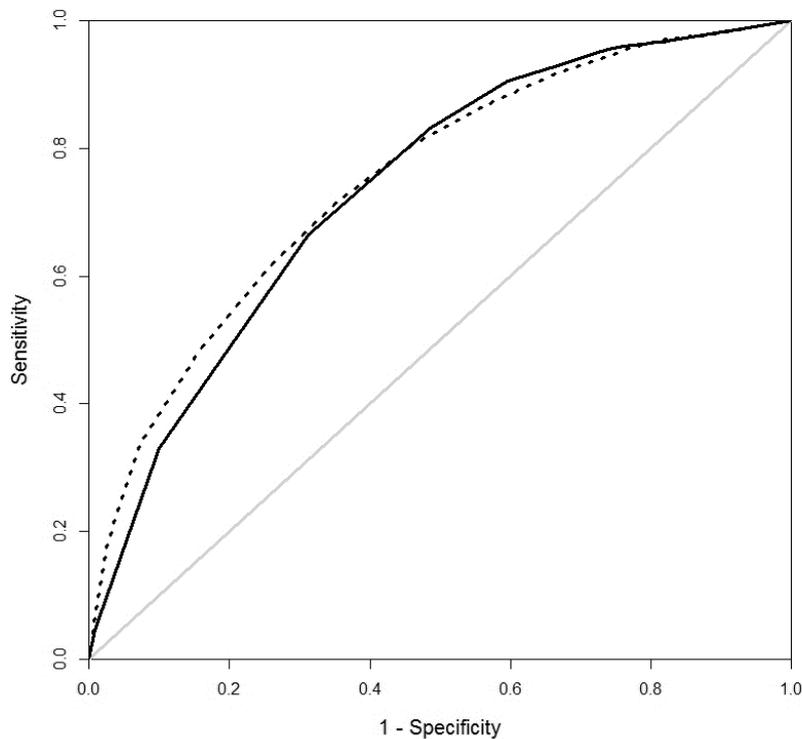


Figure 7.5: Receiver operating characteristic (ROC) curve for a model predicting the absence of cattle (dotted line) and the absence of livestock units (LSUs shown with a solid line) on rural properties in New Zealand.

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 models 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 comprising the log 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 populations 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 the appendix). We believe that these covariates still explain some of the variance (and improve prediction) without necessarily explaining an amount of variance beyond the traditional 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 7.3), predicted counts were less accurate for farms classified as contract grazing and dry stock. These results could be due to real differences in practices such as constantly changing stock numbers on dry stock 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 dry stock 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 7.4). 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.

A useful characteristic of our model is the ability to predict LSU and cattle counts in these small zones with a high degree of accuracy reflected by the lower RPS for the 3km radial zones compared with the predictions for individual farms (Table 7.3).

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).

An area of future exploration of the modelling approach described in this paper is that in addition to predicting animal populations at the farm level, it generates a measure of uncertainty for each population which can be incorporated into stochastic disease simulation models. 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 and Ryan, 2012; Tildesley et al., 2008). 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 (Thomas W 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.

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 present on a farm but it is not able to predict with total 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 (Marshall 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 as a driver for dairy farmers to increase herd size. 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 farm. 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 validation, 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. Predictions were not accurate enough to be used to plan for farm level disease control activities. Future work will provide external validation of the models using data collected from the field to fully assess the precision of the estimates of livestock numbers on farm and to test the extensibility of the derived covariates to prediction to 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 data sources by aligning data collection efforts undertaken by government and industry and by the use of a single, national farm identifier.

Deliverable 6: Model validation against external data

sources

Mary van Andel¹, Daan Vink¹, Katrin Sattler¹

¹ Ministry for Primary Industries, Wellington, New Zealand 6140

Introduction

We used multiple methodologies to validate predictions from a frequentist model of animal counts at the farm-level in a Bayesian framework and from a random forest model. These two models are described in detail in chapters 6 and 7. Both models were used to predict counts of cattle and counts of livestock units (LSUs) for each New Zealand farm. It has been mentioned more than once that data in centrally held databases may not necessarily reflect the state of the population in the field (Honhold and Taylor, 2006) and when considering the usefulness of modelled data from a model trained on such a database, ground-truthing is of great concern. In our case, even the information which we have used to validate our modelled results are likely to contain bias.

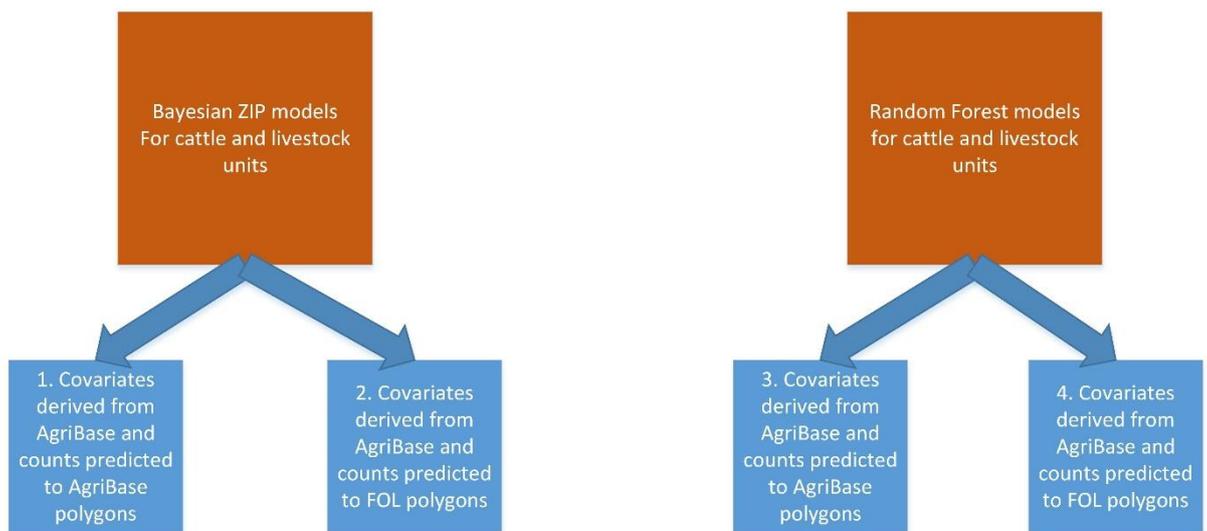
Materials and Methods

Data sources

AgriBase (Sanson and Pearson, 1997) and FarmsOnline (FOL) have been described earlier in this report (Jewell et al., 2016). The Agricultural Production Survey (APS) data collected by Statistics New Zealand survey only includes those farms registered for GST purposes (MacPherson, 2015). This means that no predictions of animal counts on non-commercial holdings are able to be validated by this means. No farm-level identifier is included in the APS survey or census; instead the system makes use of a GST registration number and a physical address. Statistics New Zealand conducts the APS census once every 5 years and surveys a stratified random subset of 30 000 properties every year which are estimated to provide a reliable estimate at the Area Unit level (a spatial aggregation of mesh blocks which is smaller than the NZ territorial local authority). The last census was in 2012. Every year the results of the survey are used to predict mesh block level results; the 2015 results are used in this analysis. Farm-level results from the APS from the most recent census (2012) are used. Figure 8.1 sets out the different comparative analyses schematically.

Mesh block level validation: These data are then aggregated to the mesh block level (the smallest level of aggregation) among larger aggregate levels for use in analysis. The size of a mesh block depends on the number of people and type of area covered. In rural areas, mesh blocks usually have 60 people and in urban areas they have 110 people. Mesh blocks are used to aggregate to area units, territorial local authorities (n=73) and finally, regional councils (n=16). As in prior chapters of this report, livestock units (LSUs) were calculated in accordance with European Union definition of a livestock unit in relation to a dairy cow (Eurostat, 2013) using only those species which were consistently present in the validation data sources. LSUs are calculated as the sum of dairy cattle + beef*0.8 + deer*0.8 + pigs*0.4 + sheep*0.1.

Farm-level validation: Validation at the farm-level used data from two sources. The first was the confidential APS farm-level data which was matched by address to FOL address fields (see Appendix 4.3 for detailed methodology) yielding 12,928 direct matches on address with farms present in each of the 73 local territories (Figure 8.3). A direct match means a match which was able to be made without any manipulation of data to link two records. Each of these 12,928 records is expected to have a high degree of reliability based on the methods presented in appendix 4.3. The second source of validation data was an unstructured survey conducted with veterinarians regarding animal counts in their practice area (see Appendix 4.4 for detailed methodology). LSUs were calculated in the same way as for the mesh block level validation.



Each of the 4 blue boxes above were then compared with different data samples for each of the cattle and live stock unit predictions:

- A. Agricultural Productions Census Data at the mesh block level* (Analysis 2&3)
- B. Agricultural Production Census Data at the property level (for ~13K properties) – for blue block 2 and 4 (Analysis 4)
- C. Vet survey data at the property level (for ~360 properties) – for blue block 2 and 4 (Analysis 5)

Analysis 1 compared the Agricultural Productions Census Data at the mesh block level* with AgriBase data (not modelled data)

Figure 8.1: Graphical representation of the different comparisons carried out between real and modelled data for chapter 8. Note that in analysis 1 the “real” AgriBase data are compared with the 2015 Agricultural Production Survey data at the meshblock level. After that, all analyses 2 and 3 data predicted to AgriBase and FOL are compared to 2015 meshblock data. In analysis 4 data predicted to FOL are compared to farm level Agricultural Production Census data (2012) and in level 5 the same predictions are compared with a survey of veterinarians. * As no farm identifier is included in any Statistics New Zealand Census or Survey activity, FOL and AgriBase polygons were spatially matched to mesh blocks and selected based on farm type for inclusion in the comparison. Methods are shown in Appendix 4.1.

Method 1: AgriBase values compared with APS data at the mesh block level

The first analysis compared counts of cattle and weighted LSUs between AgriBase properties and Statistics New Zealand mesh block level data. AgriBase properties were spatially joined to Statistics New Zealand mesh blocks by recording the mesh block identifier that intersected the weighted centroid of each farm polygon. This was done in the geographic information system QGIS (Quantum GIS Development Team, 2016). All properties in AgriBase categorised by the farm class 'lifestyle' were excluded from the analysis at this point (to limit the predicted farms to those that would be GST registered). Alternate methods tested for identifying AgriBase properties most likely to be registered for GST were examined and are listed in Appendix 4.1.

Method 2: AgriBase Random Forest and Bayesian ZIP model predicted values compared with APS data at the mesh block level from the 2015 census

Methodology 2 and 3 use models built with the outcome variable cattle count and the weighted LSUs of beef and dairy cattle, pigs, sheep and deer. Predicted counts of cattle and predicted LSUs were then compared at the mesh block level. AgriBase properties were spatially joined to Statistics New Zealand mesh blocks by recording the mesh block identifier that intersected the weighted centroid of each farm polygon. This was done in QGIS (Quantum GIS Development Team, 2016). All properties in AgriBase where the farm class was 'lifestyle' were excluded from the analysis at this point. The posterior predications for each AgriBase farm generated by the Bayesian ZIP model were then joined to this dataset and aggregated by the mesh block that they intersected. The mesh block level RPS was calculated by comparing the aggregated posteriors for each mesh block with the APS estimate for each mesh block. The mean RPS across all mesh blocks is shown in Table 8.6. Figures 8.1 and 8.2 display the mesh block level RPS values and their distribution across New Zealand. Mean absolute error (MAE) was calculated from the mean of the posterior prediction from the Bayesian models and from the predictions of the Random Forest models and is shown in Table 8.5 by comparing these predictions with the results of the Agricultural Production Census at the mesh block level..

Method 3: FOL Random Forest and Bayesian ZIP model predicted values compared with APS data at the mesh block level

FOL properties were spatially joined in QGIS to Statistics New Zealand mesh blocks by recording the mesh block identifier that intersected the weighted centroid of each farm polygon. All properties in

FOL where the final four digits of the FOL identifier were 9999 were removed (these polygons are holding places for public land between farms and have no stock on them). Categories of farms in FOL were examined. FOL has 45 land use categories. From these 45 categories dairy, pastoral, deer and pig farms were likely to be GST registered and included in the analysis. Appendix 4.2 is a table of FOL farm types and counts of farms in each type. The posterior predictions for each FOL farm generated by the Bayesian ZIP model were joined to this dataset of farms and aggregated by mesh block. The mesh block level RPS was calculated by comparing the aggregated posteriors for each mesh block with the APS mesh block estimates. Mean absolute error (MAE) was calculated from the mean of the posterior prediction from the Bayesian models and from the predictions of the Random Forest models and is shown in Table 8.5 and compared with the results of the Agricultural Production Census information at the mesh block level

Method 4: FOL Bayesian ZIP model predicted values and Random Forest predictions compared with APS data at the farm-level

The fourth analysis compared FOL property predictions of the Bayesian ZIP Random Forest models with the farm-level APS data from the 2012 census. The APS data were matched to FOL using the address field. Both the FOL and APS datasets are classified and may not be distributed or disclosed without consent. A confidential report on the matching procedure can be found in Appendix 4.3. The models built from AgriBase data for LSU and cattle were used to predict cattle and LSU for each FOL property (or a posterior distribution of predictions for the Bayesian ZIP models). The error was calculated for each farm. This was done by subtracting the observed (APS data) from predicted (from the random forest models or the mean of the posterior from the Bayesian ZIP models). Mean absolute errors could then be calculated. RPS for each farm was calculated using the APS data as the observed data for each of the 12928 farms that were matched by address (see Appendix 4.3 for details of the specific methodology employed). RPS was stratified by region and examined for evidence of a regional effect using boxplots. Farm type from FOL was recorded for each matched farm as a way to check the validity of this data. Some FOL properties (33) had links to multiple APS addresses and therefore the cattle or LSU counts for these properties were aggregated.

Method 5: FOL Bayesian ZIP model predicted values and random forest predictions compared with veterinary survey data at the farm-level

The fifth analysis compared Bayesian ZIP model predictions and Random Forest model predictions made to FOL properties with data collected at the farm-level by a survey of veterinary practitioners in peri-urban areas. Methodology of this survey and association of FOL numbers with veterinarian local knowledge are documented in Appendix 4.4. The same model that was used in method 4 was used for these predictions. The error was calculated for each farm. This was done by subtracting the observed (survey data) from the predicted (predicted from the random forest models or the mean of the posterior from the Bayesian ZIP models), to obtain the mean absolute error. RPS for each farm was calculated for the Bayes models using the survey data as the observed data for each of the 334 unique FOL farms that were identified by veterinarians (see Appendix 4.4 for details of the specific methodology employed). When a single FOL identifier was indicated by one veterinarian as being used by multiple farming operations, LSU and cattle counts on these properties were aggregated (n = 12).

Results

Data sources used – descriptive:

Table 8.1: *Descriptive statistics of counts of animals for each mesh block stratified by species in the Agricultural Production Survey (APS)*

	Sheep	Dairy	Beef	Pigs	Deer
Minimum	0	0	0	0	0
1st quartile	0	0	0	0	0
Median	39	16	49	0	0
Mean	2727	562	325	27	92
3rd quartile	1838	649	308	0	0
Maximum	100159	21323	16909	347701	17450

Table 8.2: *Count of mesh blocks per region where “islands” signifies those mesh blocks not included in the 16 New Zealand regions which make up the North and South Islands. Examples would include Stewart Island and the Chatham Islands.*

Region	Count of mesh blocks
IslandsIslands	11
Auckland Region	913
Bay of Plenty Region	779
Canterbury Region	1459
Gisborne Region	243
Hawke's Bay Region	570
Manawatu-Wanganui Region	1174
Marlborough Region	236
Nelson Region	38
Northland Region	1025
Otago Region	795
Southland Region	725
Taranaki Region	646
Tasman Region	314
Waikato Region	1911
Wellington Region	417
West Coast Region	207

Table 8.3: *Counts of species present in each region according to the APS data extrapolated from the 2015 survey.*

Region	Count of sheep	Count of dairy	Count of beef	Count of deer	Count of pigs
Auckland Region	205270	117281	117458	12647	9613
Bay of Plenty Region	322882	312856	93357	42156	4542
Canterbury Region	5347862	1200293	470731	291783	168813
Gisborne Region	1547294	17095	267599	17323	854
Hawke's Bay Region	3262468	93047	471010	69977	8116
Manawatu-Wanganui Region	5612743	475466	579565	74033	29039
Marlborough Region	547180	33218	60127	7539	3545
Nelson Region	6466	1386	1581	28	4

Northland Region	440955	397764	380719	5158	2917
Otago Region	5342846	336278	290398	153224	11041
Southland Region	4356427	670581	172150	238402	3036
Taranaki Region	434402	604383	103546	3756	19931
Tasman Region	276877	71956	40149	15540	500
Waikato Region	1776544	1831531	505869	81019	37290
Wellington Region	1664892	108174	140420	13933	14039
West Coast Region	58085	173651	29002	34176	421
Missing region	59522	722	10731	0	2
Total	31262715	6445682	3734412	1060694	313703

All AgriBase properties were examined according to their intersection with the Statistics New Zealand mesh blocks. Five AgriBase properties fell outside of the Statistics New Zealand mesh blocks. The distribution is a little right skewed with most mesh blocks having less than 10 AgriBase properties but a few having large numbers (up to 70) and some where a single farm occupies a mesh block. Despite this right skew, table 8.4 shows that the mean and median are not very different to each other. Eleven thousand three hundred and four mesh blocks are considered to have livestock by Statistics New Zealand and were the sample population in this analysis.

Table 8.4: *Descriptive measures for numbers of AgriBase farms overlapping Statistics New Zealand mesh blocks.*

Numbers of AgriBase properties in mesh blocks	
Mean	7.3
Standard Error	0.07
Median	5
Mode	1
Standard Deviation	7.1
Range	69
Minimum	1
Maximum	70
Sum	82437

Results analysis 1

The agreement between the APS data and the aggregated farm-level mesh block data by the chosen method gives an MAE of 686.9 LSU, disaggregated into 275.2 beef cattle, 392.7 dairy cattle, 117.7 deer, 32.7 pigs and 2112 sheep at the mesh block level. These results indicate the level of agreement between the data used to build the models and the data used to compare with model predictions. This agreement can be compared with the results of both the random forest models and the Bayesian ZIP models in Table 8.5.

Table 8.5: Mean absolute error calculated for the difference between 1) 2 observed data sources namely Statistics New Zealand mesh block level agricultural production survey (APS-mb) data and AgriBase data aggregated to the mesh block level for GST registered properties; 2) Aggregated results of the Random Forest (RF) model using AgriBase predictor variables compared with APS-mb for the cattle model and the livestock units (LSU) model; 3) Aggregated results of the Bayesian Zero Inflated Poisson (B-ZIP) model using AgriBase predictor variables compared with APS-mb data for the cattle model and the LSU model; 4) Aggregated results of the RF model using FarmsOnLine (FOL) predictor variables compared with APS-mb for the cattle model and LSU model; 5) Aggregated results of the B-ZIP model using FOL predictor variables compared with APS-mb for the cattle model and LSU model.

Comparison	Mean Absolute Error LSUs	Mean Absolute Error cattle
1) APS-mb and AgriBase	686.9	392.7(dairy cattle) 275.2(beef cattle)
2) APS-mb and RF (AgriBase)	690.9	567.25
3) APS-mb and B-ZIP (AgriBase)	742.1	615.4
4) APS-mb and RF (FOL)	1076.95	903.95
5) APS-mb and B-ZIP (FOL)	3036.2	2367.4

Results analysis 2

The agreement between the APS data at the mesh block level and the modelled LSU and cattle counts from the Bayesian ZIP models was found to be similar to the mean RPS values obtained for evaluation of the predictions of the Bayesian ZIP models at the farm-level. These predictions were performed from models built with AgriBase data and RPS was used to measure prediction. Mean absolute errors from the model predictions can be seen in Table 8.5 and RPS values in Table 8.6.

Table 8.6: Mean rank probability scores (RPS) for a Bayesian ZIP model built with AgriBase data and tested against mesh block level Agricultural Production Survey data and farm-level AgriBase data.

	Mean RPS at mesh block level (APS mb data as observed)	Mean RPS at farm-level (AgriBase farm data as observed)
Cattle count model	0.246	0.221
LSU model	0.251	0.265

Results analysis 3

When predictions made to the FOL dataset were aggregated to the mesh block level, the RPS value calculated for the LSU model was 0.4799 and 0.4347 for the cattle model. MAE results are displayed in Table 8.5 for both the RF and the Bayesian ZIP models.

Results analysis 4

Thirty one percent of Agricultural Production Census (APC - 2012) records could be matched to FOL address records using the methodology described in Appendix 4.3. This left 12928 properties against which we could validate our model at the farm-level. The spatial distribution of these properties can be examined in Figure 8.44. Validation farms were available from each of 73 territorial local authorities in New Zealand. Farm types from FOL are reported for all farms matched in the APCC data and are shown in Table 8.7.

Table 8.7: *Counts of recorded farm types for each of the FarmsOnLine polygons matched by address to the Agricultural Production Census (2012).*

Farm type	Count
Pastoral	4644
Lifestyle	4006
Dairy	3075
Arable	205
S-Deer	202
Horticulture	133
Other	84
H-Vines	78
S-Horses	74
H-Kiwifruit	51
H-Market Garden	41
H-Glasshouse	29
Residential	26
H-Pipfruit	24
Forestry Exotic	23
S-Poultry	20
H-Stonefruit	17
S-Pigs	17
H-Flower	14
S-Other	9
H-Citrus	8
H-Berry	6
Forestry-Indigenous	5
Sporting	5
Industrial	2
Commercial	1
Elderly	1
Forestry-Vacant	1
New Block	1
Religious	1
Reserve	1

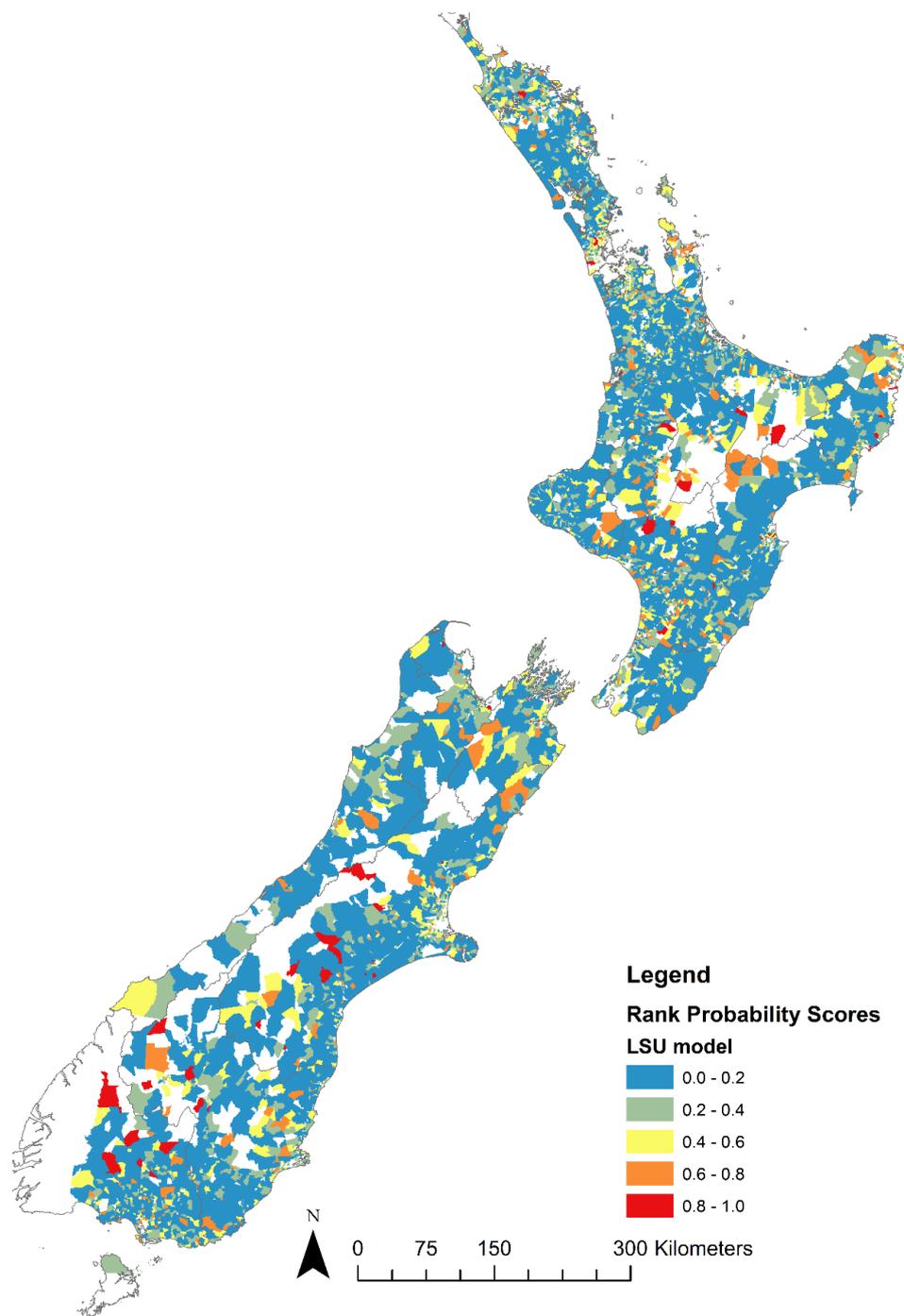


Figure 8.22: Rank probability score (RPS) at the mesh block level. Livestock unit model predictions at the mesh block level are compared with Statistics New Zealand Agricultural Production Survey data.

White areas are those meshblocks where Statistics New Zealand did not hold any Animal Production Survey information in 2015.

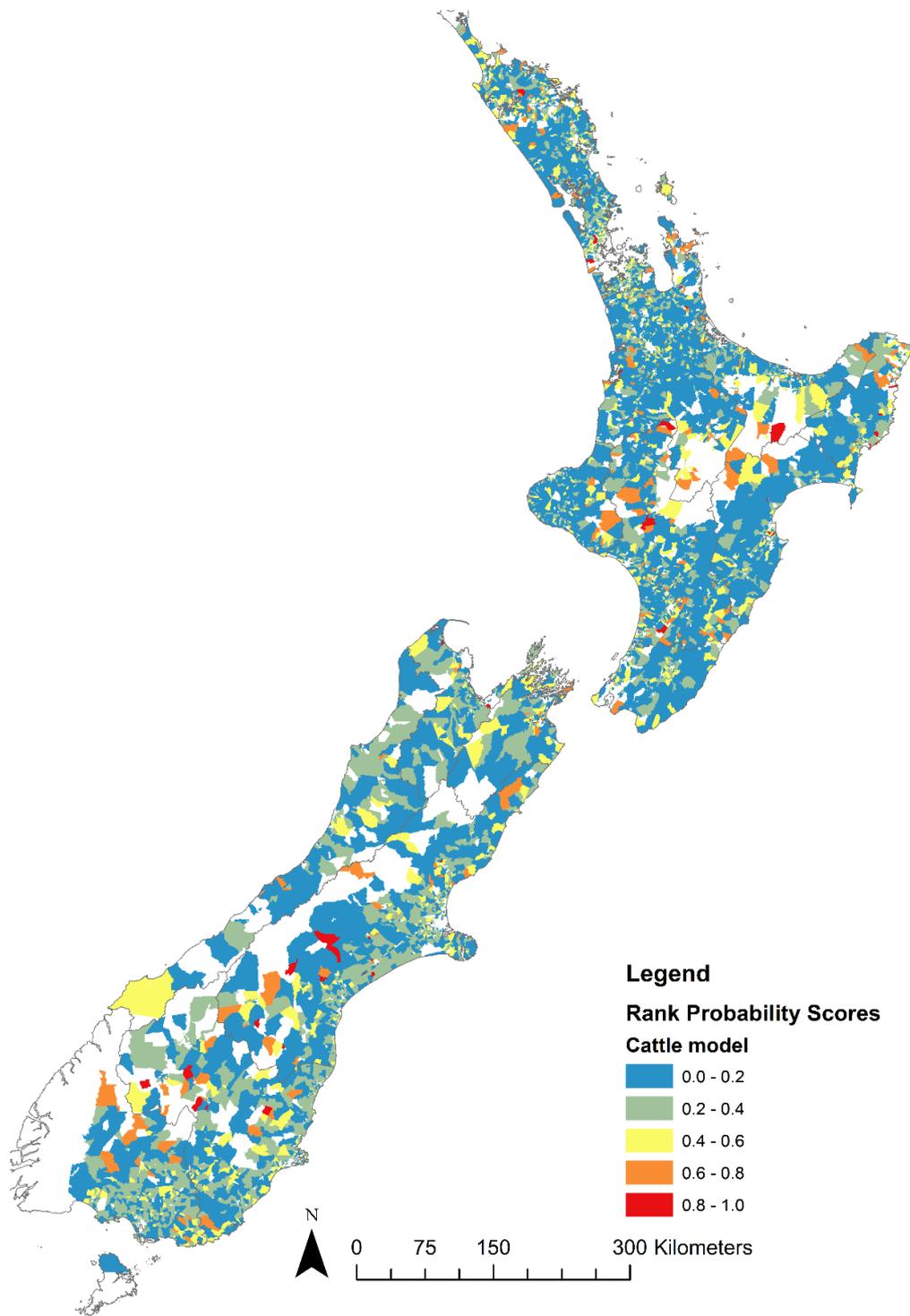


Figure 8.33: Rank probability score (RPS) at the mesh block level. Cattle model predictions at the mesh block level are compared with Statistics New Zealand Agricultural Production Survey data. White areas are those meshblocks where Statistics New Zealand did not hold any Animal Production Survey information in 2015.

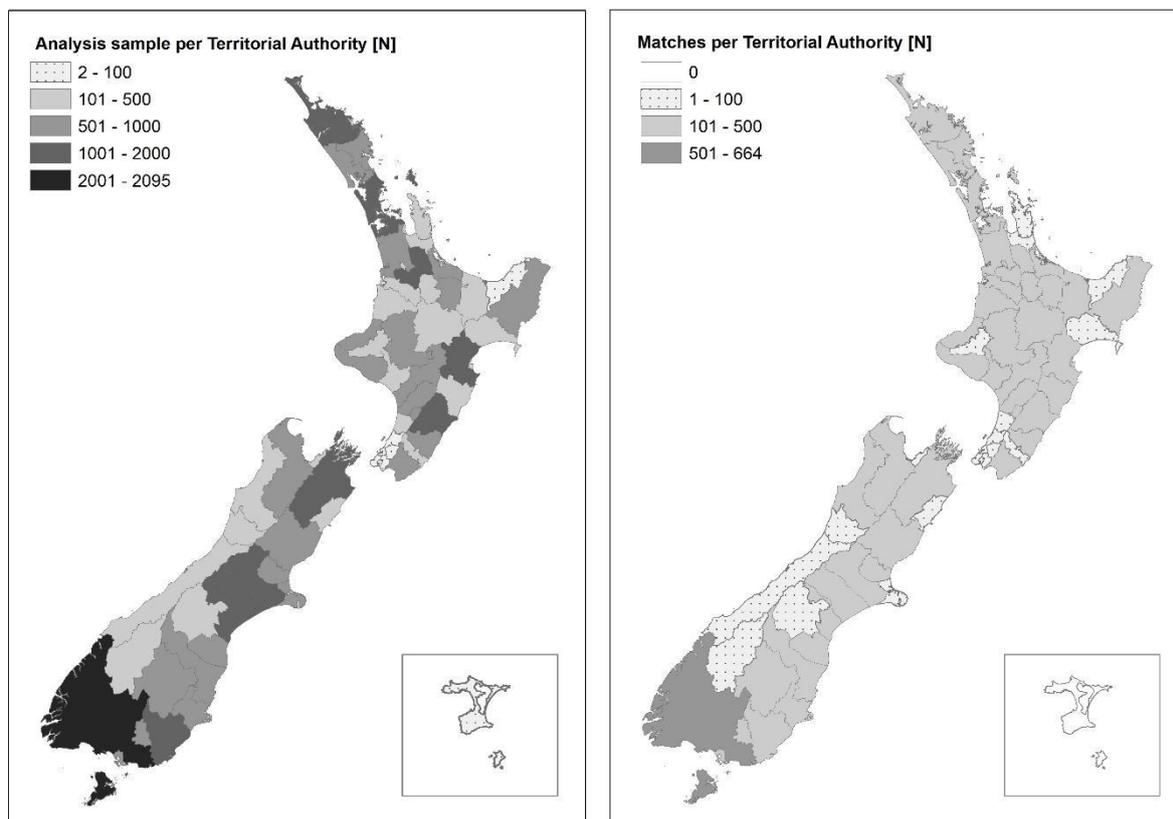


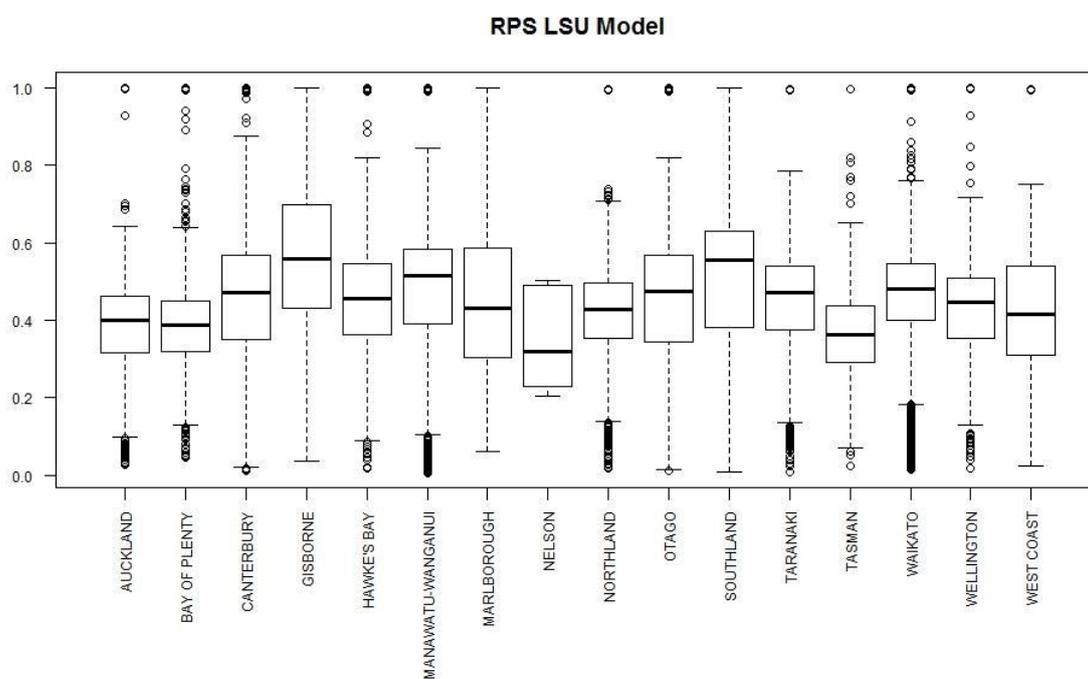
Figure 8.44: Regional distribution of Agricultural Production Census (APCCensusC) analysis sample and achieved matches per territorial authorities as per 20122. 492 entries of the APCC analysis sample ($N = 41182$) could not be unambiguously attributed to an area unit due to incorrect/insufficient location information in the APS dataset and are consequently not represented in the figures. Territorial authority boundaries were sourced from Statistics New Zealand, licensed under CC-BY.

Mean absolute error amongst all farms with APS validation data for the LSU model and the cattle model were calculated both for the Random Forest models and for the Bayesian ZIP models. The RPS

value for the posterior prediction for the Bayesian ZIP models were calculated. These are displayed in Table 8.8.

Table 8.8: Mean absolute errors (MAE) for the Random Forest and Bayesian ZIP models as well rank probability scores for the Bayesian ZIP models for the livestock unit and cattle prediction models when compared against APS data collected at the farm-level.

	Random Forest MAE	Bayesian ZIP MAE	Bayesian ZIP RPS
LSU model	164.2	375.3	0.45
Cattle model	149.3	303.9	0.38



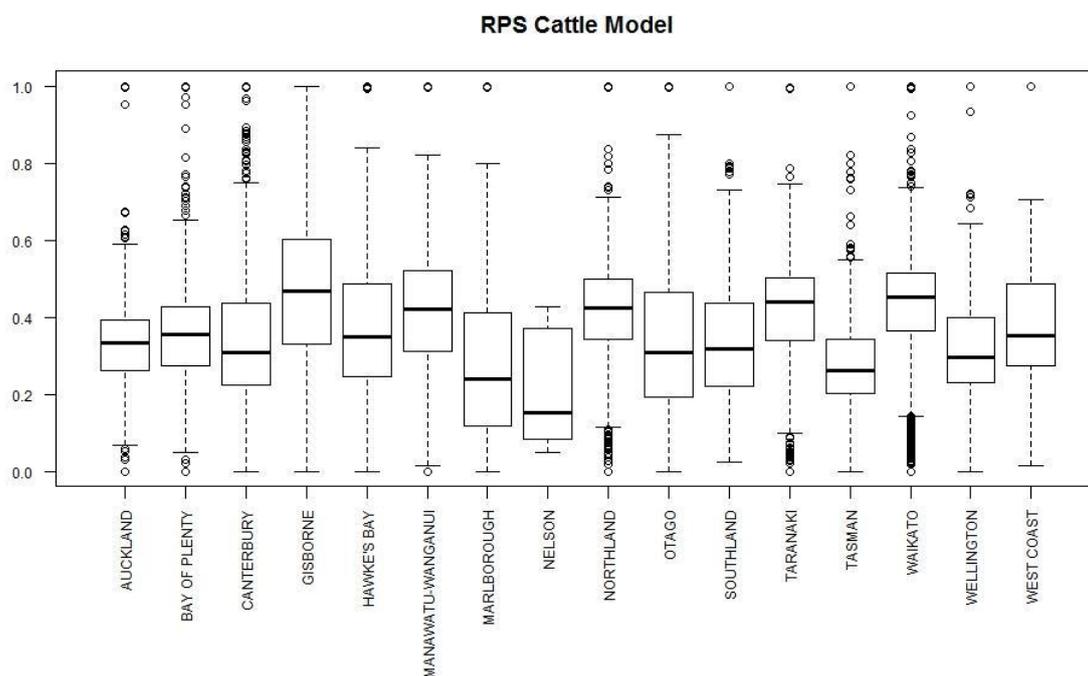


Figure 8.55: Mean rank probability score (RPS) for each farm in the APS validation data stratified by region. The Y axis represents RPS which is measured between 0 and 1 as a measure of agreement between an observed and expected value. In this case, observed values are sampled from Agricultural Production Census (2012) information and the expected are predictions from a Bayesian zero inflated poisson model built to predict counts of livestock units and cattle in New Zealand.

Results analysis 5

Mean absolute error amongst all farms with veterinary survey validation data for the LSU model and the cattle model were calculated both for the Random Forest models and for the Bayesian ZIP models. The RPS value for the posterior prediction for the Bayesian ZIP models were calculated. These are displayed in Table 8.9.

Table 8.9: Mean absolute errors (MAE) for the Random Forest and Bayesian ZIP models as well rank probability scores for the Bayesian ZIP models for the livestock unit and cattle prediction models when compared against data collected at the farm-level by surveying veterinarians regarding their knowledge of animal populations in their practice areas.

	Random Forest MAE	Bayesian ZIP MAE	Bayesian ZIP RPS
LSU model	144.8	367.2	0.444
Cattle model	145.2	344.6	0.415

Discussion

Data sources used

Considerable discussion has already been documented in this report around the limitations and overlaps of FOL and AgriBase. The APS data have their own limitations which include differences in address formatting between the APS data and the databases being matched to. Despite instructions to the contrary, survey and census respondents do not always use their enterprise address for receipt of their GST associated mail and in some cases fail to document their enterprise address on the census documentation. This results in concentrations of addresses for supposed animal production premises in cities when the property with the animals is in fact situated rurally. We believe that this particular bias will not affect our data used for validation at the farm-level as properties would have to have their physical address incorrect both in the APS data and in FOL for this to affect our outcome (to allow the inclusion in the sample comparison dataset). This non-differential bias remains when the data are used at the mesh block level. It is important to note that the Agricultural Production Census is completed every 5 years in New Zealand and a subset of GST registered farmers are surveyed every year in the Agricultural Production Survey. The Agricultural Production Survey is designed to be representative at the area unit level. Each area unit is an aggregation of mesh blocks. For this reason we chose to use Agricultural Production Census information (the census will be repeated in 2017) from 2012, rather than the more recent survey information. We acknowledge the presence of bias in our survey of veterinary practitioners. Our selection is likely biased towards larger, commercial operations and represents a convenience sample, however veterinarians were asked if they had lifestyle clients before being surveyed and many were able to indicate which holdings they knew of were lifestyle rather than commercial holdings.

Agreement analyses

Method 1 and 2 when considered together, give some idea about the relationship at an aggregated level between the AgriBase dataset and the APS data. There are likely to be biases in the way that AgriBase properties have been selected for inclusion in the comparison between APS and AgriBase, however, as these have been continued through the analysis and the same method was used to select the modelled posteriors and RF predictions to include, the results should be non-differential. AgriBase and the APS data seem to be in fair agreement when the RPS and MAE are used as measures to compare them. The ecological fallacy should be considered at this time when the good agreement between the results of the model trained on AgriBase data are aggregated to the mesh block APS data level. Data in each mesh block are collected by Statistics New Zealand and counts are aggregated based on the definition of a mesh block (see materials and methods). Accurate prediction results at an aggregated level do not guarantee accuracy or agreement at the individual property level. If individual predicted premises estimates differ with respect to livestock type (e.g. dairy versus beef) and if the composition of a region differs with respect to these premises types, then the process of aggregation will result in a differential error. However the definition of the ecological fallacy describes a situation in which risk factors identified at an aggregated level is inferred to have a causal role at a lower, disaggregated level.

Method 3 aggregated all GST registered FOL properties to the mesh block level. Model predictions for these properties were subsequently aggregated to calculate a posterior prediction for each mesh block based on the characteristics of properties in the mesh block (as was performed for method 2, but using the model trained on AgriBase data to predict to FOL properties). The results of this analysis were very poor as the quality of the farm type data in FOL made it impossible to select those properties that would be registered for GST purposes with any level of accuracy. Appendix 4.2 shows the table of farm counts in each of the farm classes considered for inclusion. Cross checking of the approximately 12 thousand APS properties was performed and their farm type in FOL recorded (Table 8.7). As the APS data is expected to contain only GST registered properties, the indication by FOL that almost a third (4006 out of 12928 matched) are lifestyle properties is most discouraging. Any database can only reasonably record the primary use of a land area however the land use information in FOL are misleading in the subset of records examined.

When the models were executed on the FOL dataset and the farm-level outcomes compared with APS data the results were poorer than was expected based on the fair performance of the demographic models on the AgriBase farm-level data and the APS mesh block levels. Thirty three APS addresses were matched as many to one with FOL addresses. When this occurred, APS cattle or LSU counts were aggregated. This is reasonable as it is possible that multiple GST registered businesses may use one FOL property for their activities. As our model aim is to predict counts of animals in a geographic area, this solution is reasonable for this small number of occurrences. A possible reason for this poor performance is that actual farm boundaries may not be being correctly represented by FOL polygons. Previous analysis showed that FOL assigns multiple FOL numbers to what we believe are single farms based on AgriBase and the National Animal Identification and Tracing (NAIT) records. As APS data are matched with single addresses which were then matched to FOL address records, errors in polygon boundaries or the assignment of multiple FOL numbers to what is in fact one farm would not be accounted for and could be a contributing factor to the poor model performance.

Validation data gained from local information gathering from veterinarians practicing in the areas of interest was predicted with slightly higher accuracy by the RF models than data from the APS, while the Bayesian ZIP model predicted LSUs from this validation source better than data from the APS, it predicted cattle counts on these properties more poorly when measured by the MAE. It is uncertain if this reflects the inherent biases of data collected in this way (see Appendix 4.4).

When the two model types were compared against both validation data sources, the RF models predicted better when the mean absolute error (MAE) was used as a measure, for every subset of validation data. Over all, the random forest model predictions both for cattle and LSUs were closer to the observed values than those of the Bayesian ZIP models. Random forests (Cutler, 2007) are relatively novel in ecology though they are widespread in other disciplines. These machine learning techniques require greater computing capacity but are generally thought to have great predictive accuracy (Elith et al., 2006; Cutler, 2007; Robinson et al., 2014) than maximum likelihood methods. These techniques can easily fit complex relationships and interactions without the need to explicitly specify them which likely explains their greater performance over traditional techniques (e.g. logistic regression) (Elith et al., 2006; Cutler, 2007; Elith and Graham, 2008). They are also extremely stable with noisy data (Cutler, 2007; Elith et al., 2006).

Agricultural Production Census data include only those properties registered for GST purposes – this implies that any farm in this dataset must generate more than 60 000NZ\$ per annum from farming and that this income must be declared to the New Zealand government. As mentioned, these census data exclude lifestyle holdings which are not commercial concerns and are by biased toward including more larger and more commercial farming operations. There is no information available to us on what level of non-registration for GST exists among the commercial farming population but it is reasonable to assume that some farms will be missed by the Statistics New Zealand data collection method (even in the census years).

AgriBase has multiple mechanisms of being updated many of which provide high value information and a reliable way of gathering farm level information from primary producers. These are listed in table 8.10. The inability to directly link AgriBase and the Agricultural Production Survey/Census is an ongoing source of frustration to those that require farm level data in MPI. While we are not able to truly ground-truth or validate our models based on the data sources at our disposal, this comparison has provided further evidence of the fragmented data landscape in New Zealand and provides the authors with the technical background to make recommendations about future steps that must be taken in New Zealand.

Table 8.10: Sources of data updates to AgriBase, a New Zealand farms database which contains farm identifiers, animal and crop populations and farmer contact details for all farm types and is maintained byASUREQuality, a New Zealand State Owned Enterprise.

User	Purpose	Notes
Dairy Industry	clean stream accord	Industry uses AgriBase to fulfil obligations for the clean streams accord for dairy
Horticulture Industry	WorkSafe requirements	Industry uses AgriBase for hazard mapping under the work safe requirements
Rural Helicopter Rescue Trust	maintaining up to date contact details	Rural helicopter rescue trust (mail out 45000 farm surveys per annum)
MPI	tagging meat samples	MPI meat inspection vets use AgriBase for tagging meat samples, and do their se based on the Animal Status Declaration (ASD) forms. The MPI meat inspection veterinary service provide feedback when meat samples are taken for chemical a antibiotic residue testing and these operational staff provide feedback on the rec they use and report any issues they encounter finding the farm, or locating the These details are then updated from the contact details on the Animal Status Declaration (ASD) form

AsureQuality	QA programmes	On-farm QA programmes in the sheep and beef sector, wool sector and dairy seed certification sector and export vegetable growing, field officers go to sites and check for regulatory purposes.
Environment Canterbury	land consent	Environment Canterbury use the AB farm boundary layer and require all consent applications to include the AB id, other regional councils are expected to follow suit.
Reasearchers	survey sample frame	Researchers who use AgriBase for sampling of farms for surveys are required (in contracts of use of the data) to provide updated details if they emerge in the course of the study.

Conclusions

Aside from the lack of animal count information in FOL, the unreliability of the farm type information poses a challenge for using this information in population level analyses. Significant differences in the formatting of FOL and APS data mean that currently only 31% of APS records can be matched to FOL address records with confidence. The inclusion of a farm-level identifier in the Statistics New Zealand census and survey activities is highly recommended. If modelled animal or LSU counts are to be employed, the use of random forest methodology for predicting point estimates is found to be more accurate in our validation exercise, however no measure of uncertainty around these estimates is available. At this time, use of a species distribution model to predict animal counts to FOL cannot be shown to be suitably accurate at the farm-level. Every effort must be made to improve the quality of animal presence and absence, land use and animal count data in the FOL database and to move toward a single farm identifier across all systems in New Zealand which will allow FOL, AgriBase and the APS data to be utilised to their full potential.

Deliverable 7: Use of model outputs in a disease simulation model (Australian Animal Disease spread model- AADIS)

Mary van Andel¹, Richard Bradhurst², Graeme Garner², Andrew Robinson³

¹ Ministry for Primary Industries, Wellington, New Zealand 6140

² Department of Agriculture and Water Resource, Canberra, ACT, Australia

³ Centre of Excellence for Biosecurity Risk Analysis, School of Biosciences, University of Melbourne, Melbourne VIC Australia 3010

Introduction

New Zealand has multiple sources of animal demographic data, and previous chapters in this report have focussed on the databases and how data may be predicted based on environmental factors. Keeping up-to-date farm demographic information is a costly task for the competent authority. This chapter of the report assesses the effect of the accuracy of farm-level information upon the utility of disease modelling outputs. Experiments in this chapter are limited to the effects of farm-level population accuracy on disease spread simulation modelling. Errors in farm-level information will have additional far reaching and costly effects during disease response.

The New Zealand Ministry for Primary Industries (MPI) has used the disease spread simulation platform InterSpread Plus (ISP) (Stevenson et al., 2013) for modelling the expected spread of Foot and Mouth Disease (FMD) and equine influenza (EI) in New Zealand (Cogger, 2011; Sanson et al., 2011). MPI has developed a set of parameters for representing FMD in ISP that is known as the New Zealand Standard Model (NZSM) (Owen et al., 2011). This parameterisation has been compared with other national level parameterisations used by the competent authorities of Australia, Canada, the United Kingdom and the United States in their preferred model platforms (Roche et al., 2015).

The focus of CEBRA project 1402C was on farm-level population numbers, their uses, availability, predictability and impact. To design an experiment that provides information on the impact of farm-level animal counts on disease modelling, a disease spread simulator that models disease spread within a herd or farm was required. ISP is a farm-level spatial simulation model, however it does not explicitly account for within-farm spread of disease. The model used in these experiments is the Australian Animal Disease spread model (AADIS) (Bradhurst et al., 2015) developed in 2015 for use

by the Australian Federal Government for disease response preparedness to FMD. AADIS is an important tool for disease spread modeling and is expected to be used in more geographic areas and for more diseases in the near future. As such, the opportunity to use AADIS in New Zealand was welcomed both as it supplied a way to ask questions about the effect of herd size on disease spread and as it presented an opportunity for collaboration and building modelling capability. AADIS allows multiple herds to be present on one farm. A *herd* is the unit of interest and disease spread is simulated within each herd, with disease spread between herds based on a probability that is high for herds co-located on a single farm (between herd, within farm spread) and decreases based on distance from an infected herd for herds not co-located (local spread between farms).

Previous work in New Zealand on the sensitivity of the NZSM in ISP to the accuracy of the count and spatial distribution of farms has been performed (Mackereth, 1998). Here the farm population was thinned to examine what percentage of farms were required to be present for accurate disease spread modelling in ISP (Mackereth, 1998). Repeated simulations on thinned datasets identified eighty percent of the true farm population as a cut point at which further omissions of properties negatively affected the quality of the information gained from the modelling exercise.

Developments to ISP are ongoing and the addition of the ability to model within herd spread is currently being worked on. While this development provides new modelling opportunities, the fragmented nature of the New Zealand farm level population information should raise questions about the impact of this information on a within herd spread model. Very little attention is paid to the sensitivity of disease spread simulation models to starting populations in the academic literature and it appears that a disaster on the scale of the 2001 UK FMD outbreak is required for this topic to be discussed.

The aim of this part of the report is to set up AADIS so that it can be used with existing NZ data and to perform initial sensitivity analyses so that further experiments around sensitivity to herd and farm composition can be undertaken.

Model modifications

In the course of the parameterisation, modifications to AADIS were made so that the model was better able to support disease modelling in the NZ situation. These changes included:

- Increased flexibility in parameterisation so that herd types, farm types and species are completely user-defined (i.e., no assumptions of Australian herd types, farm types and species).
- Increased flexibility in parameterisation so that any number of jurisdictions can be included in the model. Australia is a federation with the States and Territories responsible for managing disease outbreaks within their own areas. Consequently AADIS has been configured to allow for jurisdictional level parameterisation (for implementing control policies and resourcing). This is not relevant for NZ's situation as the whole of NZ would be managed as a single jurisdiction. Changes were made to make parameterisation for a single jurisdiction easier.
- Incorporation of a beta pertpert distribution for herd size where previously only a single estimate was used. This functionality is optional and herd size point estimates can be used if preferred
- Incorporation of seasonal heterogeneity of movement distance for animals moving between sites referred to hereafter as “direct animal movements”

Model parameterisation

The parameterisation of AADIS with NZ data was important for multiple reasons:

1. Enable AADIS to be used to simulate outbreaks in NZ
2. Opportunity to assess the data that are available for animal disease spread modelling
3. Identify which movement frequency data are in need of being updated
4. Assess data that are not currently in use in the NZSM for possible inclusion in the NZSM
5. Familiarise MPI staff with an alternate platform to ISP and the parameterisation of a disease spread model
6. Building relationships between MPI and Department of Agriculture and Water Resources (DAWR)

In point 4 above, 3 categories of data (a: movement profiles; b: herd types; c: spread mechanisms) can be identified and are further discussed below. More information is available in Appendix 6. When changes have not been specified, parameterisation was based on previous work done by

Robert Sanson (RS) and Sharon Roche (SR) in phase 1 of the CEBRA project “Using decision support tools for emergency animal disease preparedness and response: foot and mouth disease”.

a. Movement profiles

In the context of parameterising AADIS, information on animal movements is required. The specific measures required are movement frequency, movement distance and seasonality of movements. The National Animal Identification and Tracing System (NAIT) was started in July 2012 and is supported by legislation. In mid-2015, the scheme became mandatory for all NZ cattle and compliance enforcement activities were initiated. An analysis of all animal movements present in the NAIT database (which only covers cattle and deer at the present time) up until March 2016 was performed to build movement profiles for AADIS. These profiles can be stratified by season and type of property moved to and from. Movement frequency and movement distance can be defined. In particular the New Zealand phenomenon of Gypsy Day is of interest as a disease spread risk. Gypsy Day happens on the 31st of May/1st of June each year and is the period when sharemilking⁴ tenancies come to an end and new ones begin. The period of gypsy day is examined and described as it is reflected in NAIT. Gypsy Day is not currently included in AADIS, but a description of how it is represented in the NAIT data is available in Appendix 6.

Assumptions and biases

To generalise movement profiles from the NAIT data some assumptions must be made – the results of the analysis must be interpreted in the light of these assumptions and judgement made regarding if the NAIT data are complete enough to use in this way. The NAIT data have not been specifically described in this chapter (although they are described in chapter 5/Deliverable 3) so the basis for every assumption is not detailed here. Specifically, we assume that:

1. Registered People in charge of animals (PICAs) are representative of non-registered PICAs (as the number of PICAs not registered in NAIT is not known)

⁴ Sharemilkers are farmers who own their own dairy herd but not the land they farm them on. The land is leased from the land owner and typically these leases expire on May 31/June 1 – then cows are moved to a new property to continue production.

<http://maxa.maf.govt.nz/mafnet/rural-nz/profitability-and-economics/trends/sharemilking-review/httoc.htm>

2. Registered PICAs have up to date and complete movements recorded (as NAIT can be flawed in more than one way, the data can be incomplete when individuals fail to register and when movements are not registered)
3. Where multiple FarmsOnLine (FOL) farms are associated with a single NAIT⁵ number the first FOL number is the list of to and from is used to calculate the distance of the movement (using Euclidean geometry). Assume that this measurement bias is non-differential⁶.
4. The denominator for the frequency calculations includes all registered PICAs for the production type being examined.
5. Omission of the “mixed” production types has not biased the results (they are small numbers of observations – see Figure 9.1).
6. Consider that NAIT compliance is improving over time and that the average of multiple years may not be more reliable than data from the single most up to date year.

Data available

1. Data from July 1st 2012 until Feb 1st 2016 of each movement in NAIT
2. FOL with x and y in NZTM for distance calculations

AADIS tables updated from the original AADIS-NZ tables built from ISP New Zealand Standard Model for FMD (NZSM) data.

1. Movement frequencies on and off of farms of cattle and deer by season
2. Movement distances by season for cattle and deer

Issues encountered

Only cattle and deer movements are recorded in NAIT. AADIS tables for “indirect movements”, which are actually movements of low and medium risk conveyors, cannot be parameterised from

⁵ According to the NAIT Act <http://www.legislation.govt.nz/act/public/2012/0002/latest/whole.html> multiple FOL identifiers may be associated with a single NAIT number should the FOL farms be within 20 km of each other and are used for the same farming purpose. When this situation results a choice must be made about how to measure the distance between 2 groups of FOL farm ids.

⁶ Non-differential bias is a term used in veterinary epidemiology to indicate a bias that affect all data strata equally. While measurement bias is expect the resulting error is not expected to alter the inference as it applies equally to all strata.

these data, and ISP extrapolations from work done for CEBRA by Sharon Roche and Robert Sanson was used. This work is detailed in phase 1 of the CEBRA project “Using decision support tools for emergency animal disease preparedness and response: foot and mouth disease”. Briefly this work compared model outputs from the NZSM with AADIS both in New Zealand and Australia. Direct and indirect movements of non-NAIT species must be populated from existing ISP to AADIS work which is based on diary studies completed for construction of the NZSM in ISP.

NAIT has high compliance at sale yards and slaughterhouses due to the data collection structure and enforcement focusses at these points. A known area of weakness in the data source is the completeness of the farm to farm movements. If the NAIT information is used without adjustment to parameterise the cumulative frequency tables in AADIS that define the percentage of movements that go to sale yards vs farm-to-farm movements for dairy, beef and deer farms, then far too much spread from sale yards is identified. Values from previous ISP to AADIS work were used.

b. Herd Types

When simulating disease transmission, a model may use a farm, or a herd, or an individual animal as the unit of interest. For example, the New Zealand Standard Model (NZSM) for FMD which runs in ISP defines 6 farm classes. AADIS allows multiple herds to be present on one farm. It has been mentioned in the introduction that a *herd* is the unit of interest in AADIS and that herds can be grouped within farms. Animals within herds influence disease spread in the model as they are the population that drives the within herd spread component which is run for each herd as a simple compartment model (SLIR).

AgriBase (Sanson and Pearson, 1997) was used to build the herd file for the NZ parameterisation of AADIS. Farm types are required for input for AADIS, and the ISP farm types used in the NZSM were collapsed to 4 farm types (Table 9.1). The grazing and dry stock and the pastoral livestock types were merged to form a pastoral category and the two pig production types were amalgamated to a pig’s category.

Table 9.1: *Relationship between the farm types used in the New Zealand Standard Model in InterSpread Plus and how these relate to farm types used to parameterise the Australian Animal Disease Spread Model for New Zealand in this experiment.*

ISP – farm types	Description according to AgriBase for NZSM	AADIS – farm types
Dairy	Dairy farms	Dairy
Grazing and dry stock	Grazing of dairy replacement cattle and dry stock	Dairy
Pastoral livestock	Farms with beef cattle, deer and sheep	Pastoral
Lifestyle	Hobby farmers with any type of livestock	Lifestyle
Pig Breeders	Pig operations supplying young animals to other farms to grow up for slaughter	Pigs
Pig Fatteners	Pig operations that raise pigs from weaners to finishers and send for slaughter	Pigs

All farms with no animals susceptible to FMD were removed, leaving 76487 farms with one or more herds (of susceptible species present) on each farm. A total of 115618 herds were identified to be included in the model (see Table 9.2) on 76487 farms each of which could take one of four types (pastoral, dairy, lifestyle or pigs).

Ten herd types were created, a large and small herd type for each of deer, sheep, pigs, dairy and beef cattle. These herd types were assigned based on AgriBase farm types, presence of stock and the size of the herds present in the database.

The following logic was applied:

- For all farms that were identified as *lifestyle* in AgriBase, the herd size was allocated to the herd type of the animals present (dairy, beef, pigs, sheep or deer) as “small”
- Of the remainder if the herd was not on a lifestyle farm but was smaller than 50 sheep, 15 dairy, 15 beef, 15 deer and 30 pigs it was allocated to be “small”
- The remaining herds were given “large” identifiers.

The median of the herd population is 30 animals (mean 414; minimum 1; maximum 115620). A population made up for the most part of smaller herds with relatively fewer large herds having a high influence on the arithmetic mean.

Table 9.2: *Counts of herds present in the herd file constructed for use in the Australian Animal Disease Spread Model showing the number of herds of each type stratified by the farm type they are present on.*

Farm types				
Pastoral	Dairy	Pigs	Lifestyle	Total

	Large sheep	17950				17950
	Small sheep	6772	1526	43	12086	20427
	Large pigs	114	33	126		273
	Small pigs	2044	624		1770	4438
Herd types	Large deer	3236	63			3299
	Small deer	302	22		294	618
	Large dairy	2010	11806	2		13818
	Small dairy	465	24		681	1170
	Large beef	23559	1203	40		24802
	Small beef	7907	2080	22	18814	28823
	Total	64359	17381	233	33645	115618

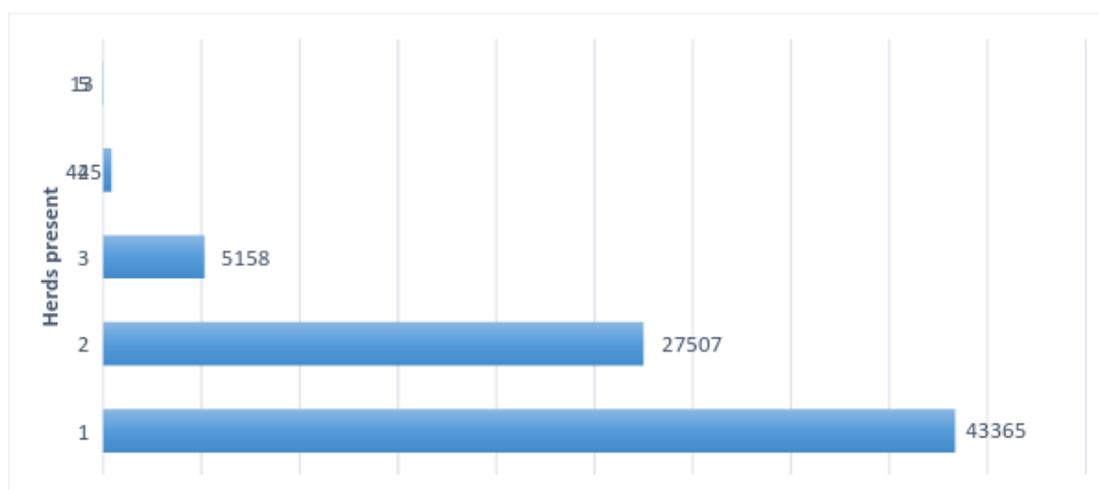


Figure 9.1: Farms in AADIS with multiple herds present on them. Farms had between 1 and 5 herds on them (inclusive) with the majority of farms 43365 of the total 76469 (56.7%) having only 1 herd present on the farm, 35.9% having 2 herds present, 6.7% having 3 herds present, 0.6% with 4 herds and only 0.02% of the population (13 farms) having 5 herds on a single farm.

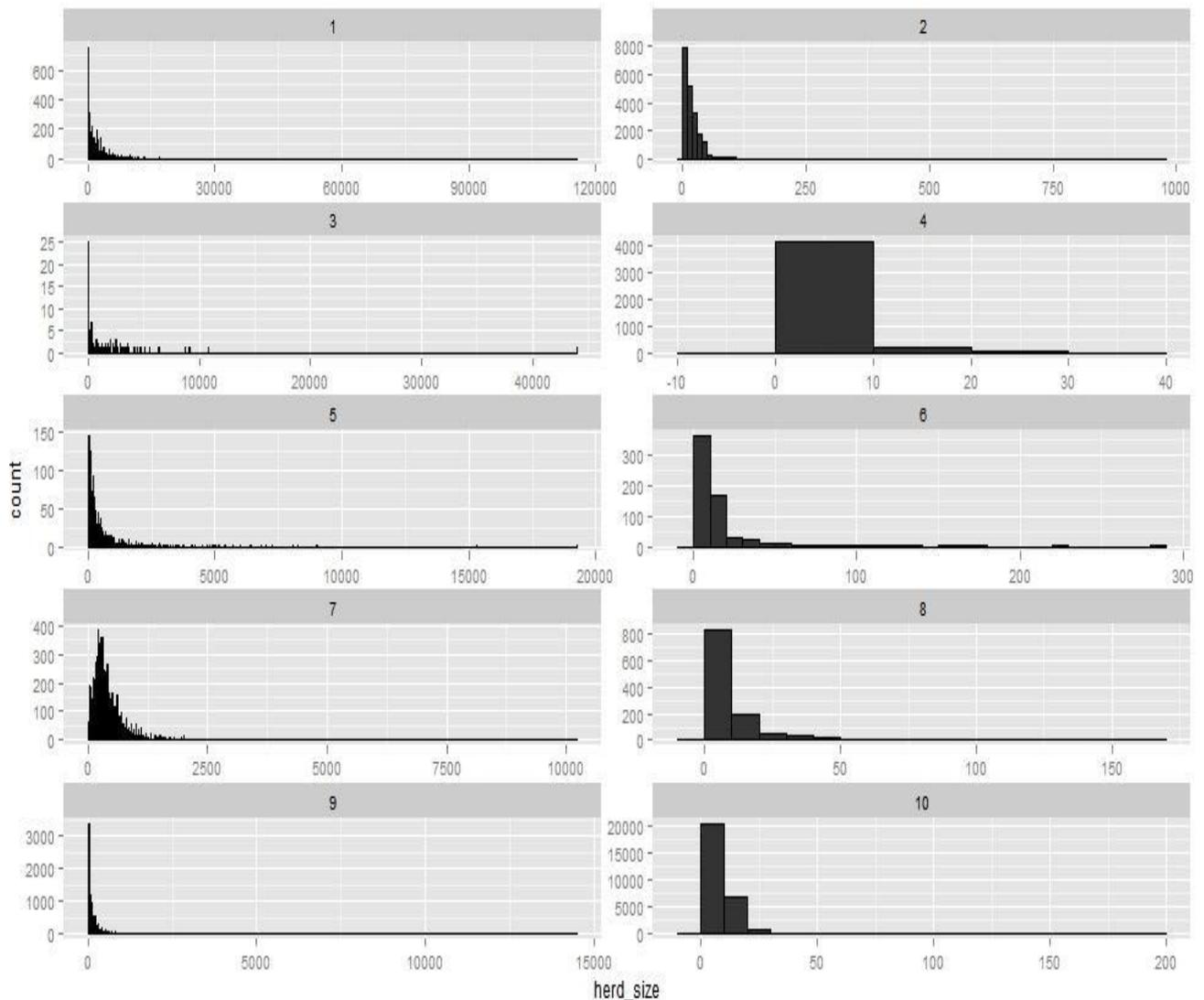


Figure 9.2: Variation in size amongst herds within the 10 herd types. From 1 to 10: Large sheep, small sheep, large pigs, small pigs, large deer, small deer, large dairy, small dairy, large beef, and small beef. Note that the x and y axes scales differ to allow better visualisation of the smaller classes although bin width is set to 10 throughout.

c. Spread classifications

AADIS requires a baseline probability of spread between herds on a single farm, a probability of local spread (between different farms), and an indirect spread probability. These are population-wide

values that affect every herd type to determine whether infection is transmitted when an infected and susceptible herd come into contact. The baseline value is modified to take into account specific attributes of source and contact herds (size, species, prevalence in source herd, etc.). We chose 0.5 for within farm between herd spread, 0.03636 for between farm between herd spread and 0.055 for indirect spread because these values are recommended by the AADIS programmers and designing epidemiologists, and resulted in ratios comparable with previous NZ and Australia comparisons (table 9.3) by the mechanisms identified. The numbers were regarded as reasonable as they were adapted from those used in Australia which are derived from the literature and from field reports and they compare favourably with those generated by the NZSM from ISP for each of the different transmission pathways (Table 9.3). A material difference between the Australian use of AADIS and the NZ parameterisation suggested here is that the current parameterisation of AADIS with NZ data uses a large number of herds, so that local spread between herds on single farms make up a larger percentage of spread than in the ISP model (or in the baseline Australian AADIS model). This is the reason that totals in the current parameterisation rows (rows 1 and 3 of Table 9.3) do not add to 100% (Table 9.3). In ISP a farm is the unit of interest and so spread between herds on a single farm is not taken into account.

Table 9.3: Percentages of new farm infections attributable to different spread pathways

	Seed	Local spread	Direct spread	Indirect spread	Sale yard spread
Current parameters in AADIS	Taranaki	54%	9%	16%	1.6%
ISP previous CEBRA ⁷	Taranaki	50%	12%	35%	22%
Current parameters in AADIS	Canterbury	21%	8%	34%	2%
ISP previous CEBRA	Canterbury	27%	15%	53%	3%

Sensitivity Analysis of AADIS to herd size

Infectivity and susceptibility weights and powers

AADIS incorporates a mechanism that allows different species of FMD susceptible animals to have varying infectivity and susceptibility weightings as identified by Kitching et al. (2006). An important

⁷ CEBRA is the Centre of Excellence for Biosecurity Risk Assessment at Melbourne University.

finding from the UK 2001 outbreak was that herd size had an effect on the infectivity and susceptibility of a particular herd to FMD (Keeling et al 2001). Herd size and infectivity/susceptibility in AADIS is configurable as either a linear or non-linear relationship (Bradhurst, 2015) (p. 61). Tildesley et al (2012) reported that non-linear scaling provided a better fit to the UK 2001 outbreak than linear scaling. In AADIS susceptibility and infectivity weights of individual herds are calculated as:

$$w_h = w_s * h^n$$

where:

w_h = herd infectivity/susceptibility weight

w_s = species relative infectivity/susceptibility

h = herd size

n = infectivity/susceptibility power

The herd infectivity/susceptibility weights are normalised across the population to give relative susceptibility and infectivity for every herd in the population.

The infectivity/susceptibility power (n) is assigned a value between 0 and 1, where 1 specifies a linear relationship between herd size and infectivity/susceptibility, and 0 specifies no relationship. A different value may be assigned for each species and for each of the susceptibility and infectivity weights. As the infectivity/susceptibility powers decrease toward 0, outbreak size increases because the infectiousness and susceptibility of very small herds is increased. As the infectivity/susceptibility powers increase toward 1, outbreak size decreases as the relationship between herd size and infectivity and susceptibility becomes linear (smaller herds become less infectious than larger herds of the same animals). This is particularly of importance in our situation where the herd file contains more smaller herds than larger herds (mean herd size is 30 animals). In the UK 2001 epidemic, values of 0.2 to 0.4 were found to have the best fit to the epidemic data (Keeling et al., 2001; Tildesley and Ryan, 2012; Tildesley et al., 2008). The baseline Australian scenario which has been parameterised in AADIS by Graeme Garner (DAWR) and Sharon Roche (DAWR) uses a value of 0.555 both for infectivity and susceptibility powers as it is assumed that spread would be slower and more associated with farm size in Australia than it was in the UK in 2001. Sensitivity analyses on this parameter have been performed by Australian epidemiologists.

Based on previous work completed on the sensitivity of AADIS to the effect of herd susceptibility and infectivity weights in the AADIS herd file, the starting point for this project was identified in Table 9.4. Table 9.4 is adjusted from the Australian parameterisation of AADIS. Note that the Australian version of AADIS does not include deer and the version of AADIS parameterised for this report for New Zealand use does not include a species⁸ “smallholders”. These values are for the Type O pan Asia strain of FMD virus.

Table 9.4: *Relative infectivity and susceptibility weights for each of the species included in the Australian Animal Disease spread model (AADIS) and used both in the Australian and New Zealand parameterisations.*

	Cattle	Pigs	Sheep	Deer
Relative infectivity weight (I_w)	2	4	1	1.5
Relative susceptibility weight (S_w)	6	0.4	1	2.5

The values in Table 9.4 are used throughout the experiment described. The AADIS model uses the species constants in Table 9.4 along with infectivity/susceptibility powers and the herd size to calculate the infectivity/susceptibility of a single herd (Bradhurst et al., 2015). While the importance of herd size is well established as a risk factor for disease spread in countries where outbreak data have been collected, the inclusion of such an effect in a disease spread model raises questions for those interpreting the outcomes of disease simulation models. Firstly, how accurate the estimation of herd size should be to provide realistic spread approximation in a disease spread model. Second, by looking at the sensitivity of disease models used for predicting optimal control strategies to inaccuracy in herd size. Our experiment is a first step toward answering these questions.

Approach to experiment

The objective of the sensitivity analysis presented here to document the response of the AADIS model in the silent spread period when parameterised with New Zealand data and when using a variety of herd sizes (real, median, beta-pert at herd type level) and compare this with the NZSM built in ISP.

⁸ Australia includes a “smallholders” category in their parameterisation which is an average of other herd type parameters.

As an additional comparison, posterior predictions of herd sizes modelled in chapter 7 (Deliverable 5) at the farm level (not the herd type level) were used to parameterise another version of AADIS under New Zealand conditions and compared with real herd sizes from AgriBase for the same 3 seeds. More details on each of the steps follow:

1. Models using the real, median and beta-pert herd sizes were compared at different infectivity and susceptibility powers⁹. The outputs compared were from the silent spread phase of the outbreak and included a comparison with the same period examined in the New Zealand Standard Model built in Interspread Plus.

- a. Real herd sizes

Herd sizes from the AgriBase dataset were used and 3000 randomly seeded simulations with 14 days of the silent spread period were run with infectivity and susceptibility powers of 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.

- b. Median herd sizes

A herd file was built and for each of the 10 herd types, the median herd size allocated for each herd type. The same simulations were performed (3000 randomly seeded simulations of a 14 day silent spread period with infectivity and susceptibility powers of 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1). This approach was used as a way to approximate the ISP approach of having a probability of transmission based on a farm type (in this case a herd type) that does not take into account herd size.

- c. Beta-pert herd sizes (herd type level)

A herd file was built where a number was picked for herd size from a beta-pert distribution for each of the 10 herd types. These beta-pert distributions were identified by testing distributions in the software ModelRisk by Vose. The same simulations were performed (3000 randomly seeded simulations of a 14 day silent spread period with infectivity and susceptibility powers of 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1).

- d. InterSpread Plus (ISP) New Zealand Standard Model (NZSM) as external check

⁹Although susceptibility and infectivity could be individually influenced for each species, for the purposes of this experiment both the species level infectivity and susceptibility are adjusted together.

As no external information is available to validate the disease modelling parameters for FMD spread in New Zealand results were compared with the NZSM in ISP for the same 14 day period. Infectivity and susceptibility are not scaleable in ISP and the NZSM uses the characteristics of 6 farm types to drive infectivity (but does not use animal counts). 2. Real vs modelled herd sizes

In the earlier chapters of this CEBRA project a farm-level model for animal populations was built. The Bayesian ZIP model generated a posterior distribution of livestock units (LSUs) for each AgriBase farm. Some alterations needed to be made to the population dataset including:

- I. The removal of 4962 farms that do not have modelled posteriors as they were part of overlapping polygons or 0 animals were predicted by the binomial part of the model
- II. A further 372 herds from 295 farms were deleted from the herd file as the Poisson part of the model predicted that no animals were present on these farms (although in the actual data they have animals present)
- III. Only the Poisson part of the model was used
- IV. LSUs made up of each herd type on each farm was calculated from the predictive models built in chapter 5.
- V. LSUs were then split by herd type on each farm and converted back to animal counts for each herd so that species specific parameters could be used.
- VI. A beta pert description was adopted for each herd where the minimum, mode and maximum of the posterior was recorded in the herd file
- VII. Modes of farms with very few animals were recorded as one based on the method used to calculate the mode (Appendix 7)– to address this problem, if the mode was 1 then it was replaced with the mid-point between the min and max values (this condition obtained for 4173 lines of data out of 107067, 3.8% of lines)

The above steps left 107167 herds with posteriors described as beta pert distributions to be compared with the same 107167 herds as an alternate real dataset. These two herd datasets were used to compare the effect of uncertainty around herd size during the silent spread phase when infectivity and susceptibility powers were set at 0.85 (keeping in mind that a setting of 0 denotes no effect of herd size and a setting of one denotes a linear relationship between herd size and infectivity/susceptibility).

Fifty days of silent spread were simulated 10000 times for each of 3 set seeds, repeated for both model types. All three seeds were small pig herds, one was situated in the Waikato, one in the Taranaki and one in the Canterbury Region.

The largest risks of FMD introduction to NZ have been reviewed and published (Pharo, 2002). In accordance with this assessment, the 2002 FAO preparation for FMD contingency plans manual (Geering and Lubroth, 2002) outlines questions to consider when producing an assessment of the threat of entry to a country which include proximity to recent outbreaks, current global FMD distribution trends and subtype specific range expansions, disease history, imports of risk species and products and border biosecurity. The review goes on to mention swill feeding of pigs, illegal movements of animals and animal products and civil unrest. Based on these suggested questions the greatest risk of introduction to New Zealand appears to be the inclusion of FMD infected material being fed to non-commercially kept pigs. For this reason seeding into 3 small pig herds was performed.

Results and discussion

a. Real vs Median vs Beta pert (herd type level) herd sizes:

In the silent spread phase the log of the counts of IPs for each of the the 3 AADIS models areare visually compared (Figure 9.3).

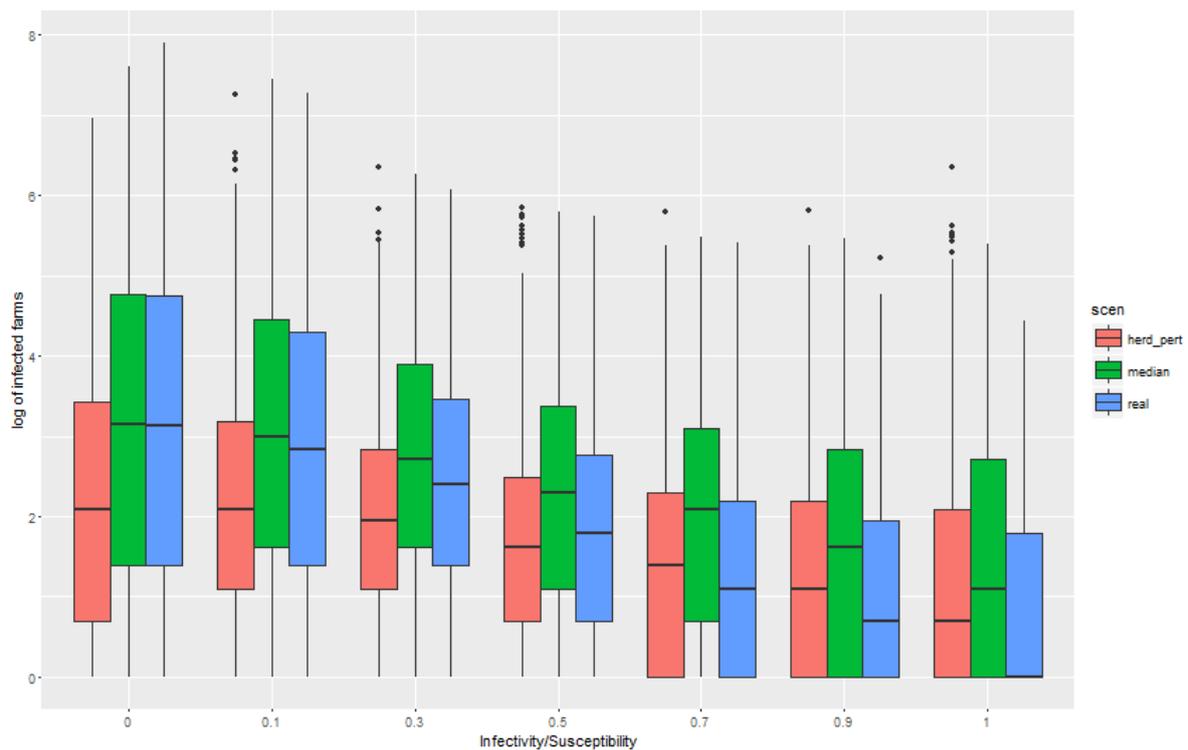


Figure 9.3: Graph of infected farms for a silent spread period of 14 days generated by the Australian Animal Disease Spread Model (AADIS) for 3 different herd datasets. The 3 herd datasets are: blue – the “real” data as recorded in AgriBase, green – a median herd size from each of the 10 herd types described in chapter 9 and in pink- picks for the size of each herd form a herd type level beta-pert distribution generated from AgriBase data. The y axis is on the log scale and the x axis shows different values of the infectivity and susceptibility settings which control the effect on herd size on the infectiousness and susceptibility of a single herd to FMD virus infection.

Table 9.5: Tabulated results of four different disease spread models built in 2 different platforms (the Australian Animal Disease Spread Model – AADIS and InterSpread Plus – ISP). The aim of this table is to show the data from figure 9.3 in tabular form and to provide a reference point where the chosen

value of 0.55 for infectivity and susceptibility can be compared against results from the same silent spread period modelled in Interread Plus using the New Zealand Standard Model. Note that for the “real” AADIS herd dataset, 3 additional datapoints were calculated (0.55, 0.6, 0.65) – these are marked with a *.

Model	herd file	infectivity/susceptibility setting	Min	1st Qu	Media n	Mean	3rd Qu	Max
ISP								
(NZSM)	real	n/a	1	3	5	7.125	9	65
AADIS	real	0	1	4	23	127.6	115	2719
AADIS	real	0.1	1	4	17	76.02	74	1434
AADIS	real	0.3	1	4	11	26.12	32	434
AADIS	real	0.5	1	2	6	13.25	16	313
AADIS	real*	0.55	1	2	5	11.32	13	177
AADIS	real*	0.6	1	1	5	10.18	12	166
AADIS	real*	0.65	1	1	4	9.4	11	209
AADIS	real	0.7	1	1	3	8.366	9	225
AADIS	real	0.9	1	1	2	6.107	7	185
AADIS	real	1	1	1	1	4.947	6	84
AADIS	herd-pert	0	1	2	8	33.96	31	1055
AADIS	herd-pert	0.1	1	3	8	24.39	24	1416
AADIS	herd-pert	0.3	1	3	7	13.9	17	570
AADIS	herd-pert	0.5	1	2	5	10.14	12	349
AADIS	herd-pert	0.7	1	1	4	8.36	10	330
AADIS	herd-pert	0.9	1	1	3	7.837	9	332
AADIS	herd-pert	1	1	1	2	7.605	8	576
AADIS	median	0	1	4	23.5	119.3	118.2	2027
AADIS	median	0.1	1	5	20	83.82	86	1717
AADIS	median	0.3	1	5	15	39.04	49	524
AADIS	median	0.5	1	3	10	22.82	29	329
AADIS	median	0.7	1	2	8	18.47	22	240
AADIS	median	0.9	1	1	5	14.89	17	235
AADIS	median	1	1	1	3	13.44	15	219

b. Real vs modelled herd sizes:

No statistically significant differences could be identified between median numbers of infected farms generated by a model using a real point estimate of herd sizes and a herd file including a distribution for each herd size in the population. T-tests for each scenario had non-significant p values at the

0.05 level in each case (Table 9.6). As expected there were differences between seeds as Canterbury, Waikato and Taranaki have different farm and species densities.

Table 9.6: The results of t-tests for comparing the median number of infected farms between a model using a point estimate of herd size with a model using a beta pert distribution for farm size after 50 days of spread.

Region	AADIS Herd id	Median infected farms for point estimate	Median infected farms for a beta pert distribution	P value for t-test
Waikato set seed	40359	1431	1430	0.9
Taranaki set seed	48004	287	285	0.86
Canterbury set seed	1145	431	454	0.2

Spread pathways were compared within each of the regional scenarios and are displayed visually in figures 10 to 12 below. The point of the comparison is to examine the influence of herd variability on spread pathways.

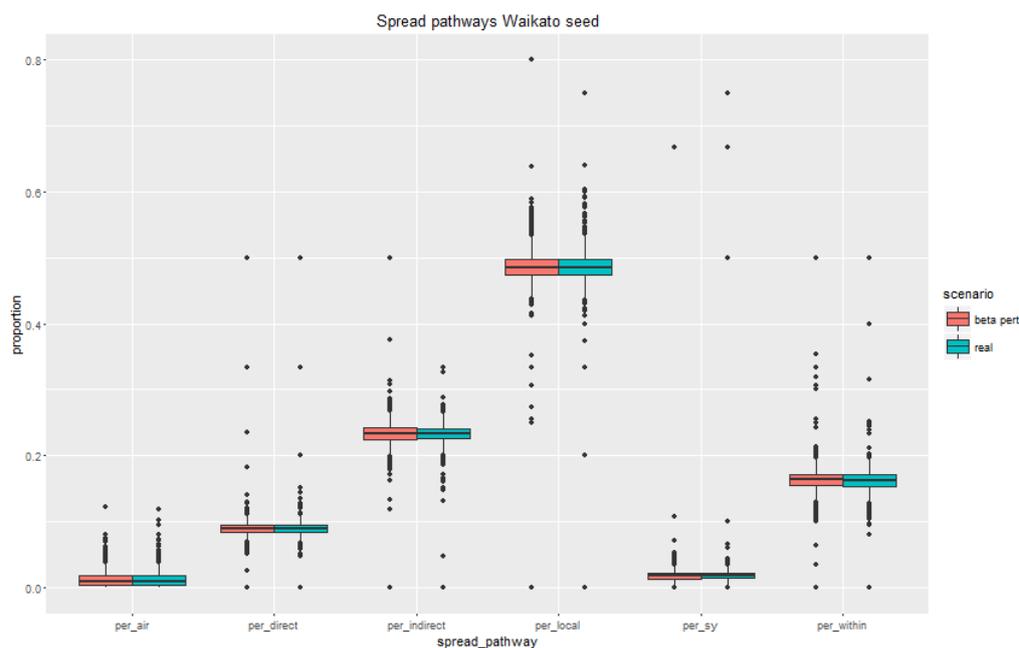


Figure 9.44: Proportions of herd infections by each spread pathway for a model run from a set seed in the Waikato that used a point estimate of farm size (blue) compared with a model that used a beta pert distribution for farm size (pink). From left to right spread pathways are airborne spread, direct spread, indirect spread, local spread between farms, sale yard spread and within farm between

herd spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.

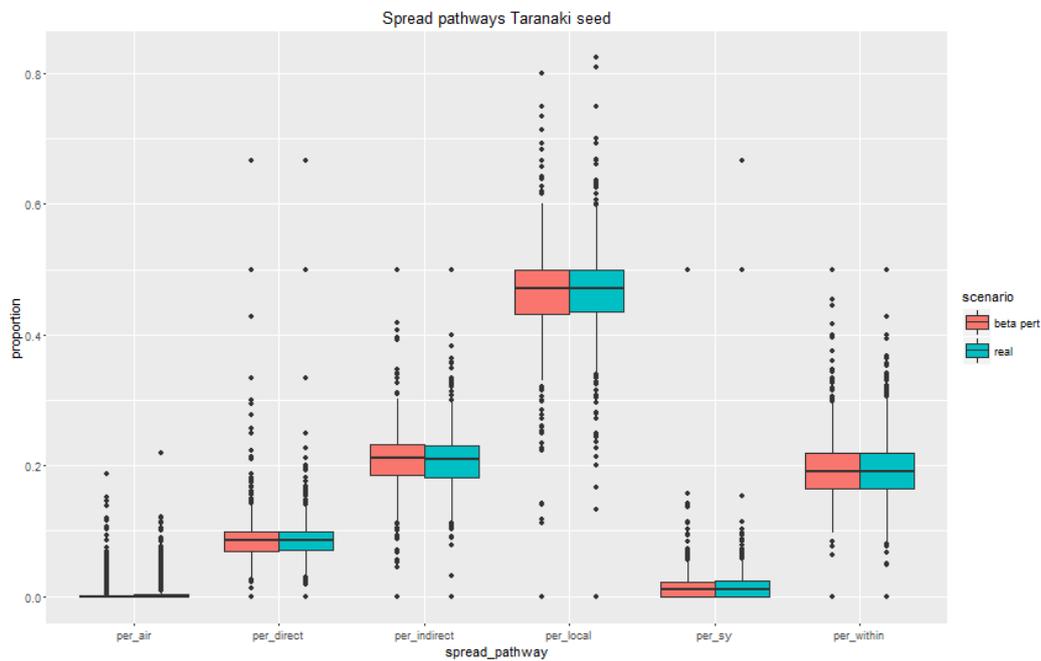


Figure 9.55: Proportions of herd infections by each spread pathway for a model run from a set seed in the Taranaki that used a point estimate of farm size (blue) compared with a model that used a beta pert distribution for farm size (pink). From left to right spread pathways are airborne spread, direct spread, indirect spread, local spread between farms, sale yard spread and within farm between herd spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.

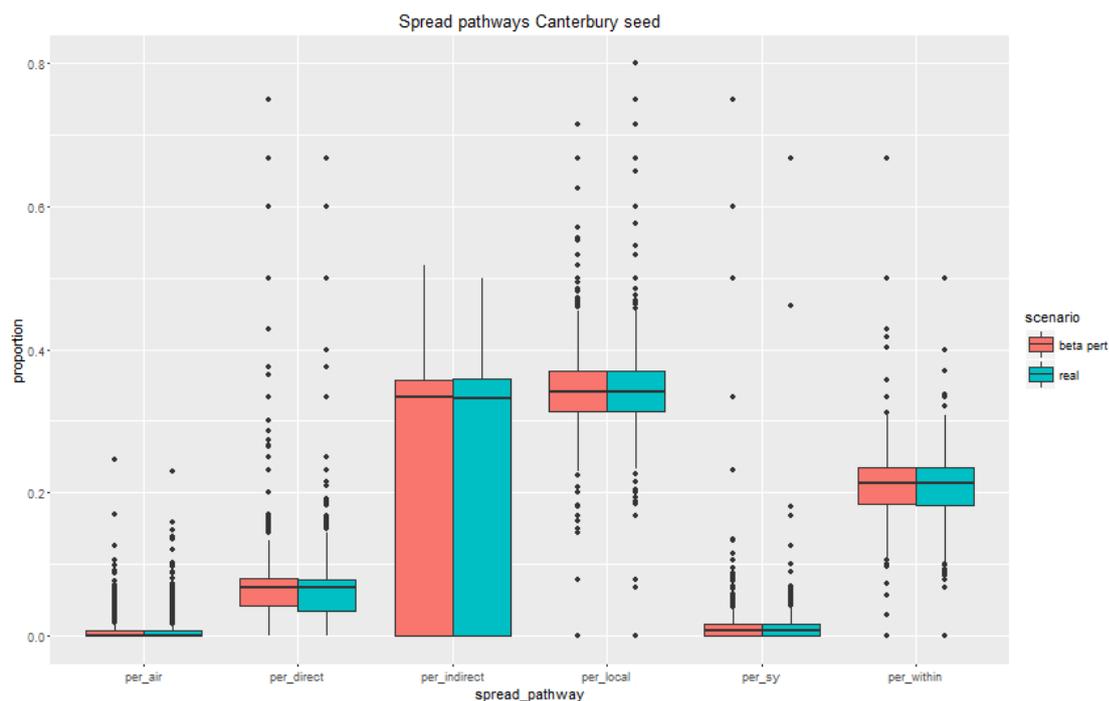


Figure 9.66: Proportions of herd infections by each spread pathway for a model run from a set seed in Canterbury that used a point estimate of farm size (blue) compared with a model that used a beta pert distribution for farm size (pink). From left to right spread pathways are airborne spread, direct spread, indirect spread, local spread between farms, sale yard spread and within farm between herd spread. The y axis represents the proportion that each mechanism makes up toward the outbreak.

For those simulations where the infectivity and susceptibility powers approach zero (i.e. in those cases where the modeller believes that herd size has little effect on infectivity or susceptibility) there is little difference in the predicted outbreak size (no value from accurate herd size data – see figure 3). As the effect of herd size on infectivity and susceptibility increases, the use of a median herd size data set tends to overestimate the size of the outbreak. Those countries that have had outbreaks in recent years have found statistically significant relationships between herd size and disease risk (infectivity and susceptibility) (Hugh-Jones, 1972; Keeling et al., 2001; Lindstrom et al., 2012; Muroga et al., 2013).

In the AADIS platform, infectivity and susceptibility powers have more effect on indirect and local spread than on direct and sale yard spread. This is to be expected as the direct and sale yard spread

pathways don't use infectivity or susceptibility weightings in calculations of the probability of infection, although they do take into account within herd prevalence and consignment sizes

Conclusions

Based on the results of a comparison between a real herd dataset and a median herd dataset, it is clear that the lack of heterogeneity in the herd file predisposes the model to over-predict the size of outbreaks (in our scenario where the medians have been used) and may bias decision makers toward strategies that favour larger outbreaks, for instance vaccination (Sanson et al., 2013). Estimating herd sizes realistically using beta-pert distributions appears to provide as good an estimate of key outbreak metrics as the true data. This finding is potentially useful as it implies that detailed farm-level census-type information is not required to provide plausible disease spread simulations. Herd-level descriptive statistics which reflect variability in herd sizes may prove to be just as useful for disease spread modelling use and are a more realistic view of the true state of the animal population.

We were able to examine how the infectivity and susceptibility setting of 0.55 used by Australian AADIS users compares during the silent spread period with the NZSM and establish that this is a setting that can be used for NZ data in future experiments.

It was not possible to complete investigations into the effect of herd size inaccuracies on the management decisions made, for example the decision to deploy vaccination or not in the event of an outbreak. However the modification to the AADIS model have made it possible for these experiments to be performed in the year ahead as part of Mary van Andel's PhD programme.

Deliverable 8: Comparison of two national livestock databases for informing decision making for a foot and mouth disease outbreak in New Zealand

Masako Wada¹, Tim Carpenter¹, Mary van Andel²

¹*Epicentre, Massey University, Palmerston North, NZ*

²*Ministry for Primary Industries, Wellington, New Zealand 6140*

Introduction

An accurate national livestock database is an important component of a country's biosecurity where livestock industries are important. Although substantial expenses are incurred to collect up-to-date, detailed information from individual livestock premises, the efforts could be offset by benefits from better management and control of infectious livestock diseases. In particular, for an outbreak of highly contagious livestock diseases, such as foot and mouth disease (FMD), highly pathogenic avian influenza (HPAI), equine influenza (EI) and classical swine fever (CSF), presence of such a database listing susceptible premises with their location, along with the records of movements of animals, would provide means of identifying premises at high risk of infection, that is, those that have made high risk contact, or locate in the proximity from the infected premises. Active surveillance on well targeted premises and early detection of disease would efficiently prevent further spread of disease, mitigating potentially catastrophic economic impacts of a disease outbreak. In addition, even in the absence of disease, the database could be actively used to identify particular geographical areas or farm types with a relatively higher risk of disease infection, where the authority should allocate more resources for disease preparedness and livestock producers should be alerted and encouraged to routinely carry out good biosecurity practices, or explore alternative control strategies by disease simulation, which may be useful for enhancing contingency planning.

Currently, New Zealand has two national geographical livestock databases: AgriBase (Sanson and Pearson, 1997) and FarmsOnline (<https://farmsonline.mpi.govt.nz/>). Both databases record the

type, geographical location, animal species present, and contact details of each property. AgriBase was developed in 1993 and is maintained byASUREQuality Limited, a New Zealand state-owned enterprise. FarmsOnLine was developed by New Zealand Ministry for Primary Industries (MPI) starting in 2010, as a non-mandated, web-based system to collect and maintain farm information. The AgriBase data are commercially available to the public, while use of FarmsOnLine is limited to the MPI for specific biosecurity purposes due to privacy and security reasons. AgriBase includes counts of animals by species, while FarmsOnLine records only presence or absence of animals by species. The number of records for premises with FMD susceptible animals in AgriBase is known to be about 20% less than that of FarmsOnLine (Jewell et al., 2016). A recent comparative study showed that the two databases had a limited compatibility, giving different pictures of the livestock geographic landscape (Jewell et al., 2016). With lack of a common identifier, information from the two databases cannot be correctly linked without extensive field surveys or unverified assumptions about the records (Jewell et al., 2016).

For control of an FMD outbreak, one of the major decisions to make is whether or not to use emergency vaccination in addition to the primary measures, known as stamping-out, i.e. immediate depopulation of a premises detected with FMD. Comparison of a range of quantifiable outcomes for stamping-out vs vaccination by use of farm-based spatial disease simulation models is useful for informing policy makers to enhance contingency planning, as well as supporting decision making in the face of an outbreak. A number of studies have identified the conditions where emergency vaccination may be more advantageous than stamping-out only (Hagerman et al., 2012; Boklund et al., 2013; Wada et al., 2016). The common understanding from previous studies is that emergency vaccination is likely to be beneficial in a situation where prompt depopulation of infected premises is hampered by the resource limitation, or/and the increasing number of new cases as the number of infectious premises increases.

In New Zealand, an FMD simulation model termed New Zealand Standard Model (NZSM) has been developed using InterSpread Plus framework (Sanson, 2006). The models consider individual farm characteristics, such as proximity to other livestock properties, animal species and farm type, based on AgriBase livestock database. Therefore, it is anticipated that any implications drawn from the results of NZSM would be influenced by the accuracy of AgriBase. To date, however, no study has examined the extent to which the presence or absence of missing records and accuracy of the records in AgriBase would influence the simulation-based

conclusions.conclusions This study was aimed at examining the sensitivity of NZSM-based policy recommendations regarding emergency vaccination for an FMD outbreak by variation in density of livestock premises and enterprise categorisation, using data from AgriBase and FarmsOnLine.

Methods

Livestock data

AgriBase

The 2011 version of AgriBase database was obtained fromASURE Quality Limited. The AgriBase data were comprised of 81,759 records of unique premises with animals susceptible to FMD. Each record was comprised of a unique premises identifier ('uniqueid'), type of the enterprise with 6 categories ('class'), counts of animals by species, and the easting and northing coordinates of the premises centroid. There were six farm types: 'PLVSTCK,' or pastoral livestock (39,747 premises: 48.6%); 'LIF,' or lifestyle blocks (24,538 premises: 30.0%); 'DAIRY,' or dairy (11,903 premises: 14.6%); 'GRADRY,' or dry stock grazing (5,206 premises: 6.4%); 'PIGF,' or pig fattening (194 premises: 0.2%), and; 'PIGB,' or pig breeding (171 premises: 0.2%). For animal counts, six animal species/production types were distinguished: beef cattle, dairy cattle, deer, sheep, pigs and goats. The mean animal counts per premises by species, farm type, and district (see Descriptive spatial analyses section) are presented in Appendix 8 Table 14.20.

FarmsOnLine

The 2016 version of FarmsOnLine database was obtained from MPI. In total, there were 244,453 records in FarmsOnLine, of which 101,539 premises had animals susceptible to FMD. The following analyses for FarmsOnLine were limited to those records with the presence of FMD susceptible animals. Each record comprised of a unique premises identifier ('FarmID'), a common identifier used in AgriBase ('AgriID'), farm types with 41 categories ('FarmType'), presence or absence of animals by species, and the easting and northing coordinates of the premises centroid. The AgriBase identifiers were only available in 63,568 records (62.6%).

For compatibility with the NZSM parameters, the six-level AgriBase farm type (class) was newly assigned to each of the FarmsOnLine records, with the resulting composition: LIF (54,993 premises: 54.2%), PLVSTCK (26,791 premises: 26.4%), DAIRY (16,868 premises: 16.6%), GRADRY (2,819

premises: 2.8%), PIGF (37 premises: <0.1%), and PIGB (31 premises: <0.1%). The criteria to assign AgriBase farm types were as follows (Figure 10.1). First, premises with the FarmsOnLine farm type: 'pastoral' with absence of dairy cattle (23,572 premises: 23.2%), 'pastoral' with presence of dairy cattle (2,662 premises: 2.6%), 'lifestyle' (51,220 premises: 50.4%), and 'dairy' (16,659 premises: 16.4%), were labelled as PLVSTCK, GRADRY, LIF, and DAIRY, respectively. For the remaining records that had common AgriBase identifiers (4,489 premises, 4.4%), the same AgriBase farm types were assigned. For the records without AgriBase identifiers, those with either 'industrial' or 'commercial,' with presence of beef cattle, sheep or deer (257 premises, 0.3%) and 'S-Deer' (213 premises, 0.2%) were labelled as PLVSTCK. 'Industrial' or 'commercial,' with presence of dairy cattle exclusively (9 premises, <0.1%) were assigned as DAIRY. The rest of the records (2,458 premises, 2.4%) did not have any labels to associate with any of the commercial farm types (i.e. PLVSTCK, GRADRY, DAIRY, PIGB or PIGF), and hence they were labelled as LIF.

The counts of animals per premises were absent in FarmsOnLine. Based on the newly assigned AgriBase farm type, the mean counts of animals by farm type and district, estimated by AgriBase database (Appendix 8 table 14.20), were used for estimation of the economic impacts of an FMD epidemic (section below on Economic Impacts) for FarmsOnLine-based simulation.

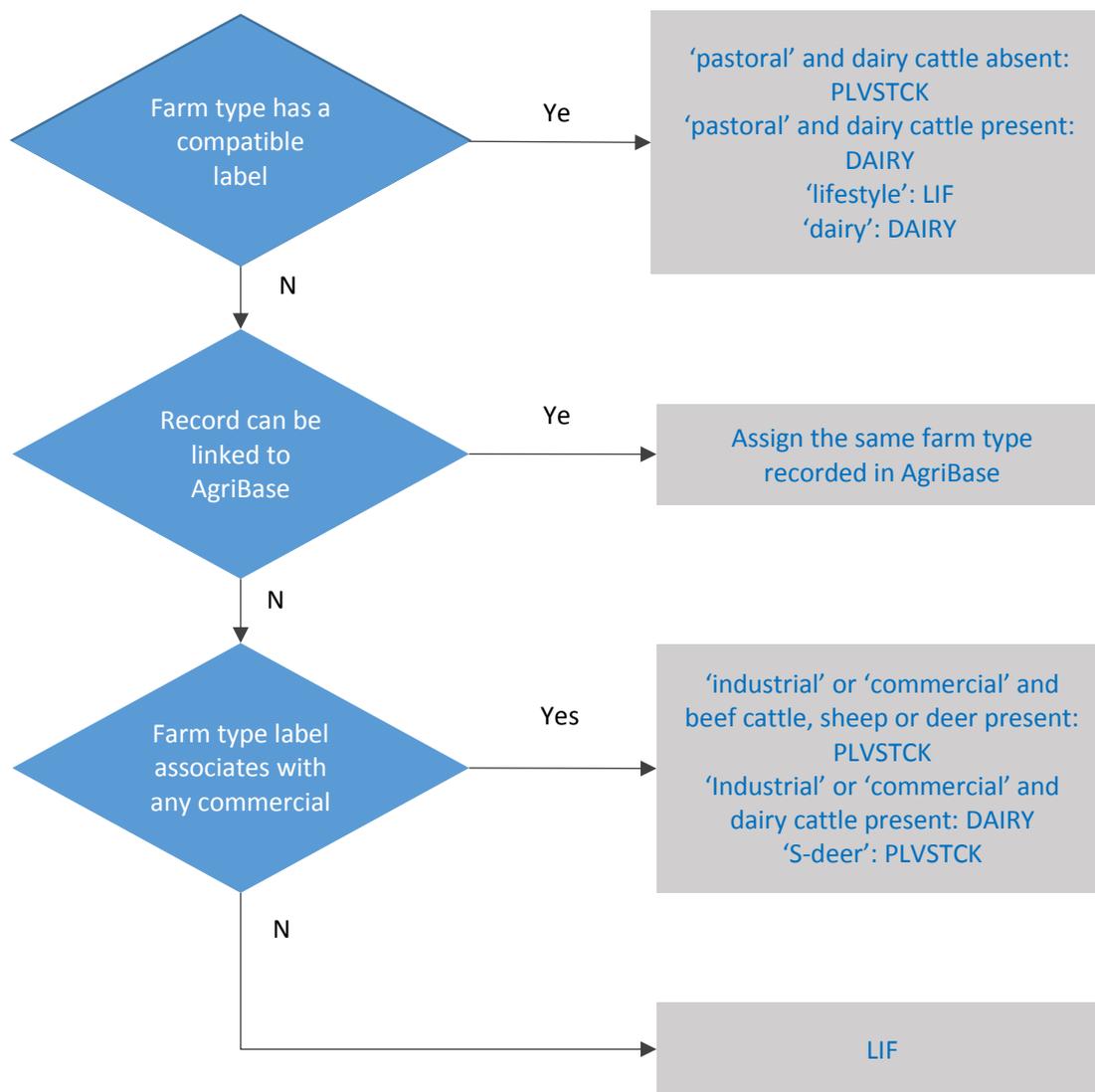


Figure 10.1: Flowchart illustrating the method of assigning AgriBase farm types to FarmsOnLine records, for use of New Zealand Standard Model for foot-and-mouth disease epidemic simulation.

Descriptive spatial analyses

Livestock premises were allocated into one of 69 New Zealand territorial authorities (hereby referred to as district) based on the location of premises centroids. Although Auckland Region has been treated as one unified district since 2009, this study distinguished the previously used district borders (Rodney, North Shore, Waitakere, Auckland City, Manukau, Papakura and Franklin). This is to avoid potential bias in the estimation of clustering of livestock premises, acknowledging its two characteristics: (i) it has a high mean density of livestock premises and (ii) it is an urban area

As an indicator for the risk of FMD local spread, the degree of clustering of susceptible livestock premises was measured for each district, using Ripley's K-function (Ripley, 1977; Ripley, 1988). For each district, the expected number of premises centroids within a distance r from an arbitrary premises centroid was calculated as $\lambda * K(r)$, where λ is the density of livestock premises in the district, and $K(r)$ is Ripley's K-function estimated for the district assuming complete spatial randomness. To correct the edge effects on $K(r)$ estimates, Ripley's isotropic correction was used (Ripley, 1988). $K(r)$ was estimated using 'spatstat' package in R (Baddeley et al., 2015).

The following epidemic simulation considered 54 districts, excluding all 15 cities due to their relatively small area size and small number of livestock premises. The mean values of $\lambda * K(5 \text{ km})$ for AgriBase and FarmsOnLine were used as a criterion to select districts for FMD epidemic simulation. Five kilometres was chosen as a sufficient distance to encompass all surrounding livestock premises directly at risk of local spread from an infected premises. That is, a distance of 3 km beyond which the risk of local spread was considered to be negligible (a probability of <0.001 cases per day), based on the analyses of the FMD outbreak in Cumbria, 2001 (Sanson et al., 2006b), plus a 2 km margin. A total of 7 districts were selected for FMD simulation, based on the rank of the districts.

FMD epidemic simulation

FMD epidemics were simulated by InterSpread Plus version 4.02.17 (Stevenson et al., 2013), using NZSM developed by Sanson et al (2006a). For each unique epidemic setting (i.e. primary case and control strategy), epidemics were simulated for 42 iterations which was considered to be the optimal number of iterations. This figure was determined by assessing the width of the 95% confidence intervals of the estimated median by the bootstrap sample size (Efron and Tibshirani

1993), over which an increase in the number of iterations did not contribute to the precision of the estimate of the median. All the InterSpread Plus parameters are provided in Appendix 8 Table 14.21. This section describes only those settings varied in this study.

Primary case

FMD epidemic simulation was initiated by seeding FMD into a randomly chosen premises to be a primary case in each of the 7 districts of selection. An epidemic was repeatedly simulated from a different random primary case, each with 42 iterations, since the unique characteristics of a primary case (e.g. species, farm types, and surrounding premises at risk) may influence the course of an epidemic, which may mask the effects of interest (e.g. control strategy and district). For each primary case, a set of simulation was run, which comprised of 42 iterations * 3 control strategies (described below) * 2 databases (AgriBase and FarmsOnLine). A list of primary cases to examine was created by randomly selecting 5% of premises for each district recorded in AgriBase, which had a matching identifier with FarmsOnLine.

Control strategies

We examined three strategies: (i) stamping-out only throughout the epidemic (SO), (ii) stamping-out, as well as emergency vaccination implemented on the 2nd week after first detection, targeting only those commercial premises excluding lifestyle blocks and pig farms (AgriBase class: PLVSTCK, DAIRY and GRADRY) within 3 km vaccination zone (SOV_{targeted}), and (iii) same as (ii) except that emergency vaccination targeted all susceptible premises (SOV). The first strategy, SO, comprised of resource-constrained depopulation of detected premises, active surveillance of premises locating within 3 km of detected premises (protection zone) in accordance with MPI's response plan (Anonymous 2011) and premises that had movements from or to detected premises, for a specified monitoring period (14 – 17 days) with a frequency dependent on the risk of infection, and varying level of movement restrictions within 50 km of detected premises. The radius of a vaccination zone is likely to be determined on an ad hoc basis, considering the epidemiological and logistic factors. In this study, the same radius as a protection zone was used. For SOV_{targeted} and SOV, emergency vaccination was implemented on the 14th day since first detection, representing time required for preparing vaccines from the overseas vaccine bank (Katie Owen, personal communication). Resources for vaccination were limited to 200 premises per day, with prioritisation given to unvaccinated premises in the outer radius within the vaccination zone. The premises-level immunity, i.e. proportion of vaccinated

premises protected from infection, was assumed to be 50%, 75%, and 80% on day 4, day 5 and day 6 onwards since vaccine administration, based on the analyses of the FMD outbreak data in Japan (Wada *et al.* 2016).

Economic costs

To quantify and compare the potential economic cost of an FMD epidemic for each alternative control policy, our approach was to separately estimate the short-term costs and macroeconomic costs, and compute the sum as a total cost of an epidemic. Similar approaches have been used for such studies in other FMD-free countries (Hagerman *et al.*, 2012; Boklund *et al.*, 2013; Buetre *et al.*, 2013). All the economic parameters used for this study are presented in Appendix 8 Table 14.22. The methods of estimation were as follows.

Short-term costs

The short-term costs, comprising of the sum of (i) resources for control, including post-epidemic surveillance for freedom of disease, and (ii) compensation, or value loss and production loss due to control measures, were estimated by multiplying the estimated unit costs of the short-term costs by the simulated outcomes of an epidemic. For an epidemic controlled by emergency vaccination, both vaccinate-to-die (VTD) and vaccinate-to-live (VTL) strategies were considered. With VTD policy, all vaccinated animals are subsequently culled as soon as resources become available, whereas vaccinated animals may live throughout their productive lives with VTL policy. Their economic impacts of the strategies are likely to be different, due to the difference in the value loss for vaccinated animals, and a waiting period to recover FMD-free status.

- (i) The costs of resources for control comprised of depopulation, active surveillance, diagnostics, vaccination, and a crisis centre. The unit costs were estimated by aggregating the costs of resources at the premises level or daily level, dependent on the variables for the resources, based on the previously developed MPI's tools¹⁰. For example, the costs of depopulation per premises represented the workload, travel and equipment for a depopulation team, comprised of restricted place manager, security officer, slaughterer, disposer, cleaner, and appraiser, and those of monitoring the premises after it is restocked,

¹⁰ The FMD Preparedness Personnel Calculator developed by Paul Bingham (2014) and Response Cost Calculator FMD large scenario 508 IPs version developed by Bex Ansell (2014).

and decontamination costs. The cost of a crisis centre was assumed to be incurred per day throughout the epidemic period, i.e. from the first detection until the last depopulation and a 3-month post-epidemic period, which represented the workload for management and operation of response teams.

Post-epidemic surveillance was not modelled by simulation, but is an important requirement for recovery of the OIE FMD-free status. The design of post-epidemic surveillance was assumed considering the procedures prescribed in the EU regulation (Anonymous, 2003) as inspection of all premises within 3 km protection zone, a sample of premises within 10 km surveillance zone, and a sample of premises in disease-free zone, divided into 16 regions of New Zealand. A sample size was calculated for each zone by a formula described by Dohoo et al.(2003) as:

$$\left(1 - \alpha^{\frac{1}{D}}\right) \times \left(n - \frac{D-1}{2}\right)$$

where α is a significance level of 0.05, n is the number of premises in the zone, D is the estimated minimum number of infected premises in the zone, assuming a herd-level prevalence of 2% in accordance with the procedures stipulated in the EU Council Directive (Anonymous 2003).

- (ii) For depopulation, 100% value loss of slaughtered animals and 100% production loss for 90 days, while the premises remain empty were estimated for each animal, which may be fully or partially compensated. Two parameters, the current market values of animals and gross margins per day for each animal species were used as a reference. For movement restriction, the inability to take livestock to market or slaughter at an appropriate time would incur direct and indirect losses, with the most dramatic form of welfare culling in 2001 UK epidemic (Schley *et al.* 2009). The effects of movement restrictions at a premises level would be dependent on various factors (e.g. animal species, management types and duration). It has been unquantified or unreported in published literature. For this study, 5% production loss for 90 days was estimated for each animal in a premises where movement restrictions were placed. For VTD, there are no stipulated plans in New Zealand whether

meat produced from vaccinated animals can be sold. According to the EU standards, meat produced from vaccinated animals can be sold in the EU market after being processed separately from that of non-vaccinated animals and after prescribed treatment (Anonymous, 2003). However, carcasses of vaccinated animals were discarded in the two historical FMD epidemics controlled by vaccinate-to-die, i.e. Netherlands, 2001 and Japan, 2010 (Bouma et al., 2003; Muroga et al., 2012). Considering those historical examples, 100% value loss and 100% production loss for 30 days, while the premises remain empty were estimated for each animal for VTD, assuming the salvage values of carcasses of vaccinated animals would offset the costs of culling and processing. This excludes vaccinated dairy cattle, for which we assumed 0% production loss, i.e. no discounting for the value of milk. For VTL, 0% value loss and 0% production loss were assumed for vaccinated animals. There is likely to be an unknown degree of value loss, due to changes in consumers' perception and restrictions in the movements of vaccinated animals throughout their lives.

Macroeconomic costs

The macroeconomic cost of an FMD epidemic was defined as the cumulative net impact on New Zealand's gross domestic products (GDP) over 8 years after the epidemic, discounted by the current standard New Zealand Treasury's discount rate of 8%, which may be considered high but set as such to encourage short-termism (Makhlouf, 2015). The macroeconomic cost represented the long-term cost associated with export bans on livestock products and subsequent disruption in the broad economic sectors in New Zealand.

Previously, Forbes and van Halderen (2014) estimated the macroeconomic costs of an FMD epidemic for three epidemic scenarios in New Zealand. The epidemic scenarios were 1, 52 and 508 IPs eradicated within 1, 50 and 191 days for small, medium and large-scale epidemic scenarios, respectively. For each of the three scenarios, macroeconomic shock assumptions were made including export loss for meat and dairy products, change in domestic consumption, and change in exchange rate over 8 years, and tourism impacts in the first year of an epidemic. Their assumptions were applied into a computable general equilibrium model (CGEM) developed by New Zealand Institute of Economic Research (NZIER). The CGEM estimated the total real GDP effects over 8 years, considering 106 New Zealand industries. The estimated macroeconomic costs were NZD 6.1 billion, 8.2 billion, and 16.2 billion, for three epidemic scenarios, where the duration between first detection and recovery of OIE's FMD freedom was assumed to be 93, 143 and 284 days.

Although desirable, directly linking the CGEM and the epidemic simulation model fell out of scope for this study. Our alternative approach was to approximate the macroeconomic costs of a given epidemic by prediction of a simple linear regression model based on MPI's estimates. It was assumed that the macroeconomic costs, $f(x)$, would increase linearly by the number of days required from first detection of disease until recovery of OIE's FMD-free status (x) as:

$$f(x) = \beta_0 + \beta_{duration} * x,$$

where β_0 is the intercept, or the baseline macroeconomic costs when $x = 0$, and $\beta_{duration}$ is the increment in the macroeconomic costs by 1 day increase in x . By fitting the three estimated figures to this regression model, $\beta_0 = 855.10$ (million NZD) and $\beta_{duration} = 53.72$ (million NZD) were obtained. This could be interpreted as the macroeconomic costs of an epidemic for which FMD-free trade-status were recovered within a day of the first detection (although it is not realistic under the current international standard) is NZD 855.10 million, and the macroeconomic costs increased by NZD 53.72 million by additional day required to recover the OIE's FMD-free status. The adjusted R-squared of the fitted model was 0.988. The previous assumptions made for the three epidemic scenarios were assumed to hold generically, regardless of outbreak locations, control strategies, or absence or presence of vaccinated animals remaining in the country.

Time to recover FMD-free status was determined as 91 days after last depopulation for stamping-out only, 91 days after last depopulation or completion of culling vaccinated animals, whichever occurred last, for VTD, and 183 days after last depopulation for VTL policy, based on the current OIE's Terrestrial Animal Health Code (Anonymous, 2014).

For VTD, culling of vaccinated animals was assumed to be conducted in the local slaughterhouse(s) designated for subsequent cull. The date of completion of culling all vaccinated animals was estimated by the total simulated number of vaccinated animals divided by the 2012 – 2015 regional average peak rate of slaughter (Appendix 8 Table 14.23), assuming slaughterhouse(s) inside the region(s) that contained vaccination zones were designated for culling vaccinated animals, and operated at its peak speed. The start date of subsequent cull was assumed to be the last date of vaccination, i.e. when no more new infections occurred outside the current vaccination zone.

Sensitivity analyses

In the NZSM, local spread represents unknown or unspecified mechanisms of infection to premises located within a few kilometres from the potential source of infection (Sanson et al., 2006a). The local spread parameters in the NZSM derives from the analyses of the outbreak in Cumbria, UK in 2001 (Sanson et al., 2006b). The same set of epidemic simulation was conducted using local spread probabilities that were 1.5 times that of the original parameters, which represented enhanced FMD local spread due to complex epidemiological factors (e.g. virus strain, public awareness, control measures, etc).

Results

Descriptive spatial analyses

The estimated counts of susceptible premises by distance within 5 km of a random livestock premises are shown in Figure 10.2. Auckland Region had four out of five districts with relatively high degree of clustering of livestock premises compared with other districts for both AgriBase and FarmsOnLine (150 - 320 premises within 5 km radius of an arbitrary premises). The estimated counts of premises within 5 km of an arbitrary premises are shown in Figure 10.3. The estimated counts by AgriBase was on average 22.1% less than that of FarmsOnLine, with the minimum 0.6% (Manukau City) and maximum 57.7% (Napier City). Based on the quartiles of the mean degree of clustering, five districts, Franklin (rank 1), Selwyn (rank 14), Tararua (rank 28), Central Hawke's Bay (rank 41), and Central Otago (rank 54) were selected for the following epidemic simulation. Two additional districts from the upper quartile, New Plymouth (rank 3) and Waikato (rank 4), were added in the selection because of their importance for the dairy industry. The summary statistics and the locations of the seven New Zealand districts are shown in Table 10.1 and Figure 10.4.

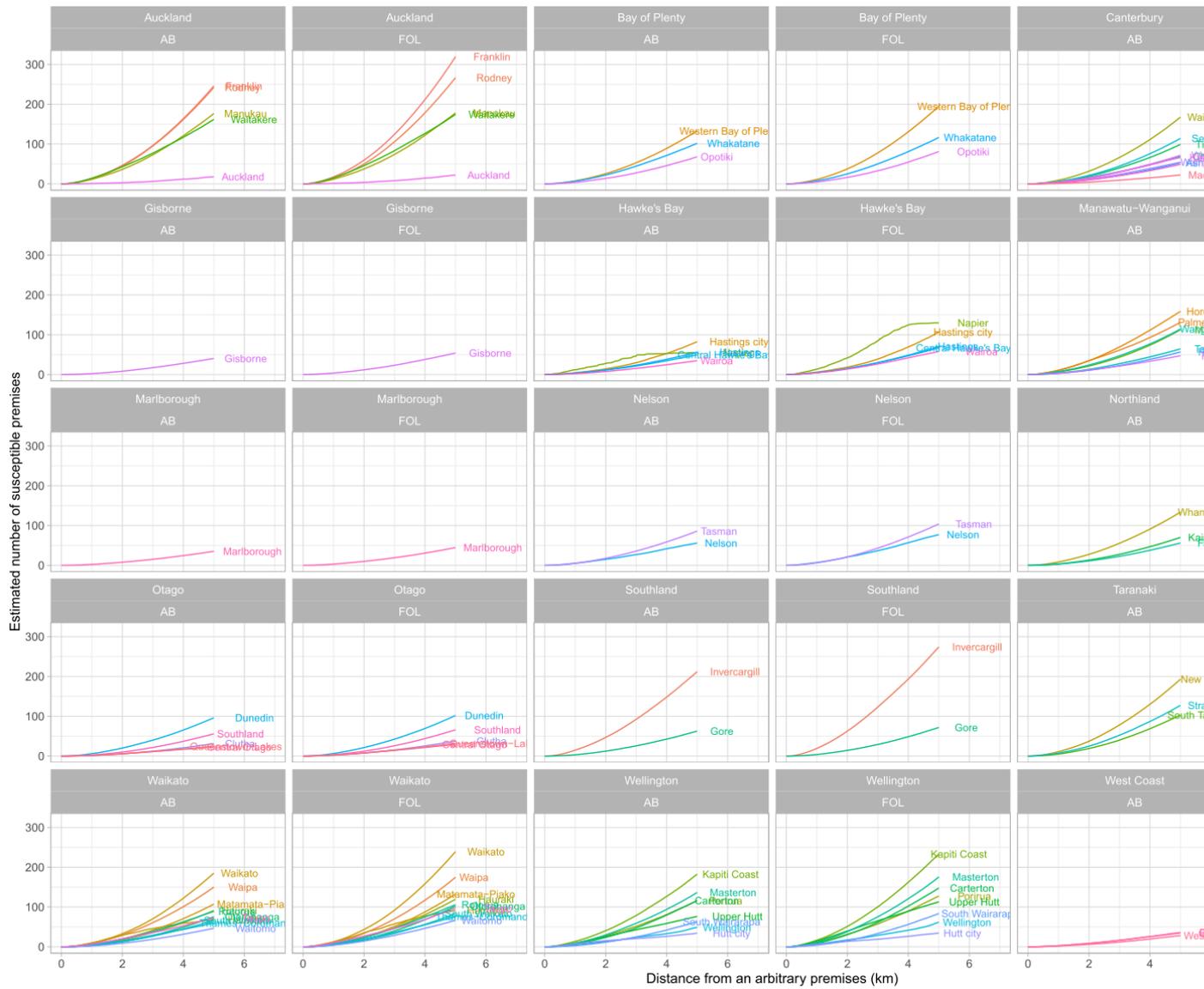


Figure 10.2: Estimated counts of livestock premises centroids by distance from an arbitrary livestock premises centroid for all districts ($n = 69$) in the 16 New Zealand regions based on AgriBase (AB) and FarmsOnline (FOL) livestock databases

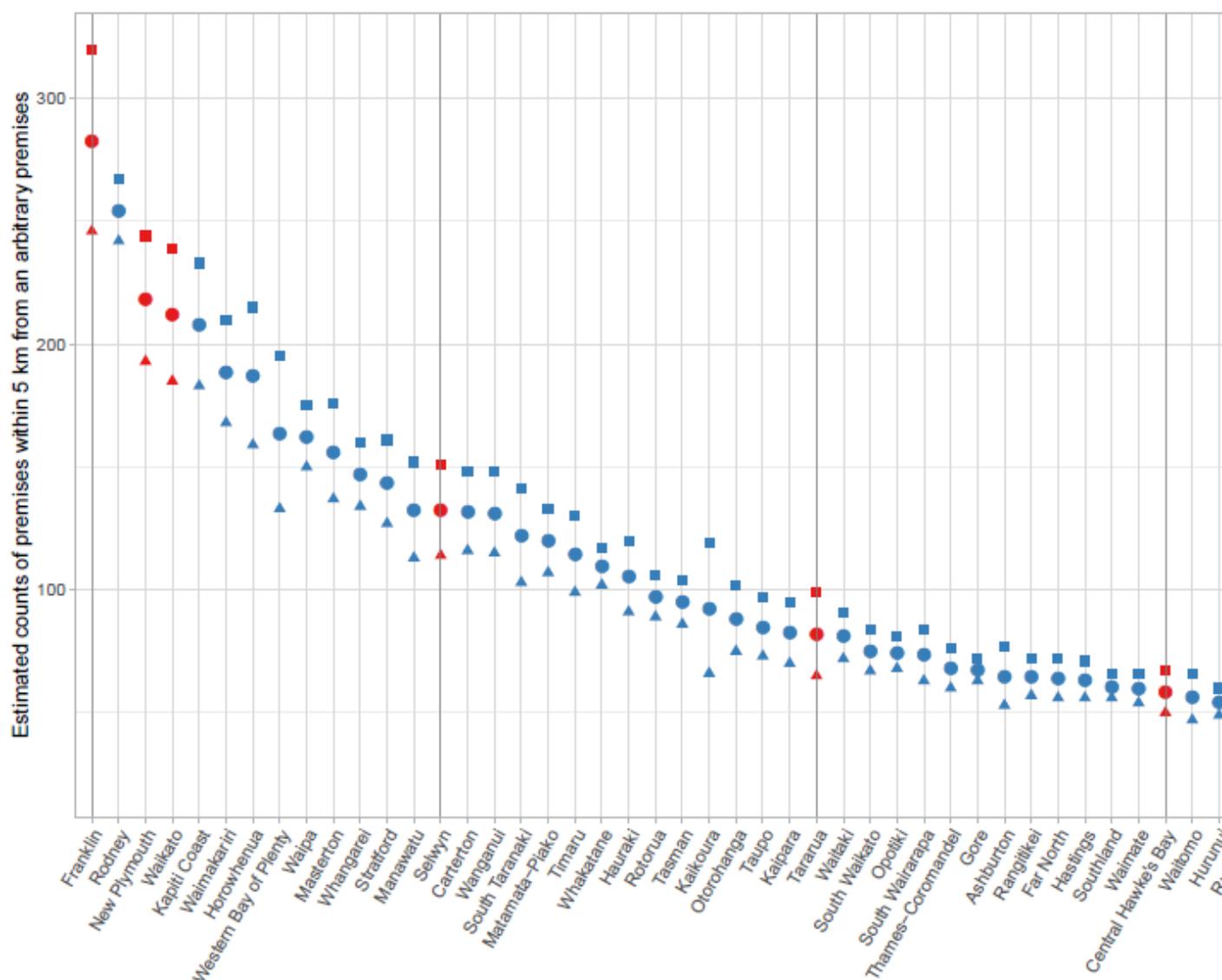


Figure 10.3: Estimated counts of centroids of livestock premises within 5 km of an arbitrary premises centroid for 54 New Zealand districts, based on AgriBase (triangle) and FarmsOnLine (square) livestock database, and the mean of the two (circle), in the descending order by the mean estimates. The seven districts selected for foot and mouth disease (FMD) simulation are shown in red.

Table 10.1: Summary statistics of the seven New Zealand districts selected for foot and mouth disease simulation by AgriBase (AB) and FarmsOnLine (FOL) livestock databases

District	Region	Total number of premises		Area size (km ²)	Estimated counts of premises within 5 km		Rank (54 total)
		AB	FOL		AB	FOL	
Franklin	Auckland	4,317	5,583	2,392	246	320	1

New Plymouth	Taranaki	2,252	2,895	2,269	193	244	3
Waikato	Waikato	3,206	4,201	3,141	185	239	4
Selwyn	Canterbury	2,215	2,862	6,544	114	151	14
Taranua	Manawatu-Wanganui	1,745	2,248	4,392	65	99	28
Central Hawke's Bay	Hawke's Bay	1,130	1,406	3,344	50	67	41
Central Otago	Otago	690	912	9,957	23	29	54

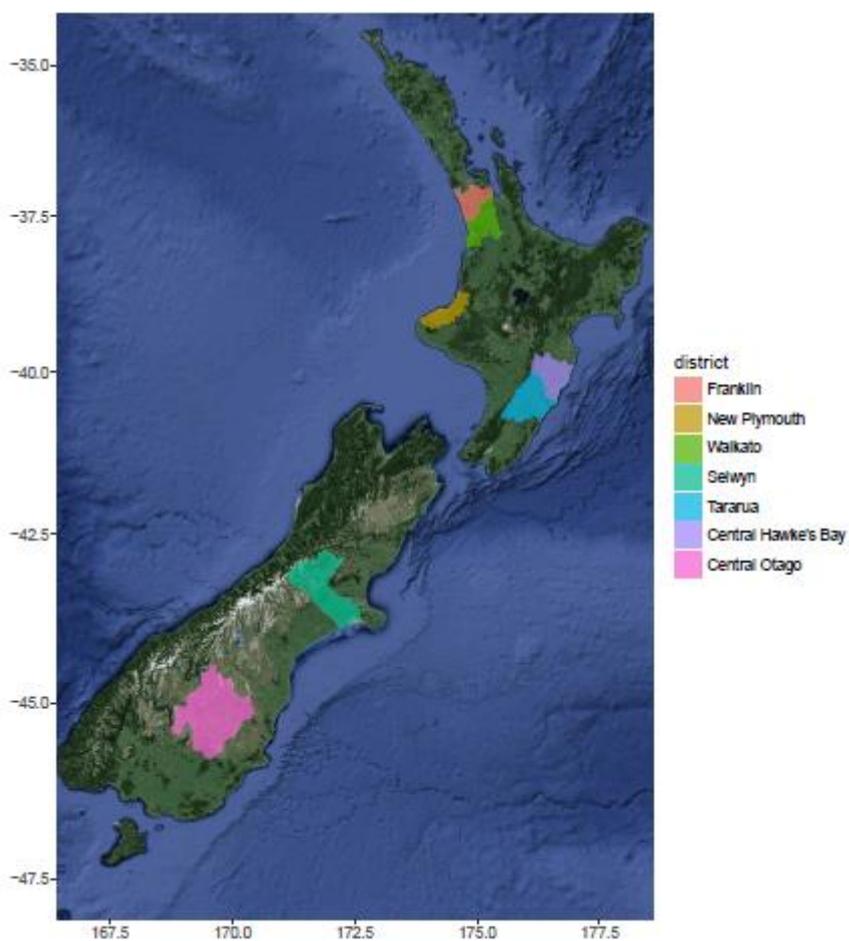


Figure 10.4: *The locations of the seven New Zealand districts selected for simulation of foot and mouth disease (FMD) epidemics*

FMD epidemic simulation

From the premises within the 7 districts of selection (Table 10.1), 777 premises in total (216, 113, 160, 111, 87, 56 and 34 premises in Franklin, New Plymouth, Waikato, Selwyn, Taranua, Central Hawke’s Bay, and Central Otago, respectively, accounting for 5% each of AgriBase records), were selected as a primary case. In total, 195,804 (777 primary cases * 42 iterations * 3 strategies * 2 databases) runs of epidemic simulation were conducted.

The median and the 95th percentile of the cumulative number of detected premises for each incursion scenario are plotted in Figure 10.5. Based on the median results, SOV was more effective in reducing the cumulative number of infections compared with SO for both livestock databases, in the three districts with high degree of clustering (Franklin, New Plymouth and Waikato), by up to 1 – 7 (AgriBase) or 2 – 14 detected premises (FarmsOnLine). In the other districts, SOV had only marginal (<1 premises) or no effects on reducing the median cumulative number of detection. SOV_{targeted} reduced the median cumulative number of detection by up to 2 in Franklin for both databases, while there were only marginal or no effects in other districts.

In some iterations, emergency vaccination was not required to control the epidemics at the presumed time of decision making (i.e. day 14 after first detection). The percentages of iterations where an epidemic died off without being detected were <6.1% and <0.4%, and those eradicated within 14 days with SO strategy were 18.1 – 54.8% and 15.0 – 49.8%, for AgriBase and FarmsOnLine, respectively (Table 10.2). The epidemiological outcomes of the epidemics lasting for ≥ 14 days are shown in Figure 10.6. There was, in general, a decreasing trend in the scale of an epidemic, as measured by the number of detected premises and epidemic duration, as control strategies changed from SO, SOV_{targeted}, and SOV. The decreasing trend was stronger in the districts with higher degree of clustering (i.e. Franklin, New Plymouth and Waikato).

Table 10.2: *Number of iterations (and percentage) where emergency vaccination was not required to eradicate a simulated foot and mouth disease epidemic in the seven New Zealand districts for AgriBase (AB) and FarmsOnLine (FOL) databases*

District	Region	Number of undetected ^{*1} epidemics (%)	Number of epidemics eradicated within 14 days (%)	Total number of iterations ^{*2}
----------	--------	---	--	--

		AB	FOL	AB	FOL	
Franklin	Auckland	218 (2.4%)	1 (<0.0%)	1,642 (18.1%)	1,361 (15.0%)	9,072
New Plymouth	Taranaki	90 (1.9%)	2 (<0.0%)	1,248 (26.3%)	1,032 (21.8%)	4,746
Waikato	Waikato	91 (1.4%)	9 (0.1%)	1,818 (27.1%)	1,694 (25.2%)	6,720
Selwyn	Canterbury	93 (2.0%)	4 (0.1%)	1,562 (33.5%)	1,449 (31.1%)	4,662
Tararua	Manawatu-Wanganui	7 (0.2%)	2 (0.1%)	1,455 (39.8%)	1,155 (31.6%)	3,654
Central Hawke's Bay	Hawke's Bay	5 (0.2%)	1 (<0.0%)	1,069 (45.5%)	923 (39.3%)	2,352
Central Otago	Otago	87 (6.1%)	6 (0.4%)	783 (54.8%)	711 (49.8%)	1,428

*1 A primary case was not detected with infection for 30 days.

*2 Total number of iterations was based on 5% of the total number of premises in each district recorded in AgriBase (primary case) * 42 repetitive iterations.

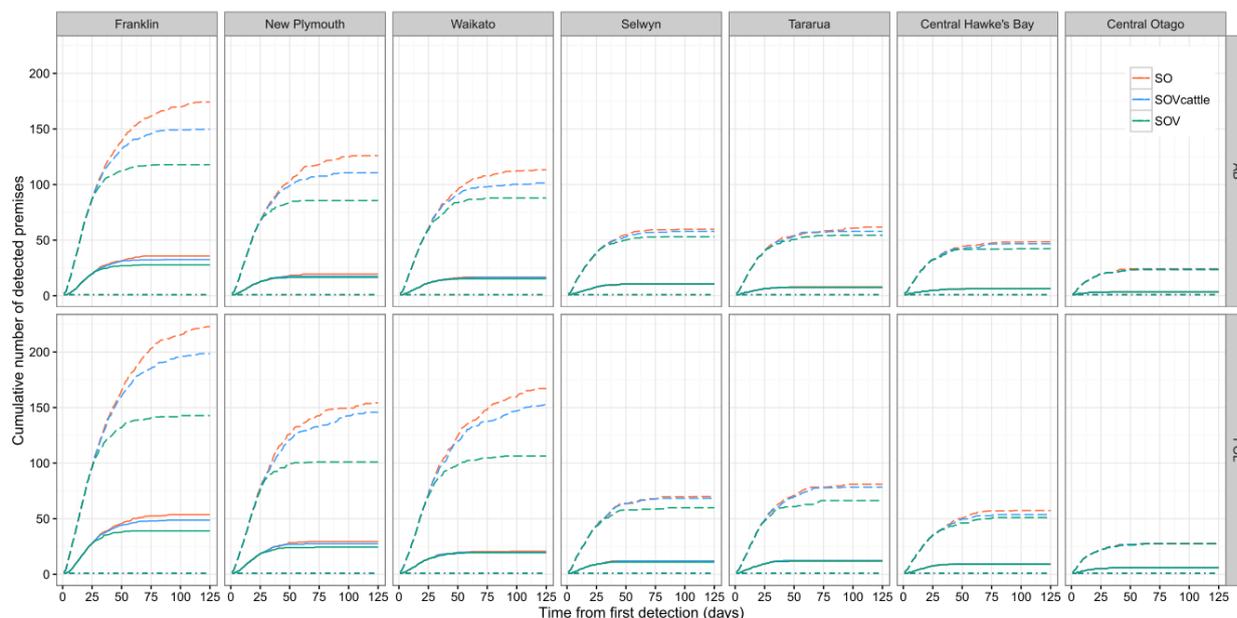


Figure 10.5: The cumulative number of premises detected with foot and mouth disease (FMD), by

number of days elapsed since first detection for three control policies and seven districts, based on disease simulation using AgriBase (AB) and FarmsOnLine (FOL) data. The policies are stamping-out only (SO), SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises ($SOV_{targeted}$) or all premises (SOV). The median (solid line) and the 5th and the 95th percentile (dash and dot-dash line) are shown.

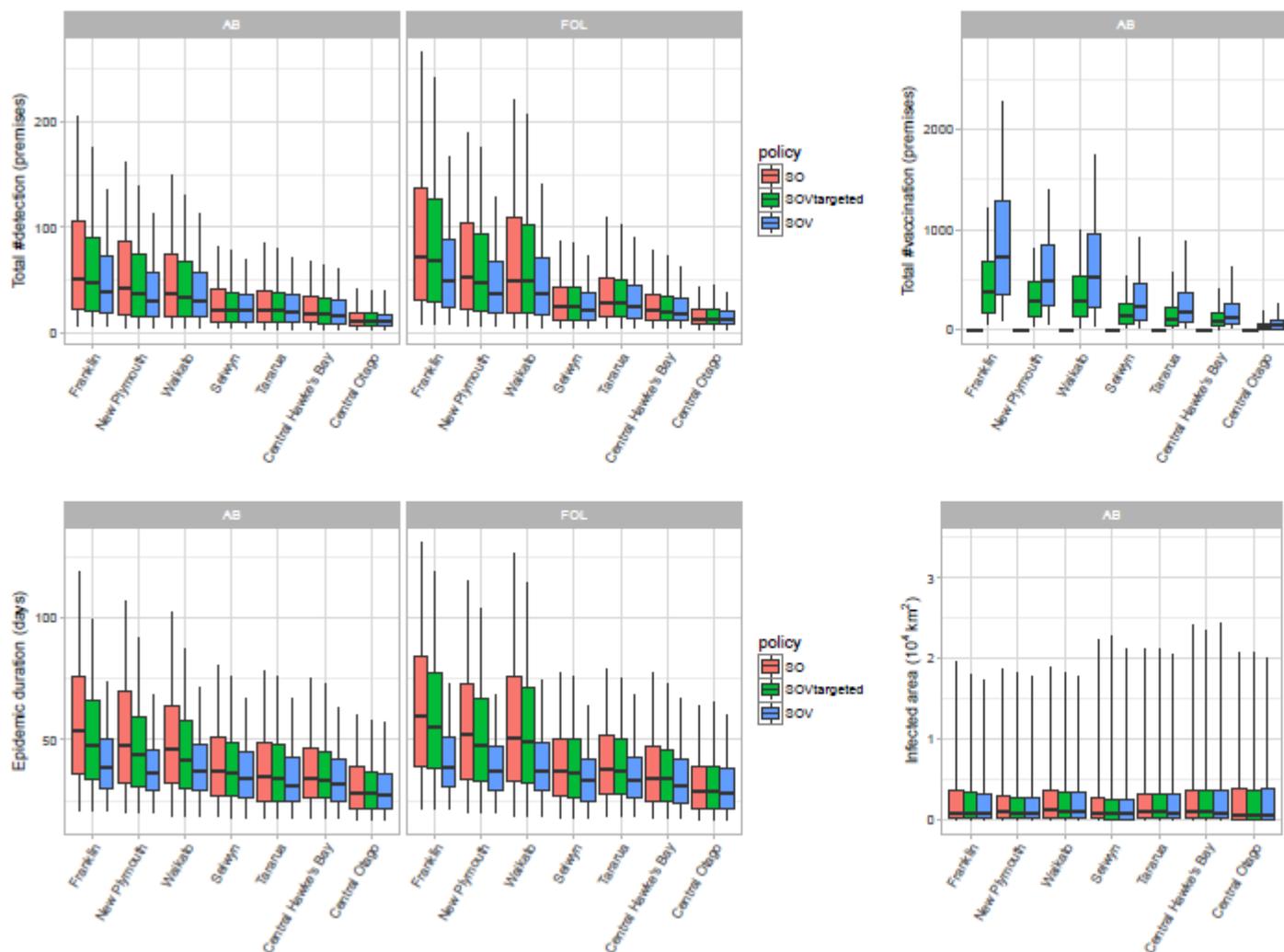


Figure 10.6: The simulated epidemiological outcomes of foot and mouth disease (FMD) epidemics lasting for ≥ 14 days for three control policies and seven New Zealand districts, using AgriBase (AB) and FarmsOnLine (FOL) data (total iterations: $n = 22,466$ and $24,284$ for AB and FOL). The policies are stamping-out only (SO), SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises ($SOV_{targeted}$) or all premises (SOV). Box represents the first and the third quartiles, and whisker the 5th and 95th percentiles, and outliers ($<5^{th}$ and $>95^{th}$ percentiles) were removed.

Economic outcomes

The estimated costs of the simulated FMD epidemics lasting for ≥ 14 days by policy, district and database are shown in Figure 10.7. The median short-term costs and total costs of an epidemic lasting for ≥ 14 days and controlled by SO were in the range of NZD 11.3 – 26.9 million and NZD 7.3 – 9.0 billion, with the lowest in Central Otago (AgriBase) and highest in Franklin (FarmsOnLine), respectively. For SOV_{targeted} and SOV with VTD or VTL, the median short-term costs were NZD 19.9 – 98.2 million (SOV_{targeted}-VTD), 12.3 – 29.9 million (SOV_{targeted}-VTL), 20.4 – 98.4 million (SOV-VTD), and 12.5 – 41.6 million (SOV-VTL), respectively, whereas the median total costs were NZD 7.3 – 9.4 billion (SOV_{targeted}-VTD), 12.2 – 14.3 billion (SOV_{targeted}-VTL), 7.3 – 8.6 billion (SOV-VTD), and 12.2 – 13.5 billion (SOV-VTL), respectively. The median short-term costs and total costs of an epidemic lasting for ≥ 14 days and controlled by SO were 4.3 – 16.8 times and 1.2 – 1.5 times that of an epidemic eradicated within 14 days, respectively (data not shown).

Figure 10.8 shows the marginal benefit cost ratio (B/C) of the vaccination strategies relative to SO by district and database. The median B/C was greater than 1 with SOV with VTD (SOV-VTD), in Franklin by FarmsOnLine (1.22) and New Plymouth by either AgriBase (1.16) or FarmsOnLine (1.18). Otherwise, the median B/C was less than 1, i.e., SO was the most economically preferable policy.

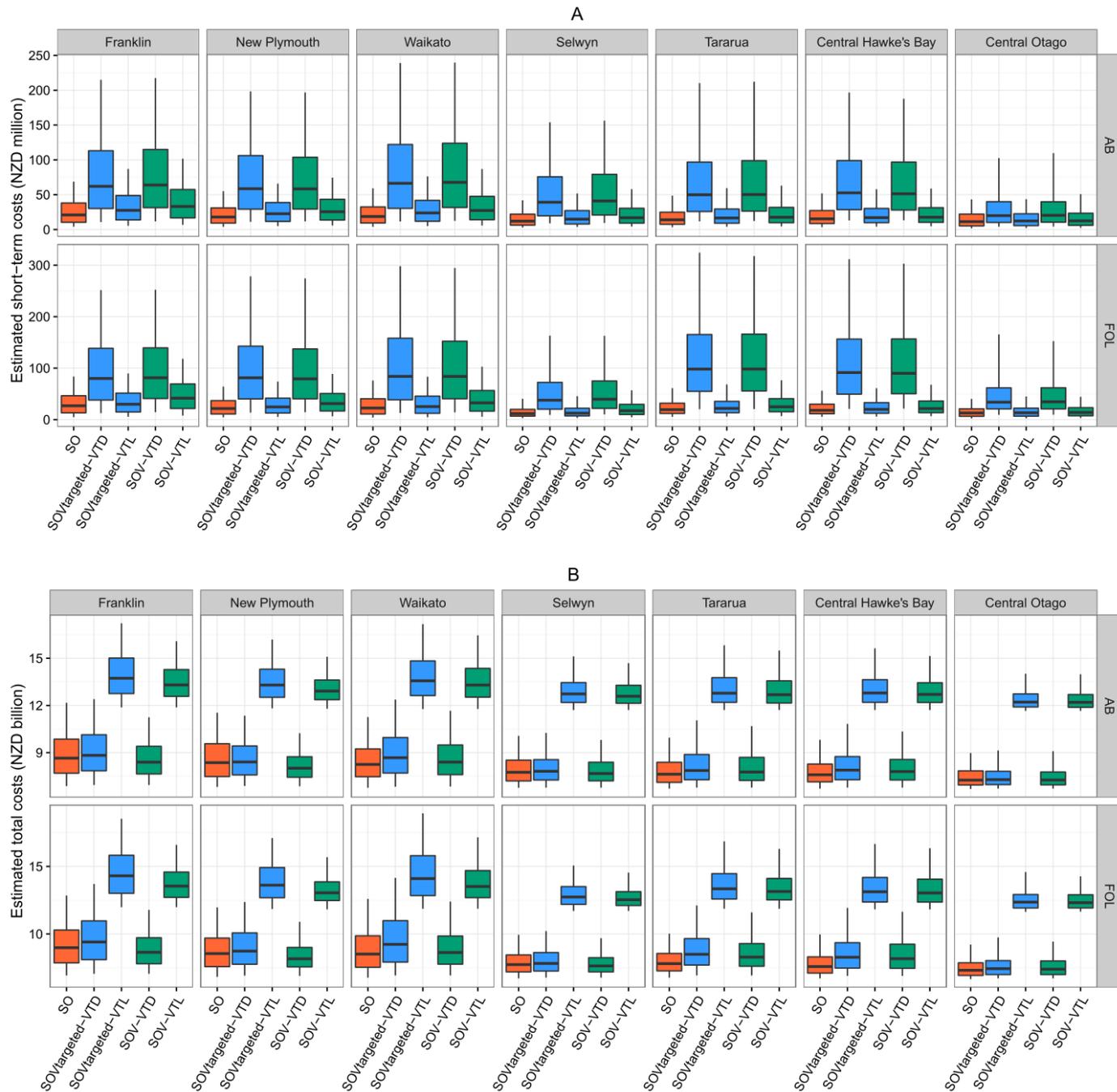


Figure 10.7: The estimated short-term costs [A] and total costs [B] of simulated foot and mouth disease (FMD) epidemics in seven New Zealand districts, with seven control strategies, based on AgriBase (AB) and FarmsOnLine (FOL) databases. The strategies are stamping-out only (SO), SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises ($SOV_{targeted}$) or all premises (SOV), with vaccinate-to-die (VTD) or vaccinate-to-live (VTL)

policy. Box represents the first and the third quartiles, and whisker the 5th and 95th percentiles, and outliers (<5th and >95th percentiles) were removed.

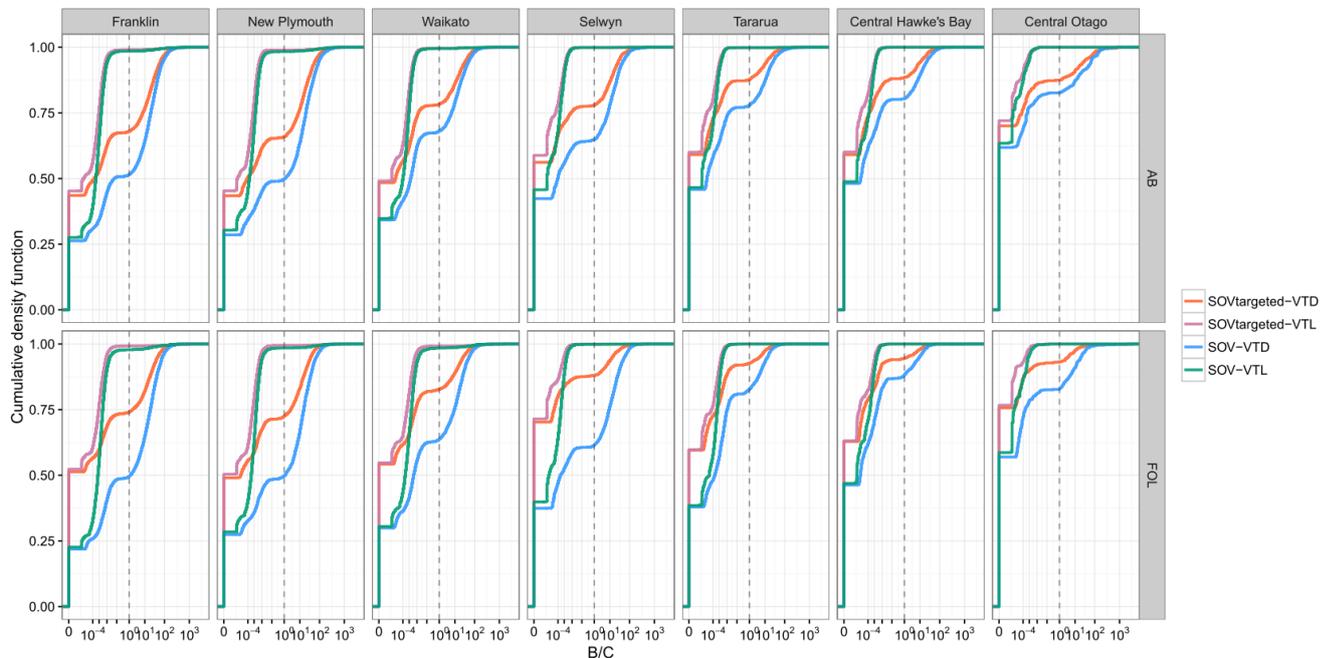


Figure 10.8: Cumulative density functions of the estimated marginal benefit cost ratio (B/C) of four vaccination-based foot and mouth disease (FMD) control strategies relative to stamping-out alone (SO), for simulated epidemics lasting for ≥ 14 days in seven New Zealand districts, based on AgriBase (AB) and FarmsOnLine (FOL) databases. The strategies are SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises ($SOV_{targeted}$) or all premises (SOV), with vaccinate-to-die (-VTD) or vaccinate-to-live (VTL). The dotted line shows $B/C = 1$, above which the benefits exceed the additional costs (i.e. i.e. addition of emergency vaccination is more beneficial than SO).

Sensitivity analyses

The estimated B/C of the vaccination strategies relative to SO with enhanced local spread probabilities are shown in Figure 10.9. The median NPV was greater than 0 and B/C greater than 1 with SOV with vaccinate-do-die (SOV-VTD), in Franklin by AgriBase (NPV: NZD 1.6 billion and B/C : 21.30) and FarmsOnLine (NPV: NZD 14.9 billion and B/C : 29.18), New Plymouth by AgriBase (NPV:

NZD 1.5 billion and B/C: 14.85) and FarmsOnLine (NPV: NZD 1.0 billion and B/C: 14.91), Waikato by AgriBase (NPV: 0.7 billion and B/C: 2.68) and FarmsOnLine (NPV: NZD 0.1 billion and B/C 7.85), and Selwyn by AgriBase (NPV: NZD 0.1 billion and B/C: 1.91) and FarmsOnLine (NPV: <0.1 billion and B/C: 3.30). Otherwise, the median NPV was smaller than 0 and B/C smaller than 1, i.e. SO was preferred by both databases.

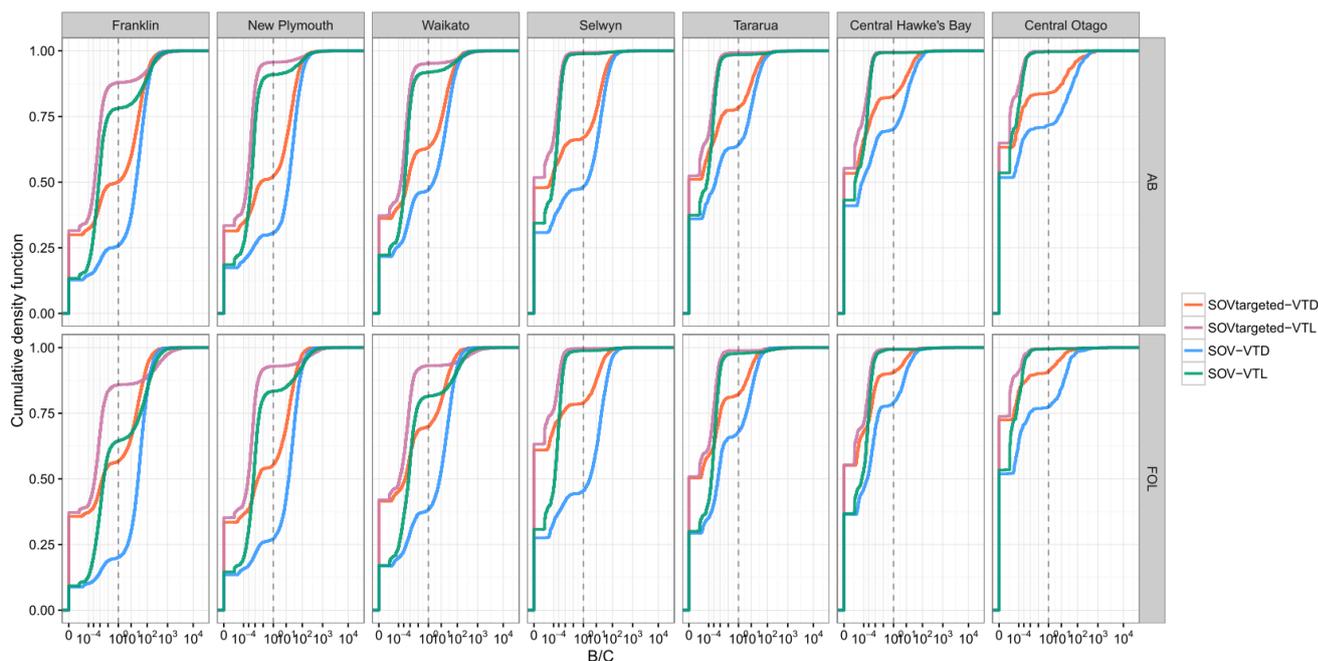


Figure 10.9: Sensitivity analyses for local spread probabilities 1.5 times that of the original values: cumulative density functions of the estimated marginal benefit cost ratio (B/C) of four vaccination-based foot and mouth disease (FMD) control strategies using vaccination relative to stamping-out alone (SO), for simulated epidemics lasting for ≥ 14 days in seven New Zealand districts, based on AgriBase (AB) and FarmsOnLine (FOL) databases. The strategies are SO with emergency vaccination within 3 km vaccination zone targeting pastoral, dairy and dry stock grazing premises (SOV_{cattle}) or all premises (SOV), vaccinate-to-die (-VTD) or vaccinate-to-live (VTL). The dotted line shows $B/C = 1$, above which the benefits exceed the additional costs (i.e. i.e. addition of emergency vaccination is more beneficial than SO).

Discussion

In this study, two agricultural premises databases (AgriBase and FarmsOnLine) were compared for informing preparedness for response to an outbreak of an exotic disease in New Zealand such as FMD. Although simulation of FMD epidemics using NZSM was carried out in various other studies (Sanson *et al.* 2006; Owen *et al.* 2011; Owen *et al.* Unpublished results), this was, to the best of our knowledge, the first study which compared the results of NZSM between AgriBase and FarmsOnLine. It was shown that ~20% difference in the density of livestock premises and inconsistent distributions of farm types between the two databases were not critical for determining the most economically beneficial strategies for FMD. It should be noted, however, that these results were based on the assumption that the information in each database was accurate. The simulation model did not consider the effects of having an inaccurate livestock information. For example, if active surveillance was conducted only on 80% of the actual premises at risk of infection (i.e. AgriBase underestimated the number of premises), 20% of infected premises would be undetected, contributing to the spread of an epidemic. If the list of premises at risk contained duplicated or non-susceptible premises (i.e. FarmsOnLine overestimated the number of premises), resources would be spent inefficiently on low-risk premises, and there may be a potential consequential delay in surveillance activities on premises truly at risk of infection.

The spatial descriptive analyses identified districts with relatively high degree of clustering of livestock premises, i.e. higher risk of local spread if FMD was introduced into the district. The estimated counts of livestock premises within 5 km of an arbitrary premises by AgriBase was found to be on average 22% less than that of FarmsOnLine, which was similar to the difference in the number of recorded premises susceptible to FMD, i.e. AgriBase (n = 81,750) has 19% less records than FarmsOnLine (n = 101,539). This difference is likely to be important for outbreak response activities in the area with a high level of clustering if, as mentioned above, the information on livestock premises used for response activities were inaccurate. In Franklin, for instance, 74 out of 320 premises within 5 km from an FMD-infected premises would be missed by AgriBase alone, despite their risk of secondary infection by local spread, if the records of livestock premises in FarmsOnLine were assumed to be accurate. The difference between the two databases may be due to the different ways data were collected, particularly around how owners are assigned to land parcels, as discussed by Jewell *et al.* (2016). It should also be noted that this study used AgriBase

collected in 2011, while FarmsOnLine in 2016. Although it is desirable to use two datasets collected at the same point of time, the level of difference between the two datasets was in general similar to that collected at the same point of time (i.e. about 20%) (Jewell et al., 2016).

In Franklin, the economically preferred FMD control strategies were not consistent between AgriBase and FarmsOnLine; with FarmsOnLine, SOV with VTD (SOV-VTD) was preferred by a median NPV of NZD 14.9 million and median B/C of 1.2, whereas with AgriBase, any of the vaccination-based policies were suboptimal compared with SO with a median NPV of <−NZD 7.4 million and a median B/C of <0.1. Although the difference in the NPVs (<NZD 15 million) would be important for the stakeholders who would bear the costs, i.e. MPI, consumers and livestock producers, it was minimal in comparison with the scale of the total costs of an FMD epidemic (NZD >8 billion).

The results of this study are strictly limited to the epidemiological and economic assumptions made for this study. Particularly, due to the limitation in the linkage between the FMD simulation model and the macroeconomic model, a simplistic linear assumption was used for estimation of the macroeconomic costs. However, a relative importance in the estimate of the macroeconomic costs was emphasised, by high ratios of the macroeconomic costs to the short-term costs (range: 20 – 33,000). It is recommended that future studies should be directed to refining the estimates for the macroeconomic costs (rather than the short-term costs which accounts for <6% of the macroeconomic costs) by elaborating the linkage between epidemiological and macroeconomic models, and account for other factors, such as region, policy (SO/VTD/VTL), and changes in domestic and international markets. It should also be noted that the whole approach used in this study is predicated on the only important aspect represented by economic terms. It is entirely possible that other aspects of an epidemic (animal/human welfare, environmental issues, business continuity, etc), not quantified in the economic terms, may be valued highly by some stakeholders.

The sensitivity analysis illustrated how the benefits of vaccination strategies may be influenced by change in the epidemiological conditions. In this study, variation in local spread was examined as it was found to be the major mechanism of disease transmission in the previous FMD outbreaks in the UK in 2001 and Japan in 2010 (Gibbens and Wilesmith, 2002; Wada et al., 2016). The benefits of vaccination strategies were in general higher with the enhanced local spread (NPV: ≤NZD 3.1 billion and B/C: ≤29.2) than that of the original local spread parameters (NPV: ≤NZD 14.9 million and B/C: ≤1.2), resulting in vaccination-based strategies preferred to SO in more districts (Franklin, New

Plymouth, Waikato and Selwyn). While the NZSM depends on the local spread estimated from the 2001 UK outbreak, there is no guarantee that the pattern of local spread of a future FMD outbreak in New Zealand (if it ever occurred) would be similar to, or 1.5 times that of the UK. It may well vary because local spread is determined by the complex factors including virus strain, public awareness and efficiency of restrictions to prohibit such movements (Wilesmith et al., 2003). Before being used to inform decision making, further research is required to examine variation in the benefits of vaccination by local spread, for informing decision making. Although not examined in this study, other conditions that may increase the relative benefits of vaccination strategies include a higher proportion of vaccinated premises protected from infection (>80%), and stricter constraints in the resources for prompt operation of stamping-out activities.

When the decision to vaccinate is made, vaccination of cattle only could be considered, as an alternative strategy to focus resource use on the sub-population which plays an important role in FMD epidemiology (Hickey, 2014). However, due to the limitation in the InterSpread Plus programme, it was not possible to model vaccination based on animal species present in premises. In this study, we approximated this by targeting commercial enterprises where cattle were important (pastoral, dairy and dry stock grazing farms). Due to differences in the distribution of farm types between AgriBase and FarmsOnline, a greater proportion of premises were targeted for vaccination by SOV_{targeted} for AgriBase (53.4 – 84.9%) than FarmsOnline (24.5 – 65.4%). Regardless of the difference in the distributions of farm types between the two databases, SOV_{targeted} was shown to be a suboptimal option. This may be because the herd immunity gained by targeted vaccination was insufficient to achieve the benefits (e.g. less number of infection and shorter duration) that would exceed the additional costs of vaccination themselves. More research is needed to explore a cattle only vaccination option.

Estimation of the economic impact of FMD involves a great deal of uncertainty. While estimation of the operational costs of control measures is relatively straightforward, knowledge on the economic impacts of control activities on livestock enterprises is limited, both in terms of productivity and product values. In particular, value loss of vaccinated animals by FMD emergency vaccination is likely to be an important determinant for the benefit of control policies using vaccination. However, there is no stipulated plan regarding processing of products derived from vaccinated animals, or no similar experiences to base assumptions upon. This study simply added the whole market value of vaccinated animals for VTD and 0% for VTL. However, the costs of vaccination for VTD may be

reduced by effective processing and marketing of vaccinated animals, as stipulated in the EU standards (Anonymous, 2003). While this may apply for VTD in New Zealand, it requires sufficient infrastructure to process a large number of meat from vaccinated animals in a short term (e.g. workforce, transport, slaughterhouses, and storage), and excessive supply during imposition of export bans by current trading partners may result in significant reduction in price. To illustrate this, the Taiwanese pork industry experienced up to a 70% reduction in the price of pigs during the export bans after an FMD outbreak in 1997 (Yang et al., 1999; Chang et al., 2006). In addition, there could be loss of perceived product value due to consumer fear towards animals treated with new and unfamiliar vaccines (Scudamore, 2007), without consumers' correct knowledge about FMD vaccination. VTL was always not preferred to VTD in terms of the total costs of an epidemic, as the increase in the macroeconomic costs due to a longer waiting period (6 months instead of 3 months) to recover FMD-free status was much greater than savings in the value of vaccinated animals.

Consideration should be given to various externalities excluded from the economic criteria. For an FMD epidemic, these negative effects include psychological distress in livestock owners and field veterinarians directly involved in the culling activities (Hunter, 2001; Hannay and Jones, 2002; Van Haaften et al., 2004; Olff et al., 2005; Peck, 2005), external value loss from culling pedigree breeding stock that have been genetically selected over many years (Power and Harris, 1973; Nishiura and Omori, 2010), companion animals that had a greater value than their market price, environmental pollution due to incineration and burial (Anderson, 2002), and animal welfare. These effects are challenging to quantify in economic terms due to lack of data and a difficulty in assigning economic values, but are likely to be important particularly at the individual level (Anderson 2002).

Conclusions

This study has demonstrated that both AgriBase and FarmsOnline produced consistent policy recommendations for FMD control based on the predicted benefits, despite the difference in the counts of livestock premises at risk of infection (22%). The results may provide policymakers with indication regarding areas where efforts should be focused in increasing the accuracy of national livestock data for New Zealand.

Deliverable 9: Capacity development and information transfer

Mary van Andel¹

¹ *Ministry for Primary Industries, Wellington, New Zealand 6140*

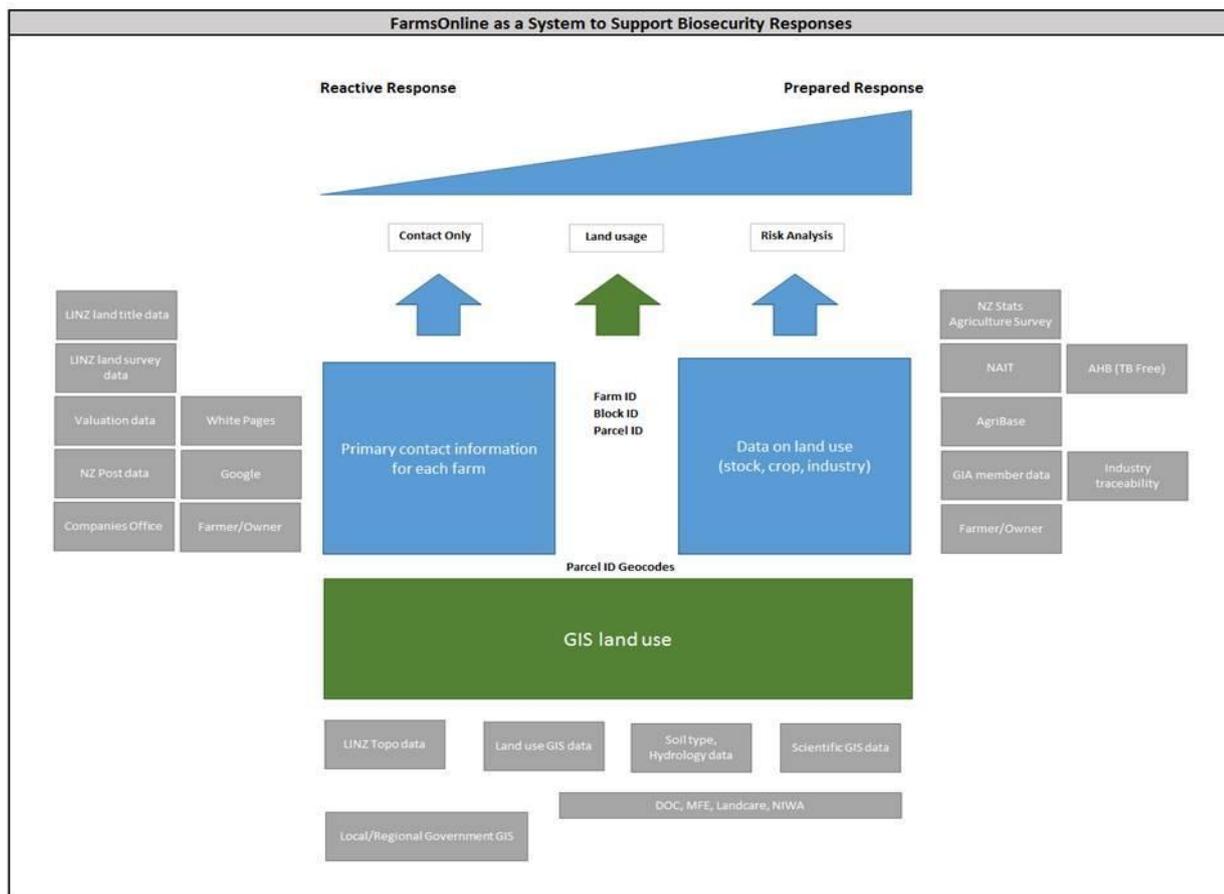
This CEBRA project has greatly increased the ability of MPI and CEBRA staff involved in the project to provide advice on the strengths and weaknesses of alternative approaches to both MPI and DAWR. There have been opportunities to provide input to ongoing MPI projects and to initiate new data sharing directions. Aside from the new directions, the staff involved have been able to supply advice and direction on existing data sharing projects (for example the ongoing data sharing arrangements between FarmsOnLine and the National Animal Identification and Tracing System - NAIT).

We highlight examples to illustrate this:

1. Ongoing discussions with Statistics New Zealand to align census data collection with FarmsOnLine (FOL) polygon information at the farm-level. Discussions with Statistics New Zealand have been reopened to examine solutions that will allow the use of statistics New Zealand census and survey information for biosecurity purposes. With this project as background we have been able to explain our needs and identify possible technical issues early in the process. A formal proposal to Statistics New Zealand has been made, MPI's legal team has provided advice and interprets the Statistics Act to mean that MPI should be allowed to use data collected in the Agricultural Production Census for biosecurity purposes (including preparedness and response). Challenges in this space include historical interpretation of the Statistics Act, large numbers of stakeholders and perceived reputational risk on the part of Statistics New Zealand. The Statistics Act is interpreted by the legal department of Statistics New Zealand to mean that data collected by Statistics New Zealand may not be used for Biosecurity purposes which include preparedness and response. Negotiations and discussions are ongoing and the last update is that Statistics New Zealand may require a request to be made at the ministerial level for a change of use of the Agricultural Production Survey/Census data for MPI.
2. MPI launched and completed an internal project to identify future uses and opportunities for the data housed in FOL. CEBRA project team members were able to provide detailed

technical information based on Deliverable 3 to this project team and were also able to provide details of other New Zealand role holders who would be able to contribute to the MPI project. We believe that this is a success story for our CEBRA project team as the depth of advice we were able to supply is a direct benefit from 1402C. As a point of clarity, the Surveillance and Incursion Investigation, Animals and Marine Team (SIAM) is the team that Mary van Andel and Daan Vink are in at MPI. FOL is FarmsOnLine, the MPI database established for biosecurity purposes which houses spatial information and occupier information on all rural properties in New Zealand.

3. Following the completion of item 2, MPI is in the process of building a plan for the future of FOL in which Mary is consulted on a weekly basis. Project 1402C has provided the background needed to be able to contribute information to this process and findings from this project are being incorporated into the project team's recommendations. The following diagram of how a common understanding of how FOL data should be used has been built by Mary and the FOL project team to build common understanding of the data required during response. .



A reactive biosecurity response is one where contact can be made with the farm owner but there is little knowledge of animals present and land use
 A prepared biosecurity response has knowledge of animals and plants on the farms and an understanding of how diseases or organisms spread based on environmental and land use risk analysis

Figure 11.1: Concept diagramme to build understanding of the data needs during a response. Attributed to Chris Taylor, a senior business analyst in Planning and Readiness in the Intelligence, Planning and Coordination Directorate.

4. Findings from data analysis performed for 1402C and 1502C have allowed highly detailed feedback to be provided to OSPRI, the parent company of the National Animal Identification and Tracing System (NAIT) about data limitations and needs. This knowledge will inform advice on the inclusion of this data in the New Zealand Standard Model (NZSM) of FMD spread.

5. A high-level view of the utility of each of the national data sources was able to be provided to MPI decision makers in a concise manner following this project as illustrated in the table below.

Table 11.1: *High-level view of the utility of each of the national data sources available in New Zealand*

Required for Response	FOL	AgriBase	NAIT	APS
Full contact detail of KDM ¹¹	✓	✓	✓	✗ ¹²
spatial location (point & polygon)	✓ ¹³	✓	✗ ¹⁴	✗ ¹⁵
animal numbers by species & class	✗	✓ ¹⁶	✓ ¹⁷	✓ ¹⁸
hectares of different plantings	✗	✓	✗	✓ ¹⁹

¹¹ Key Decision Maker (KDM) - depending on how the information is collected, this may differ between datasets. Example – NAIT collects the details of the person in charge of animals (PICA). This is not always the owner of the animals or the owner of a property.

¹² The Agricultural Production Survey (APS) is based on GST registration. The details of the person paying tax on a farming business may not be the same as the details of a key decision maker on a day to day basis.

¹³ Both AgriBase and FOL farms are built from Land information New Zealand (LINZ) parcels. The databases could be merged. A match between farms at the parcel level that was updated at an agreed interval would allow animal numbers to be transferred to FOL but would divert resource from a single source of truth.

¹⁴ A NAIT number is a relationship between a FOL number and a PICA. NAIT can be spatialized by using FOL information. There is a caveat where it is possible to have many to many relationships between NAIT and FOL numbers. FOL numbers within 20km of each other that are used in the same business may have a single NAIT number.

¹⁵ The only current way to spatialize the APS data is to geocode the address field in the data or to match the address field with the address field in an existing farms database. There is about 30% matching rate by this method as shown in Deliverable 6. Inclusion of a farm-level identifier (provision must be made for more than one farm identifier per GST registered business as one business may operate on multiple land blocks or farms) is a priority.

¹⁶ AgriBase has animal counts based on farmer generated information. This is collected by survey and by farmers logging onto the system to complete tasks required by their local councils or industry bodies.

¹⁷ NAIT has numbers of cattle and deer associated to a NAIT number. Not all animal species are covered and the many-to-many relationship issue means that in about 16% of cases animals cannot be localised to a single FOL number (could be on a selection of FOL numbers within 20km of each other). Even in the event of more species being added this will not provide crop information.

¹⁸ The Agricultural Production Survey (APS) and Census covers only those individuals registered for GST purposes. In a census year, about 70 000 farms are surveyed.

Land use type

✓²⁰

✓

✓²¹

-

¹⁹ The Agricultural Production Survey (APS) and Census covers only those individuals registered for GST purposes. In a census year, about 70 000 farms are surveyed.

²⁰ FOL has information on land use type which is assigned ad hoc by the local council. This information is not reliable. In AgriBase farmers allocate their primary land use type.

²¹ NAIT may provide land use type however this is limited to dairy, beef and deer production types. Even if more species being added this will not provide crop information.

Recommendations

This project provides a comprehensive report on the current state of agricultural data in NZ and the assessment of and potential for different methods for filling data gaps. The value of not having this information is also discussed. There are several key points and recommendations arising from the report:

1. While remote sensing of individual animals shows promise in the next 5 years, it's not a current solution
2. SDMs are not sufficiently detailed for operational purposes although they could be useful inputs for disease models and for some preparedness activity.
3. We have documented the exact uses of farm level data and what it is required for in the context of biosecurity preparedness and response for transboundary animal diseases.
4. Every effort must be made to align existing data sources to a single sustainable source of NZ data
5. NZSM is somewhat robust to variation in density of farms

Further work will inform the robustness of within farm spread models to herd size and the effect of this on decision support.

References

- Alexandersen, S., Brotherhood, I., Donaldson, A.I., 2002. Natural aerosol transmission of foot-and-mouth disease virus to pigs: minimal infectious dose for strain O-1 Lausanne. *Epidemiology and Infection* 128, 301-312.
- Alexandersen, S., Donaldson, A.I., 2002. Further studies to quantify the dose of natural aerosols of foot-and-mouth disease virus for pigs. *Epidemiology and Infection* 128, 313-323.
- Anderson, I., 2002. Foot and mouth disease 2001: Lessons to be learned inquiry report. House of Commons London, UK.
- Anonymous, 2003. Council Directive 2003/85/EC of 29 September 2003 on Community measures for the control of foot-and-mouth disease repealing Directive 85/511/EEC and Decisions 89/531/EEC and 91/665/EEC and amending Directive 92/46/EEC. European Union.
- Anonymous. (2011), WHOLE-OF-GOVERNMENT BIOSECURITY RESPONSE GUIDE.
- Anonymous, 2014. Terrestrial Animal Health Code. Foot and mouth disease. Office International des Epizooties (OIE).
- Araujo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33, 1677-1688.
- Baddeley, A., Rubak, E., Turner, R., 2015. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press London.
- Barber-Meyer, S.M., Kooyman, G.L., Ponganis, P.J., 2007. Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology* 30, 1565-1570.
- Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., Meza, R.A., Roberts, D.R., Paris, J.F., Spanner, M.A., 1994. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American Journal of Tropical Medicine and Hygiene* 51, 271-280.
- Belton, D.J., 2004. The macro-economic impact of a foot-and-mouth disease incursion in New Zealand. *Dev Biol (Basel)* 119, 457-461.
- Bessell, P.R., 2009. *The spatial epidemiology of Foot and Mouth Disease in Great Britain*. The University of Edinburgh.
- Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E.J., 2010a. Estimating risk factors for farm-level transmission of disease: Foot and mouth disease during the 2001 epidemic in Great Britain. *Epidemics* 2, 109-115.
- Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E.J., 2010b. Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain. *International Journal of Infectious Diseases* 14, e210-e215.

- Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., 2013. The global distribution and burden of dengue. *Nature* 496, 504-507.
- Bivand, R., Keitt, T., Rowlingson, B., 2014. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-1 <http://CRAN.R-project.org/package=rgdal>.
- Bivand, R., Lewin-Koh, N., 2014. mapproj: Tools for reading and handling spatial objects. R package version 0.8-30 <http://CRAN.R-project.org/package=mapproj>.
- Bivand, R., Rundel, C., 2014. rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-8 <http://CRAN.R-project.org/package=rgeos>.
- Blake, A., Sinclair, M.T., Sugiyarto, G., 2003. Quantifying the impact of foot and mouth disease on tourism and the UK economy. *Tourism Economics* 9, 449-465.
- Bogh, C., Lindsay, S.W., Clarke, S.E., Dean, A., Jawara, M., Pinder, M., Thomas, C.J., 2007. High spatial resolution mapping of malaria transmission risk in the Gambia, west Africa, using LANDSAT TM satellite imagery. *The American Journal of Tropical Medicine and Hygiene* 76, 875-881.
- Boklund, A., Halasa, T., Christiansen, L.E., Enoe, C., 2013. Comparing control strategies against foot-and-mouth disease: will vaccination be cost-effective in Denmark? *Preventive Veterinary Medicine* 111, 206-219.
- Bortolot, Z.J., Prater, P.E., 2009. A first assessment of the use of high spatial resolution hyperspectral imagery in discriminating among animal species, and between animals and their surroundings. *Biosystems Engineering* 102, 379-384.
- Bouma, A., Elbers, A.R., Dekker, A., de Koeijer, A., Bartels, C., Vellema, P., van der Wal, P., van Rooij, E.M., Pluimers, F.H., de Jong, M.C., 2003. The foot-and-mouth disease epidemic in The Netherlands in 2001. *Preventive Veterinary Medicine* 57, 155-166.
- Bradhurst, R.A., 2015. Modelling the spatiotemporal spread and control of viral disease in livestock using a hybrid equation-based and agent-based approach.
- 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. *Frontiers in Environmental Science* 3, 1-20.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5-32.
- Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. *Monthly weather review* 78, 1-3.
- Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature* 511, 228-231.
- Brown, G., 2012. An empirical evaluation of the spatial accuracy of public participation GIS (PPGIS) data. *Applied Geography* 34, 289-294.
- Bruhn, M.C., Munoz, B., Cajka, J., Smith, G., Curry, R.J., Wagener, D.K., Wheaton, W.D., 2012. Synthesized Population Databases: A Geospatial Database of US Poultry Farms. Methods report (RTI Press), 1-24.
- Buermann, W., Saatchi, S., Smith, T.B., Zutta, B.R., Chaves, J.A., Milá, B., Graham, C.H., 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography* 35, 1160-1176.

- Buetre, B., Wicks, S., Kruger, H., Millist, N., Yainshet, A., Garner, G., Duncan, A., Abdalla, A., Trestrail, C., Hatt, M., 2013. Potential socio-economic impacts of an outbreak of foot-and-mouth disease in Australia. Australian Bureau of Agricultural and Resource Economics and Sciences. Canberra, ACT, Australia.
- 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-e91724.
- 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.
- 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. *Journal of Veterinary Diagnostic Investigation* 23, 26-33.
- 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. *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 16, 11-16.
- Cartín-Rojas, A., 2012. Transboundary animal diseases and international trade. INTECH Open Access Publisher.
- Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In, *Proceedings of the 23rd international conference on Machine learning*, 161-168.
- Chang, H.S., Hsia, C.J., Griffith, G., 2006. The FMD Outbreak in the Taiwan Pig Industry and the Demand for Beef Imports into Taiwan. *Australasian Agribusiness Review* 14.
- Chrétien, L.P., Théau, J., Ménard, P., 2016. Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system. *Wildlife Society Bulletin* 40, 181-191.
- Cogger, N., 2011. Modelling the spread of equine influenza in New Zealand.
- Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 37-46.
- Conn, P.B., Ver Hoef, J.M., McClintock, B.T., Moreland, E.E., London, J.M., Cameron, M.F., Dahle, S.P., Boveng, P.L., 2013. Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea. *Methods in Ecology and Evolution* 5, 1280-1293.
- Coops, N.C., Wulder, M.A., Iwanicka, D., 2009. Exploring the relative importance of satellite-derived descriptors of production, topography and land cover for predicting breeding bird species richness over Ontario, Canada. *Remote Sensing of Environment* 113, 668-679.
- Cox, L.A., Popken, D.A., VanSickle, J.J., Sahu, R., 2005. Optimal tracking and testing of U.S. and Canadian herds for BSE: A value-of-information (VOI) approach. *Risk Analysis* 25, 827-840.
- CSIRO, 2015. Measuring cattle body condition remotely. *Precision Livestock Management*. CSIRO.
- Cunningham, D.J., Anderson, W.H., Anthony, R.M., 1996. An image-processing program for automated counting. *Wildlife Society Bulletin* 24, 345-346.

- Cutler, D.R., 2007. Random Forests for Classification in Ecology. *Ecology* 88, 2783-2792.
- Davies, G., 2002. The foot and mouth disease (FMD) epidemic in the United Kingdom 2001. *Comparative Immunology, Microbiology and Infectious Diseases* 25, 331-343.
- De Gourville, E., Duintjer Tebbens, R.J., Sangrueje, N., Pallansch, M.A., Thompson, K.M., 2006. Global surveillance and the value of information: The case of the global polio laboratory network. *Risk Analysis* 26, 1557-1569.
- de Rueda, C.B., de Jong, M.C., Eble, P.L., Dekker, A., 2014. Estimation of the transmission of foot-and-mouth disease virus from infected sheep to cattle. *Veterinary Research* 45, 58-58.
- Descamps, S., Béchet, A., Descombes, X., Arnaud, A., Zerubia, J., 2011. An automatic counter for aerial images of aggregations of large birds. *Bird Study* 58, 302-308.
- Desclée, B., Bogaert, P., Defourny, P., 2006. Forest change detection by statistical object-based method. *Remote Sensing of Environment* 102, 1-11.
- Diggle, P., 1985. A kernel method for smoothing point process data. *Applied Statistics*, 138-147.
- Diggle, P., Ribeiro, P.J., 1998. Model-based Geostatistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 47, 299-350.
- Disney, W.T., Peters, M.A., 2003. Simulation modeling to derive the value-of-information for risky animal disease-import decisions. *Preventive Veterinary Medicine* 61, 171-184.
- Dohoo, I., Martin, W., Stryhn, H. (Eds.), 2003. *Veterinary Epidemiologic Research*. AVC Inc Charlottetown, Prince Edward Island, Canada.
- Dohoo, I., Martin, W., Stryhn, H., 2009. *Veterinary Epidemiologic Research*. VER Inc. Charlottetown, Prince Edward Island, Canada.
- Donaldson, A.I., Alexandersen, S., Sorensen, J.H., Mikkelsen, T., 2001. Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Veterinary Record* 148, 602-604.
- Dube, 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. *New Zealand Veterinary Journal* 55, 280-288.
- Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118, 259-272.
- Eames, K.T.D., Keeling, M.J., 2003. Contact tracing and disease control. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 2565-2571.
- Edirisinghe, A., Hill, M., Donald, G., Hyder, M., 2011. Quantitative mapping of pasture biomass using satellite imagery. *International Journal of Remote Sensing* 32, 2699-2724.
- Efron B, Tibshirani RJ. (1993), *An introduction to the bootstrap*. Chapman & Hall/CRC
- Elith, J., Graham, C.H., 2008. Reasons for differing performances of species distribution models. Australia Centre of Excellence for Risk Analysis (ACERA) Report. University of Melbourne, Melbourne, Australia.
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J.,

- Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802-813.
- Ellis-Iversen, J., Smith, R., Gibbens, J., Sharpe, C., Dominguez, M., Cook, A., 2011. Risk factors for transmission of foot-and-mouth disease during an outbreak in southern England in 2007. *Veterinary Record: Journal of the British Veterinary Association* 168.
- Emelyanova, I., Donald, G., Miron, D., Henry, D., Garner, M., 2009. Probabilistic modelling of cattle farm distribution in Australia. *Environmental Modeling & Assessment* 14, 449-465.
- Eurostat, 2013. Glossary: Livestock unit (LSU). Eurostat.
- 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.
- Firestone, S.M., Christley, R.M., Ward, M.P., Dhand, N.K., 2012. Adding the spatial dimension to the social network analysis of an epidemic: Investigation of the 2007 outbreak of equine influenza in Australia. *Preventive Veterinary Medicine* 106, 123-135.
- Flood, J.S., Porphyre, T., Tildesley, M.J., Woolhouse, M.E.J., 2013. The performance of approximations of farm contiguity compared to contiguity defined using detailed geographical information in two sample areas in Scotland: implications for foot-and-mouth disease modelling. *BMC Veterinary Research* 9, 198-198.
- Forbes, R., van Halderen, A., 2014. Foot-and-mouth disease economic impact assessment: What it means for New Zealand. Ministry for Primary Industries (MPI), Wellington, NZ.
- Forsberg, O.I., Guttormsen, A.G., 2006. The Value of Information in Salmon Farming. Harvesting the Right Fish At the Right Time. *Aquaculture Economics & Management* 10, 183-200.
- Franceschini, G., Robinson, T.P., Morteo, K., Dentale, D., Wint, W., Otte, J., 2009. The Global Livestock Impact Mapping System (GLIMS) as a tool for animal health applications. *Veterinaria Italiana* 45, 491-499.
- Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B., Ratcliffe, N., Fox, A.J., Fleming, A.H., Porter, C., Trathan, P.N., 2012. An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PloS One* 7, e33751.
- Fretwell, P.T., Staniland, I.J., Forcada, J., 2014. Whales from space: counting southern right whales by satellite. *PloS One* 9, e88655.
- 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. *Veterinary Research* 45, 110-110.
- Geering, W. a, Lubroth, J., 2002. Prepartation of Foot-and-mouth disease contingency plans, FAO Animal Health Manual. FAO, Rome.
- Gibbens, J.C., Wilesmith, J.W., 2002. Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain. *The Veterinary Record* 151, 407-412.
- Gilbert, M., Golding, N., Zhou, H., Wint, G.R.W., Robinson, T.P., Tatem, A.J., Lai, S., Zhou, S., Jiang, H., Guo, D., Huang, Z., Messina, J.P., Xiao, X., Linard, C., Van Boeckel, T.P., Martin,

- V., Bhatt, S., Gething, P.W., Farrar, J.J., Hay, S.I., Yu, H., 2014. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat Communications* 5.
- Goetz, S., Prince, S., Small, J., 2000. Advances in satellite remote sensing of environmental variables for epidemiological applications. *Advances in Parasitology* 47, 289-307.
- Groom, G., Krag Petersen, I., Anderson, M.D., Fox, A.D., 2011. Using object-based analysis of image data to count birds: mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa. *International Journal of Remote Sensing* 32, 4611-4639.
- Groom, G., Stjernholm, M., Nielsen, R.D., Fleetwood, A., Petersen, I.K., 2013. Remote sensing image data and automated analysis to describe marine bird distributions and abundances. *Ecological Informatics* 14, 2-8.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993-1009.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-186.
- 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 Management* 31, 767-778.
- Hagerman, A.D., McCarl, B.A., Carpenter, T.E., Ward, M.P., O'Brien, J., 2012. Emergency Vaccination to Control Foot-and-mouth Disease: Implications of its Inclusion as a U.S. Policy Option. *Applied Economic Perspectives and Policy* 34, 119-146.
- 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. *Preventive Veterinary Medicine* 112, 194-202.
- Hannay, D., Jones, R., 2002. The effects of foot-and-mouth on the health of those involved in farming and tourism in Dumfries and Galloway. *The European Journal of General Practice* 8, 83-89.
- Hansen, G.J.a., Jones, M.L., 2008. The Value of Information in Fishery Management. *Fisheries* 33, 340-348.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: data mining, inference and prediction*. Second Edition. Springer-Verlag New York.
- Hay, S.I., Snow, R.W., Rogers, D.J., 1998. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 92, 12-20.
- 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. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 121-127.
- Haydon, D.T., Kao, R.R., Kitching, R.P., 2004. The UK foot-and-mouth disease outbreak - the aftermath. *Nature Reviews Microbiology* 2, 675-681.
- Hess, G.D., Garner, M.G., Yang, X., 2008. A Sensitivity Analysis of an Integrated Modelling Approach to Assess the Risk of Wind-borne Spread of Foot-and-mouth Disease Virus from Infected Premises. *Environmental Modeling & Assessment* 13, 209-220.

- Hickey, K., 2014. Foot-and-Mouth Disease Vaccination Policy, Evidence Based Discussion Document. Ministry for Primary Industries, Wellington, NZ.
- Hijmans, R.J., 2014a. geosphere: Spherical Trigonometry. R package version 1.3-11 <http://CRAN.R-project.org/package=geosphere>.
- Hijmans, R.J., 2014b. raster: raster: Geographic data analysis and modeling. R package version 2.3-12 <http://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2014. dismo: Species distribution modeling. R package version 1.0-5. <http://CRAN.R-project.org/package=dismo>.
- Holm, S., 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scandanavian Journal of Statistics 6, 65-70.
- 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. Society for Veterinary Epidemiology and Preventive Medicine, 152-163.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195-213.
- Hugh-Jones, M.E., 1972. Epidemiological studies on the 1967-1968 Foot and Mouth Epidemic: Attack Rates and Cattle Density. Research in Veterinary Science 13, 411-417.
- Hunter, M., 2001. Public health concerns grow over foot and mouth outbreak. British Medical Journal 322, 881.
- Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS Journal of Photogrammetry and Remote Sensing 80, 91-106.
- Jachmann, H., 2002. Comparison of aerial counts with ground counts for large African herbivores. Journal of Applied Ecology 39, 841-852.
- Jansen, V.S., Kolden, C.A., Taylor, R.V., A. Newingham, B., 2016. Quantifying livestock effects on bunchgrass vegetation with Landsat ETM+ data across a single growing season. International Journal of Remote Sensing 37, 150-175.
- Jensen, J.R., Cowen, D.C., 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. Photogrammetric Engineering and Remote Sensing 65, 611-622.
- 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., van Andel, M., Vink, W., McFadden, A., 2015. Compatibility Between Livestock Databases Used for Quantitative Biosecurity Response in New Zealand. New Zealand Veterinary Journal VJ, 1-17.
- Jewell, C.P., 2016a. Bayesian ZIP model with covariates on the zero probability. Lancaster.
- Jewell, C.P., 2016b. Variogram. Lancaster.
- Jewell, C.P., Andel, M.V., Vink, W.D., McFadden, A.M.J., 2016. Compatibility between livestock databases used for quantitative biosecurity response in New Zealand. New Zealand Veterinary Journal 64, 158-164.

- Jewell, C.P., Brown, R.G., 2015. Bayesian data assimilation provides rapid decision support for vector-borne diseases. *Journal of The Royal Society Interface* 12, 20150367.
- Jewell, C.P., Keeling, M.J., Roberts, G.O., 2009a. Predicting undetected infections during the 2007 foot-and-mouth disease outbreak. *Journal of the Royal Society InterFace* 6, 1145-1151.
- Jewell, C.P., Kypraios, T., Neal, P., Roberts, G.O., 2009b. Bayesian Analysis for Emerging Infectious Diseases. *Bayesian Analysis* 4, 465-496.
- Jin, X., Davis, C.H., 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Applied Signal Processing* 2005, 2196-2206.
- 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. *International Journal of Environmental Research and Public Health* 10, 7126-7143.
- Kadlec, J.A., Drury, W.H., 1968. Aerial estimation of the size of gull breeding colonies. *The Journal of Wildlife Management* 32, 287-293.
- Kalluri, S., Gilruth, P., Rogers, D., Szczur, M., 2007. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. *PLoS Pathog* 3, 1361-1371.
- Kao, R.R., 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends in microbiology* 10, 279-286.
- Kawamura, K., Akiyama, T., Yokota, H.-o., Tsutsumi, M., Yasuda, T., Watanabe, O., Wang, S., 2005. Quantifying grazing intensities using geographic information systems and satellite remote sensing in the Xilingol steppe region, Inner Mongolia, China. *Agriculture, Ecosystems & Environment* 107, 83-93.
- Ke, Y., Quackenbush, L.J., Im, J., 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sensing of Environment* 114, 1141-1154.
- Keeling, M.J., 2005. Models of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences* 272, 1195-1202.
- Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappay, 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.
- Keisler, J.M., Collier, Z.A., Chu, E., Sinatra, N., Linkov, I., 2014. Value of information analysis: The state of application. *Environment Systems and Decisions* 34, 3-23.
- Kitching, R.P., Thrusfield, M., Taylor, N.M., 2006. Use and abuse of mathematical models: an illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Link: Link to publication record in Edinburgh Research Explorer Published In. Revue scientifique et technique (International Office of Epizootics)* 25, 293-311.
- Klain, T.J., Lusk, J.L., Tonsor, G.T., Schroeder, T.C., 2014. An experimental approach to valuing information. *Agricultural Economics (United Kingdom)* 45, 635-648.
- Koller, R., Samimi, C., 2011. Deforestation in the Miombo woodlands: a pixel-based semi-automated change detection method. *International Journal of Remote Sensing* 32, 7631-7649.

- Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., 2013. Challenges and opportunities in mapping land use intensity globally. *Current Opinion in Environmental Sustainability* 5, 484-493.
- Laliberte, A.S., Ripple, W.J., 2003. Automated wildlife counts from remotely sensed imagery. *Wildlife Society Bulletin* 31, 362-371.
- Landcare Research, 2015. LCDB v3.0 Accuracy Assessment. Landcare Research, Wellington, New Zealand.
- Landis, J., Koch, G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159-174.
- LaRue, M., Lynch, H., Lyver, P., Barton, K., Ainley, D., Pollard, A., Fraser, W., Ballard, G., 2014. A method for estimating colony sizes of Adélie penguins using remote sensing imagery. *Polar Biology* 37, 507-517.
- LaRue, M.A., Rotella, J.J., Garrott, R.A., Siniff, D.B., Ainley, D.G., Stauffer, G.E., Porter, C.C., Morin, P.J., 2011. Satellite imagery can be used to detect variation in abundance of Weddell seals (*Leptonychotes weddellii*) in Erebus Bay, Antarctica. *Polar Biology* 34, 1727-1737.
- LaRue, M.A., Stapleton, S., Porter, C., Atkinson, S., Atwood, T., Dyck, M., Lecomte, N., 2015. Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution. *Wildlife Society Bulletin* 39, 772-779.
- Lawrence, K., McFadden, A., Gias, E., Pulford, D., Pomroy, W., 2016. Epidemiology of the epidemic of bovine anaemia associated with *Theileria orientalis* (Ikeda) between August 2012 and March 2014. *New Zealand Veterinary Journal* 64, 38-47.
- Leathwick, J., McGlone, M., Walker, S., 2012. NZ's Potential Vegetation Pattern. Department of Conservation, Wellington, NZ.
- Leathwick, J., Morgan, F., Wilson, G., Rutledge, D., McLeod, M., Johnston, K., 2002. Land Environments of NZ: A Technical Guide. Ministry for the Environment, Wellington, NZ.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar Remote Sensing for Ecosystem Studies. *BioScience* 52, 19-30.
- Leonard, R.M., Fish, E.B., 1974. An aerial photographic technique for censusing lesser sandhill cranes. *Wildlife Society Bulletin* 2, 191-195.
- Levine, J.M., D'Antonio, C.M., 2003. Forecasting biological invasions with increasing international trade. *Conservation Biology* 17, 322-326.
- Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkönig, I., Skidmore, A., 2007. Capturing the fugitive: Applying remote sensing to terrestrial animal distribution and diversity. *International Journal of Applied Earth Observation and Geoinformation* 9, 1-20.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2(3), 18-22.
- Lindstrom, T., Lewerin, S.S., Wennergren, U., 2012. Influence on disease spread dynamics of herd characteristics in a structured livestock industry. *Journal of The Royal Society Interface* 9, 1287-1294.
- Linthicum, K. J., Anyamba, A., Tucker, C. J., Kelley, P. W., Myers, M. F., & Peters, C. J. (1999). Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science*, 285(5426), 397-400.
- Löffler, E., Margules, C., 1980. Wombats detected from space. *Remote Sensing of Environment* 9, 47-56.

- Lynch, H., LaRue, M., 2014. First global census of the Adélie Penguin. *The Auk* 131, 457-466.
- Lynch, H.J., White, R., Black, A.D., Naveen, R., 2012. Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar biology* 35, 963-968.
- Lysons, R., Gibbens, J., Smith, L., 2007. Progress with enhancing veterinary surveillance in the United Kingdom. *Veterinary Record: Journal of the British Veterinary Association* 160.
- Mackereth, G.F., 1998. Spatial data requirements for animal disease management in New Zealand. Massey University, Palmerston North, NZ, 1-82.
- MacPherson, L., 2015. Agricultural Production Statistics June 2014. Statistics New Zealand, Wellington, New Zealand.
- Makhlouf, G., 2015. Guide to Social Cost Benefit Analysis. The Treasury, Wellington, NZ.
- Mansley, L.M., 2004. The challenge of FMD control in the 2001 UK FMD epidemic. In: Madec, F. (Ed.), *International Society for Animal Hygiène*, Saint-Malo, 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. *Wildlife Society Bulletin* 34, 609-619.
- Mathieu, R., Freeman, C., Aryal, J., 2007. Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. *Landscape and Urban Planning* 81, 179-192.
- McClintock, B.T., Moreland, E.E., London, J.M., Dahle, S.P., Brady, G.M., Richmond, E.L., Yano, K.M., Boveng, P.L., 2015. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. *Marine Mammal Science* 31, 1057-1076.
- McMahon, C.R., Howe, H., van den Hoff, J., Alderman, R., Brotsma, H., Hindell, M.A., 2014. Satellites, the All-Seeing Eyes in the Sky: Counting Elephant Seals from Space. *PloS One* 9, e92613.
- McNeill, S., Barton, K., Lyver, P., Pairman, D., 2011. Semi-automated penguin counting from digital aerial photographs. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 4312-4315.
- Meinshausen, N., 2006. Quantile regression forests. *The Journal of Machine Learning Research* 7, 983-999.
- Mejias, L., Duclos, G., Hodgson, A., Maire, F., 2013. Automated marine mammal detection from aerial imagery. In, *2013 OCEANS - San Diego*, San Diego, CA, 2013, 1-5.
- Métrás, R., Jewell, C., Porphyre, T., Thompson, P.N., Pfeiffer, D.U., Collins, L.M., White, R.G., 2015. Risk factors associated with Rift Valley fever epidemics in South Africa in 2008–11. *Scientific Reports* 5, 9492-9492.
- Meyerson, L.A., Reaser, J.K., 2002. Biosecurity: Moving toward a Comprehensive Approach *BioScience* 52, 593-600.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Weir, L.A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* 92, 1422-1428.
- Ministry for Primary Industries, 2012. Exercise Taurus 2012 Final Report. Wellington, NZ.

- Ministry for Primary Industries, 2013. Preparing for and responding to biosecurity incursions. Wellington, NZ.
- Ministry for the Environment, 2012. Land-Use and Carbon Analysis System: Satellite imagery interpretation guide for land-use classes (2nd edition). In: Environment, M.f.t. (Ed.), Wellington.
- Moore, A., Vickery, P., Hill, M., Donnelly, J., Donald, G., 1999. Combining satellite data with a simulation model to describe spatial variability in pasture growth at a farm scale. *Animal Production Science* 39, 285-300.
- Moore, J.L., Runge, M.C., 2012. Combining Structured Decision Making and Value-of-Information Analyses to Identify Robust Management Strategies. *Conservation Biology* 26, 810-820.
- Morris, R.S., Wilesmith, J.W., Stern, M.W., Sanson, R.L., Stevenson, M.A., 2001. Predictive spatial modeling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Veterinary Record* 149, 137-144.
- 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. *Journal of Applied Ecology* 45, 649-658.
- Mullahy, J., 1986. Specification and testing of some modified count data models. *Journal of Econometrics* 33, 341-365.
- Muroga, N., Hayama, Y., Yamamoto, T., Kurogi, A., Tsuda, T., Tsutsui, T., 2012. The 2010 foot-and-mouth disease epidemic in Japan. *Journal of Veterinary Medical Science* 74, 399-404.
- 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 Veterinary Research* 9, 1-9.
- Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indexes. *IEEE Transactions of Geoscience and Remote Sensing* 33, 481-486.
- Neumann, K., Elbersen, B.S., Verburg, P.H., Staritsky, I., Pérez-Soba, M., de Vries, W., Rienks, W.A., 2009. Modelling the spatial distribution of livestock in Europe. *Landscape Ecology* 24, 1207-1222.
- Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2008. Land Resource Information System Spatial Data Layers: Data Dictionary. Landcare Research NZ Ltd, Palmerston North, NZ.
- Nicolas, G., TP Robinson, G.R.W. Wint, G. Conchedda, G. Cinardi, Gilbert, M., 2016. Using random forests to improve the downscaling of global livestock census data. *PLoS ONE* In Review.
- Nishiura, H., Omori, R., 2010. An epidemiological analysis of the foot-and-mouth disease epidemic in Miyazaki, Japan, 2010. *Transboundary and Emerging Diseases* 57, 396-403.
- Nosrati, M.S., Saeedi, P., 2009. A novel approach for polygonal rooftop detection in satellite/aerial imageries. In, 16th IEEE International Conference on Image Processing (ICIP), 1709-1712.
- Oishi, Y., Matsunaga, T., 2014. Support system for surveying moving wild animals in the snow using aerial remote-sensing images. *International Journal of Remote Sensing* 35, 1374-1394.
- Olf, M., Koeter, M.W., Van Haaften, E.H., Kersten, P.H., Gersons, B.P., 2005. Impact of a foot and mouth disease crisis on post-traumatic stress symptoms in farmers. *The British Journal of Psychiatry* 186, 165-166.

- Otte, M., Nugent, R., McLeod, A., 2004. Transboundary Animal Diseases : Assessment of socio-economic impacts. Livestock Policy Discussion Paper No. 9. Food and Agriculture Organisation, Rome, Italy, 1-46.
- Owen, K., Stevenson, M.A., Sanson, R.L., 2011. A sensitivity analysis of the New Zealand standard model of foot and mouth disease. *Revue Scientifique et Technique - Office International des Epizooties* 30, 513-526.
- 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. *Preventive veterinary medicine; Lead papers of the eleventh symposium of the International Society for Veterinary Epidemiology and Economics (ISVEE)*, Cairns, Australia 81, 117-134.
- 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. *Euro Surveill* 15, 1-9.
- Papadopoulou, C.-A., Giaoutzi, M., 2014. Crowdsourcing as a Tool for Knowledge Acquisition in Spatial Planning. *Future Internet* 6, 109-125.
- Paudel, K.P., Andersen, P., 2010. Assessing rangeland degradation using multi temporal satellite images and grazing pressure surface model in Upper Mustang, Trans Himalaya, Nepal. *Remote Sensing of Environment* 114, 1845-1855.
- Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. *R News* 5 (2) <http://cran.r-project.org/doc/Rnews/>.
- Peck, D.F., 2005. Foot and mouth outbreak: lessons for mental health services. *Advances in Psychiatric Treatment* 11, 270-276.
- Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E., Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment* 115, 1301-1316.
- Peña, J.M., Gutiérrez, P.A., Hervás-Martínez, C., Six, J., Plant, R.E., López-Granados, F., 2014. Object-Based Image Classification of Summer Crops with Machine Learning Methods. *Remote Sensing* 6, 5019-5041.
- Petráš, V., 2012. Building detection from aerial images in GRASS GIS environment. Masters Thesis. Faculty of Civil Engineering, Department of Mapping and Cartography. Czech Technical University Prague.
- Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology* 51, 839-848.
- Pharo, H.J., 2002. Foot-and-mouth disease: an assessment of the risks facing New Zealand. *N. Z. Vet. J.* 50, 46–55. doi:10.1080/00480169.2002.36250
- 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, 486-492.

- Porphyre, T., Auty, H.K., Tildesley, M.J., Gunn, G.J., Woolhouse, M.E.J., 2013. Vaccination against Foot-And-Mouth Disease: Do Initial Conditions Affect Its Benefit? *PLoS ONE* 8, e77616-e77616.
- Power, A.P., Harris, S.A., 1973. A cost-benefit evaluation of alternative control policies for foot-and-mouth disease in Great Britain. *Journal of Agricultural Economics* 24, 573-597.
- Probert, W.J.M., Shea, K., Fonnesebeck, C.J., Runge, M.C., Carpenter, T.E., Dürr, S., Garner, M.G., Harvey, N., Stevenson, M.A., Webb, C.T., Werkman, M., Tildesley, M.J., Ferrari, M.J., 2015. Decision-making for foot-and-mouth disease control: Objectives matter. *Epidemics* 15, 1-11.
- Prosser, D.J., Wu, J., Ellis, E.C., Gale, F., Van Boeckel, T.P., Wint, W., Robinson, T., Xiao, X., Gilbert, M., 2011. Modelling the distribution of chickens, ducks, and geese in China. *Agriculture, Ecosystems & Environment* 141, 381-389.
- Quantum GIS Development Team, 2016. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.
- R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL <http://www.R-project.org/>.
- Rainsford, D., Mackaness, W., 2002. Template matching in support of generalisation of rural buildings. In: Richardson, D.E., van Oosterom, P. (Eds.), *Advances in Spatial Data Handling. 10th International Symposium on Spatial Data Handling*. Springer Berlin Heidelberg, 137-151.
- Recio, M.R., Mathieu, R., Hall, G.B., Moore, A.B., Seddon, P.J., 2013. Landscape resource mapping for wildlife research using very high resolution satellite imagery. *Methods in Ecology and Evolution* 4, 982-992.
- 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. *Journal of Vegetation Science* 5, 703-714.
- Ridgeway, G., 2013. gbm: Generalized Boosted Regression Models. R package version 2.1. <http://CRAN.R-project.org/package=gbm>.
- Ripley, B.D., 1977. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society (B)* 39, 172 - 212.
- Ripley, B.D., 1988. *Statistical inference for spatial processes*. Cambridge University Press.
- Robinson, S.E., Everett, M.G., Christley, R.M., 2007a. Recent network evolution increases the potential for large epidemics in the British cattle population. *Journal of the Royal Society InterFace* 4, 669-674.
- Robinson, T.P., Franceschini, G., Wint, W., 2007b. The Food and Agriculture Organization's gridded livestock of the world. *Veterinaria Italiana* 43, 745-751.
- Robinson, T.P., Wint, G.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aiotti, 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., Dube, C., Patyk, K.A., Stevenson, M.A., Yu, Z.D., Rawdon, T.G., Gauntlett, 2015. Evaluating vaccination strategies to control foot-and-mouth disease: a model comparison study. *Epidemiology and Infection* 143, 1256-1275.

- Rottensteiner, F., 2003. Automatic generation of high-quality building models from lidar data. *IEEE Computer Graphics and Applications* 23, 42-50.
- Sanson, R., 1993. The development of a decision support system for an animal disease emergency. PhD Thesis. Massey University, Palmerston North, New Zealand.
- Sanson, R., 1994. The epidemiology of foot-and-mouth disease: implications for New Zealand. *New Zealand Veterinary Journal* 42, 41-53.
- Sanson, R., Pearson, A., 1997. Agribase - a national spatial farm database. *Epidemiologie et Sante Animale*, 31-32.
- 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. In, *ISVEE13: Proceedings of the 13th International Symposium on Veterinary Epidemiology and Economics*, Belgium, Netherlands.
- Sanson, R.L., Harvey, N., Garner, M.G., Stevenson, M.A., 2011. Foot and mouth disease model verification and 'relative validation'. *Revue Scientifique Et Technique De L'Office International Des Epizooties* 30, 527-540.
- Sanson, R.L., Stevenson, M.A., Mackereth, G.F., Moles-Benfell, N., 2006a. The development of an interspread plus parameter set to simulate the spread of FMD in New Zealand. In, *International Symposium on Veterinary Epidemiology and Economics*, 682-682.
- Sanson, R.L., Stevenson, M.A., Moles-Benfell, N., 2006b. T4-2.3.1 - Quantifying local spread probabilities for foot-and-mouth disease. In, *International Symposium on Veterinary Epidemiology and Economics*, Cairns, Australia.
- Sarandopoulos, J., 2015. Early Predictors of the Size and Duration of Foot-and-Mouth Disease Epidemics.
- 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. *Journal of The Royal Society Interface* 4, 235-241.
- Sbalzarini, I.F., Koumoutsakos, P., 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology* 151, 182-195.
- Schley D, Gubbins S, Paton DJ. (2009), Quantifying the risk of localised animal movement bans for foot-and-mouth disease. *Plos One* 4, e5481, doi:10.1371/journal.pone.0005481
- Scudamore, J.M., 2007. Consumer attitudes to vaccination of food-producing animals. *Revue Scientifique et Technique (International Office of Epizootics)* 26, 451-459.
- Segurado, P., Araujo, M.B., 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31, 1555-1568.
- Seto, K., Fleishman, E., Fay, J., Betrus, C., 2004. Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing* 25, 4309-4324.
- Shea, K., Tildesley, M.J., Runge, M.C., Fonnesebeck, C.J., Ferrari, M.J., 2014. Adaptive Management and the Value of Information: Learning Via Intervention in Epidemiology. *PLoS Biology* 12, 9-12.

- Sheeren, D., Bastin, N., Ouin, A., Ladet, S., Balent, G., Lacombe, J.-P., 2009. Discriminating small wooded elements in rural landscape from aerial photography: a hybrid pixel/object-based analysis approach. *International Journal of Remote Sensing* 30, 4979-4990.
- Shekhar, S., Chawla, S., 2003. *Spatial databases: a tour*. Prentice Hall Upper Saddle River, NJ, USA.
- Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10, 989-1003.
- Slater, M.R., 2001. The role of veterinary epidemiology in the study of free-roaming dogs and cats. *Plenary Sessions of the International Society of Veterinary* 48, 273-286.
- Sorensen, J.H., Mackay, D.K., Jensen, C.O., Donaldson, A.I., 2000. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. *Epidemiology and Infection* 124, 577-590.
- Stage, A.R., 1976. An expression for the effect of aspect, slope, and habitat type on tree growth. *Forest Science* 22, 457-460.
- Stage, A.R., Salas, C., 2007. Interactions of elevation, aspect, and slope in models of forest species composition and productivity. *Forest Science* 53, 486-492.
- Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C., Atwood, T., 2014. Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. *PLoS One* 9, e101513.
- Ster, I.C., Dodd, P.J., Ferguson, N.M., 2012. Within-farm transmission dynamics of foot and mouth disease as revealed by the 2001 epidemic in Great Britain. *Epidemics* 4, 158-169.
- Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS one* 10, e0107042.
- Stevens, K.B., Gilbert, M., Pfeiffer, D.U., 2013. Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia: a spatial multicriteria decision analysis approach. *Spatial and Spatio-temporal Epidemiology* 4, 1-14.
- Stevenson, M.A., 2003. *The Spatio-temporal Epidemiology of Bovine Spongiform Encephalopathy and Foot-and-Mouth Disease in Great Britain*. PhD Thesis. . Massey University, Palmerston North, NZ, 249-249.
- Stevenson, M.A., Sanson, R.L., Miranda, A.O., Lawrence, K.A., Morris, R.S., 2007. Decision support systems for monitoring and maintaining health in food animal populations. *New Zealand Veterinary Journal* 55, 264-272.
- Stevenson, M.A., Sanson, R.L., Stern, M.W., O'Leary, B.D., Sujau, M., Moles-Benfell, N., Morris, R.S., Sanson, A.L., Stern, M.W., O'Leary, B.D., Sujau, M., Moles-Benfell, M., Morris, R.S., 2013. InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. *Preventative Veterinary Medicine* 109, 10-24.
- Streutker, D.R., Glenn, N.F., 2006. LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment* 102, 135-145.
- Strong, D.M., Lee, Y.W., Wang, R.Y., 1997. Data quality in context. *Communications of the ACM* 40, 103-110.
- Strong, L.L., Gilmer, D.S., Brass, J.A., 1991. Inventory of wintering geese with a multispectral scanner. *The Journal of Wildlife Management* 55, 250-259.

- Sumption, K., Rweyemamu, M., Wint, W., 2008. Incidence and Distribution of Foot-and-Mouth Disease in Asia, Africa and South America; Combining Expert Opinion, Official Disease Information and Livestock Populations to Assist Risk Assessment. *Transboundary and Emerging Diseases* 55, 5-13.
- Taylor, B.F., Dini, P., Kidson, J., 1985. Determination of seasonal and interannual variation in New Zealand pasture growth from NOAA-7 data. *Remote sensing of environment* 18, 177-192.
- Temme, A., Verburg, P., 2011. Mapping and modelling of changes in agricultural intensity in Europe. *Agriculture, Ecosystems & Environment* 140, 46-56.
- Terletzky, P., Ramsey, R.D., 2014. A Semi-Automated Single Day Image Differencing Technique to Identify Animals in Aerial Imagery. *PloS One* 9, e85239.
- Terletzky, P., Ramsey, R.D., Neale, C.M., 2012. Spectral Characteristics of Domestic and Wild Mammals. *GIScience & Remote Sensing* 49, 597-608.
- Terletzky, P.A., 2013. Utilizing Remote Sensing and Geospatial Techniques to Determine Detection Probabilities of Large Mammals. PhD Dissertation. Wildland Resources. Utah State University.
- The Scottish Government, 2014. Information of Livestock Units. The Scottish Government.
- Thompson, D., Muriel, P., Russell, D., Osborne, P., Bromley, A., Rowland, M., Creigh-Tyte, S., Brown, C., 2002. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Revue Scientifique et Technique-Office International des Epizooties* 21, 675-685.
- Thrusfield, M., 2007. *Veterinary Epidemiology*. Blackwell Science Oxford, UK.
- 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. *The Veterinary Record* 156, 229-252.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.
- Tildesley, M., Savill, N., Shaw, D., Deardon, R., Brooks, S., Woolhouse, M., Grenfell M, B., Keeling, 2006. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* 440, 83-86.
- 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. *Proceedings of the Royal Society B: Biological Sciences* 275, 1459-1468.
- Tildesley, M.J., House, T.A., Bruhn, M.C., Curry, R.J., O'Neil, M., Allpress, J.L.E., Smith, G., Keeling, M.J., 2010. Impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences* 107, 1041-1046.
- Tildesley, M.J., Ryan, S.J., 2012. Disease Prevention versus Data Privacy: Using Landcover Maps to Inform Spatial Epidemic Models. *PLoS Computational Biology* 8, 1-13.
- 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. *Preventive Veterinary Medicine* 54, 301-324.
- Trathan, P.N., 2004. Image analysis of color aerial photography to estimate penguin population size. *Wildlife Society Bulletin* 32, 332-343.

- Trivedi, M.M., Wyatt, C.L., Anderson, D.R., 1982. A multispectral approach to remote detection of deer. *Photogrammetric Engineering and Remote Sensing* 48, 1879-1889.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and distributions* 13, 397-405.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* 18, 306-314.
- 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.
- Van Boeckel, T.P., Prosser, D., Franceschini, G., Biradar, C., Wint, W., Robinson, T., Gilbert, M., 2011. Modelling the distribution of domestic ducks in Monsoon Asia. *Agriculture, Ecosystems & Environment* 141, 373-380.
- Van Boeckel, T.P., Thanapongtharm, W., Robinson, T., D'Aiotti, L., Gilbert, M., 2012. Predicting the distribution of intensive poultry farming in Thailand. *Agriculture, Ecosystems & Environment* 149, 144-153.
- Van Haaften, E.H., Olf, M., Kersten, P.H., 2004. The psychological impact of the Foot and Mouth Disease crisis on Dutch dairy farmers. *NJAS - Wageningen Journal of Life Sciences* 51, 339-349.
- Velasco, M., 2009. A Quickbird's eye view on marmots. Masters Thesis. International Institute for Geo-information science and Earth Observation, Enschede, The Netherlands.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.
- Verbesselt, J., Robinson, A., Stone, C., Culvenor, D., 2009. Forecasting tree mortality using change metrics derived from MODIS satellite data. *Forest Ecology and Management* 258, 1166-1173.
- Vickery, P., Hill, M., Donald, G., 1997. Satellite derived maps of pasture growth status: association of classification with botanical composition. *Animal Production Science* 37, 547-562.
- Volkova, V., Bessell, P., Woolhouse, M., Savill, N., 2011. Evaluation of risks of foot-and-mouth disease in Scotland to assist with decision making during the 2007 outbreak in the UK. *The Veterinary Record* 169, 124.
- Volkova, V.V., Howey, R., Savill, N.J., Woolhouse, M.E.J., 2010. Sheep Movement Networks and the Transmission of Infectious Diseases. *PLoS ONE* 5, e11185-e11185.
- Waage, J., Mumford, J., 2008. Agricultural biosecurity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 863-876.
- Wada, M., Stevenson, M., Cogger, N., Carpenter, T., 2016. Evaluation of the Control Strategy for the 2010 Foot-and-Mouth Disease Outbreak in Japan Using Disease Simulation. *Transboundary and Emerging Diseases*.
- West, M., 2014. *Bayesian Forecasting*. John Wiley & Sons, Ltd.
- White, D.H., Lubulwa, G.A., Menz, K., Zuo, H., Wint, W., Slingenbergh, J., 2001. Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities. *Environment International* 27, 181-187.

- White, G.C., 2005. Correcting wildlife counts using detection probabilities. *Wildlife Research* 32, 211-216.
- Wilesmith, J.W., Stevenson, M.A., King, C.B., Morris, R.S., 2003. Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. *Preventative Veterinary Medicine* 61, 157-170.
- Williams, B.K., Eaton, M.J., Breininger, D.R., 2011. Adaptive resource management and the value of information. *Ecological Modelling* 222, 3429-3436.
- Woolhouse, M.E.J., 2003. Foot-and-mouth disease in the UK: What should we do next time? *Journal of Applied Microbiology* 94, 126-130.
- World Organisation for Animal Health, 2013. World Organisation for Animal Health (OIE) Terrestrial Animal Health Code.
- Wyatt, C.L., Anderson, D., Harshbarger, R., Trivedi, M., 1985. Deer census using a multispectral linear array instrument. In, *Proceedings of the 18th International Symposium on Remote Sensing of Environment*, Paris, France, 1475-1488.
- Yang, P.C., Chu, R.M., Chung, W.B., Sung, H.T., 1999. Epidemiological characteristics and financial costs of the 1997 foot-and-mouth disease epidemic in Taiwan. *The Veterinary Record* 145, 731-734.
- Yang, Z., 2012. Evaluating high resolution GeoEye-1 satellite imagery for mapping wildlife in open savannahs. PhD Thesis. Enschede, The Netherlands University of Twente.
- Yang, Z., Wang, T., Skidmore, A.K., de Leeuw, J., Said, M.Y., Freer, J., 2014. Spotting East African mammals in open savannah from Space. *PloS One* 9, e115989.
- Yokota, F., Thompson, K.M., 2004. Value of information literature analysis: a review of applications in health risk management. *Medical Decision Making* 24, 287-298.
- Zepeda, C., Salman, M., Ruppanner, R., 2001. International trade, animal health and veterinary epidemiology: challenges and opportunities. *Plenary Sessions of the International Society of Veterinary* 48, 261-271.
- Zhang, K., Yan, J., Chen, S.-C., 2006. Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing* 44, 2523-2533.
- Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39, 561-577.

Appendices

Appendix 1: Predictor variables – Summary of data and data sources

GIS data for NZ are available from Land and Information NZ (LINZ) and Landcare research's Land Resource Information Systems (LRIS). Most of these data are at the national level, updated regularly and free to download. An overview and summary of the available GIS datasets and livestock data relevant to this study is outlined below including information on resolution and accuracy. Datasets are available for buildings, fences, stockyards, water troughs, shelterbelts and cattle stops which are all indicators of livestock presence. These layers are incomplete and would currently be ineffective explanatory variables in any models. As they improve over time they should be reconsidered as model predictors.

Predictor variables GIS

Below are descriptions of the GIS data from various NZ government sources which are applicable to national-level livestock estimation and used as predictor variables. The information is sourced from the supporting documentation provided with each layer unless otherwise referenced.

Digital Elevation Model (North Island and South Island)

Data are a 25 metre resolution elevation grid available from the LINZ website and generated from the LINZ 1:50,000 topographic data layers (20m contours, spot heights, lake shorelines and coastline) using Landcare Research's interpolation software. The projection of the datum is WGS 84 (DEM supporting material). Downloaded as 'nzdem-north-island-25-metre (AAIGrid)' and 'nzdem-south-island-25-metre (AAIGrid).'

NZ Land Cover Database (LCDB) Version 4

The LCDB is derived from satellite imagery to classify land cover into 33 target classes across the whole of NZ's surface area. The vector polygon layer has a minimum mapping unit (MMU) of 1ha, in line with earlier versions of LCDB to enable land cover change analysis. Classification is tiered, starting with seven first order classes referencing broad vegetation type such as forest, shrubland,

grassland, and with increasingly more detailed classes at lower levels. Target class descriptions are available in the layer documentation. It is updated 5 yearly and the most recent version was released in June 2014 using imagery from 2011/2012. The accuracy of LCDB3 is just over 96% (LCDB website). Downloaded at <https://iris.scinfo.org.nz> as 'lcdb-v40-land-cover-database-version-40 (ESRI Shapefile).' Categories relevant to livestock are:

- *High Producing Exotic Grassland*: Often intensively managed grazing for sheep, beef, dairy and deer. Areas of high productivity due to high soil fertility and annual rainfall or by from the use of fertiliser and irrigation.
- *Low Producing Grassland*: typical of steep hill and high country and often intermixed with high producing exotic grassland. Common of farms grazing sheep for wool and meat, and beef.
- *Depleted Grassland*: Quality and biomass substantially lower than low producing grassland due to repeated burning or overgrazing. Areas of bare ground common and typical of south island high country.
- *Tall Tussock Grassland*: Usually high altitude; common of South Island high country. Supports summer grazing and paddock sizes are usually extensive.

Land environments of NZ (LENZ)

The LENZ raster layer is derived from the multivariate classification of fifteen climate, soil and landform variables into environments with similar characteristics and irrespective of proximity. The variables were based on those likely to affect species abundance and distributions and may act as a proxy for different ecosystem types. They are all available as individual data layers and are described below. Data resolution is 25 meters with WGS 84 projection. The data used were 25 metre DEM from Landcare Research, the Land Resource Inventory (LRI) database and long-term climate data from weather stations. There are four separate layers available with increasing levels of classification detail: 20, 100, 200 and 500 classification groups, and they are available in grid or polygon formats (Leathwick et al. 2002). Downloaded at www.koordinates.com.

LENZ Layers

Below is a brief description of the fifteen data layers used in the creation of LENZ. All are available to download separately at <http://iris.scinfo.org.nz/>. More information on the layers and how they were derived is contained in Leathwick et al (2002). Of the fifteen layers used for the creation of LENZ,

seven are climate layers, seven are soil attribute layers based on NZ Land Resource Inventory (NZLRI) data, and the last is a slope layer:

- *Mean annual temperature*: Data summaries are 12 monthly averages of daily average temperature derived from 300 climate stations across NZ between 1950 and 1980, and interpolated to 25 meters.
- *Mean minimum temperature of the coldest month*: Estimates of the average minimum temperature for July. Data are derived from 346 meteorological stations between 1950-1980 and interpolated to 25 meters.
- *Mean annual solar radiation*: Data were available for 98 meteorological stations either through direct measurement of solar radiation or indirect measurement of sunshine hours. Monthly humidity data was used to improve the surface fit. Data averaged to 1980.
- *Winter solar radiation*: As for annual solar radiation. Solar radiation is at a minimum in June.
- *October vapour pressure deficit*: Vapour pressure deficit (δe) was calculated from temperature and humidity data from 287 meteorological stations. The month of October was used due to the more persistent westerly winds which result in strong geographic variations in vapour pressure deficits. Data averaged to 1980.
- *Annual water deficit and monthly water balance ratio*: Derived from monthly data describing daily average temperatures, daily solar radiation and monthly rainfall. Evaporation was estimated from temperature and solar radiation data and then compared with rainfall to give the water deficit and accumulated through the year. The ratio of rainfall to evaporation calculated for each month and then averaged gave the monthly water balance ratio.
- *Slope*: Created from the 25 metre digital elevation model.
- *National Soils Database*: Chemical and physical composition of soils from 1500 test sites across NZ. Inaccuracies in the soil layers are outlined in Leathwick et al (2002). These data was used to derive the seven soil layers used in LENZ.
- *Drainage*: Categorical data layer on drainage quality, based on the NZ Soil Classification and soil attributes which arise under different drainage conditions.
- *Soil fertility*: Four categorical data layers were derived to describe soil fertility in LENZ based on the parent material (the material from which the soil develops): acid soluble phosphorous classes, exchangeable calcium classes, induration classes and particle size classes.

- *Soil age*: A variable derived from NZ Soil Classification with two categories: young and old. It is based on morphology and development of the soil rather than on a chronological timeline.
- *Chemical limitations to plant growth*: Factors influencing chemical accumulation inhibiting plant growth represented as a categorical variable with three classes: low, medium and high.

NZ Mainland Contours

Contour layer for all of NZ to provide slope and elevation information. Data are at 1:50,000 scale and available from Land and Information NZ (LINZ).

Mainland river polygons

Data layer of large or significant rivers that are wide enough to be described in 1:50,000 as a feature. Layer updated October 2012. This layer is also available as a centreline layer.

Aspect/elevation/slope

Derived from the 25m DEM.

Deprivation index

Data layer from Statistics New Zealand census data at the meshblock level, the highest resolution data collected by Statistics New Zealand. The deprivation index is a decile measure of social-economic status of the meshblock rated 1-10 with 1 being the lowest and 10 being the highest.

Other available GIS data

Other sources of GIS data that were considered for this study are outlined below. These datasets are either incomplete or are derived from the analysis of datasets that will be used in the models.

Applicable data layers for livestock estimation are available through the Topo50 series (available at linz.govt.nz) but are currently populated with significant points only; cattlestop points, stockyard points, tank points, gate points. These should be considered for future models when detail improves.

Potential vegetation of NZ

This layer relates to forest composition derived from regressions of environmental variables with canopy tree species. Quantitative assessments could not be made of non-forest community composition (Leathwick et al., 2012).

NZ Land Resource Inventory (NZLRI)

The NZLRI is a single polygon data layer that segments NZ into approximately 100,000 polygons and describes each land parcel polygon in terms of the following physical attributes: soil, slope, rock, extent of erosion and vegetation (Newsome et al., 2008). Uncertainties are identified in all physical data and caution is recommended when applying to large scales. The minimum polygon size is ~10ha and the average is 98ha.

LUCAS NZ Land Use Map

A vector polygon layer of land use classifications created to meet the NZ's reporting obligations to the United Nations Framework on Climate Change and the Kyoto Protocol. Land use classifications used are described in the supporting documentation but fall under the broad categories of forest, grassland, cropland, wetland, settlements and other (Ministry for the Environment, 2012). Data for forest types and grassland with woody biomass are derived from 10m spatial resolution satellite imagery taken between November 2006 and April 2008. All other classifications are derived from existing databases outlined above (NZLRI and LCDB), and therefore this data layer has limited applicability as it adds little new information. The most recent version was released in June 2012 and covers all of NZ. An updated version based on 2012 data will be available in the near future. Layer can be downloaded at www.koordinates.com.

Mainland vegetation polygons

Polygon data are available for large or significant areas of different vegetation types at 1:50,000: scrub (vegetation <3m), exotic, native. All data released August 2012. The LCDB outlined above will provide data for areas of vegetation which makes these layers superfluous for this application.

WorldClim

Worldclim is a set of global climate predictors with a 1km² resolution. There is gridded data for minimum temperature, maximum temperature, mean temperature, precipitation, bioclim and altitude. Data are available for past, current and future conditions. They can be downloaded at <http://www.worldclim.org/>.

Fundamental Soils Layers

Several soils attributes layers are available from Landcare Research. They fall into broad categories describing chemical attributes, physical characteristics, drainage parameters, environmental parameters and moisture properties. A soil classification layer is also available. The data are at

1:50,000 scale and derived from either results held in the National Soils Database or estimates from pedologists.

NZ Property Parcels

This vector polygon layer represents all cadastral parcels as well as descriptive attribute data such as size and purpose. Accuracy of the layer is 0.1-1m in urban areas and 1-100m in rural areas.

Mainland fence centrelines (Topo50)

This is a selection of fences that are the representative pattern of land use. Fences along roads, tracks, drains and in residential areas are not included in the layer. Dataset is incomplete. Most recent version released August 2012.

Mainland shelterbelt centrelines

This is a data layer for shelter belts, a line of trees/vegetation planted for wind protection. Not all are represented, particularly shorter shelter belts. Those that are bordering roads maybe be offset by up to 30m. Dataset is incomplete. Release date August 2014.

Protected natural areas

This is a vector layer identifying parcels of land held in the conservation estate and including regional parks. Accuracy is high due to land surveys.

NZ mainland road Centrelines

This is a vector line layer of all-weather roads. Roads that are un-named in rural areas and less than 300m in length are excluded. Layer updated October 2012.

NZ mainland building points

This is a point layer that identifies permanent roofed constructions. It contains buildings that are significant by size or location, and a point may represent one or more buildings that are in close proximity. May be offset from a road by 30m. Dataset is incomplete. Layer updated October 2012.

Livestock databases

AgriBase

AgriBase is a voluntary database developed for biosecurity and emergency management applications. It holds contact data and geospatial information on agricultural and horticultural properties.

Farms Online (FOL)

A database of contact details, location and some stock data for an estimated 90% of agricultural holdings in NZ. The database has ~220,000 records and covers all types of holdings including commercial and lifestyle blocks. Participation by landholders is voluntary and the database has some binary and some numeric data, none of which is considered to be accurate. The data are available as a GIS polygon layer.

National Animal Identification and Tracing scheme (NAIT)

NAIT is a compulsory register to track movements of cattle and deer, implemented for the monitoring of TB. The database uses FOL contact and property information. Much of the data are incomplete with data only for the point of departure or arrival location.

Predictor data from remotely sensed imagery

Several national-level remotely sensed imagery sets are available. First, the NZ government invested heavily in obtaining Quickbird II satellite imagery for the whole country in a one-off program. The data resolution is 2.4m multispectral and 0.61m panchromatic. At this resolution the detection of large bodied organisms and other similar sized objects is becoming more feasible (Turner et al., 2003). Second, regional councils collect aerial photography every five years. These datasets are at a resolution of 1-2m but are now available for many regions at resolutions of less than 1 metre. Third, Landsat satellite imagery has been available for 40 years, providing key data for land use change detection (Turner et al., 2003), but the resolution of Landsat 8 imagery is low at 15m panchromatic and 30m visible and near infrared. Detection of individual objects in the size range of livestock species or other predictors, with the exception of large milking and shearing sheds, could not be achieved. The low resolution limits imagery for use in vegetation analysis and it has the advantage of high temporal resolution, in the order of a few days.

The remotely sensed imagery will be best used for vegetation index mapping. For example, NDVI provides a 'greenness' index. This can be used to measure grazing pressure on land. Identification of objects related to the presence of livestock is feasible but only for large distinctive objects such as

sheds, which are commonly built with tin and therefore have a distinctive spectral reflectance signature. Spectral reflectance libraries are freely available for download. Smaller objects are difficult to detect, particularly over large spatial scales. GIS layers developed by LINZ are available for objects including cattle stops, stockyards, fencelines and water troughs. These datasets are incomplete but are being updated regularly from satellite imagery. The time and resources necessary for this project to complete analysis of satellite imagery for these types of objects, already being progressed by LINZ, would be considerable.

Landsat imagery

Landsat 8 imagery is available for free download from several sites including <http://earthexplorer.usgs.gov/> and <http://glovis.usgs.gov/>

MODIS imagery and products

MODIS (Moderate Resolution Imaging Spectroradiometer) is an instrument aboard both the Terra and Aqua satellites. They collect imagery across the entire earth every 1-2 days, recording up to 36 spectral bands at 1000m, 500m and 250m resolution. The imagery and derived products from 1999 to present day is freely available to download from <https://mrtweb.cr.usgs.gov/> after creating an account. NZ is covered by 5 tiles. MODIS products that are relevant to this study include the Normalised Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI). EVI is considered to perform better than NDVI due to higher sensitivity from the inclusion of the blue band (Huete et al., 2002). The NDVI is more sensitive to chlorophyll differences and EVI has greater sensitivity to canopy structure including leaf area and canopy type (Huete et al., 2002). NDVI and EVI products come in both 16 day and 1 month aggregations which reduce the effects of cloud cover.

Normalised Difference Vegetation Index (NDVI)

NDVI is a measure of the condition and quantity of vegetation derived from multispectral data. NDVI is a ratio calculated from:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

Near infrared (NIR) and visible infrared (RED) are spectral reflectance measurements. The value is always between -1 and 1 with larger numbers representing vegetation in good condition with high density, bare ground approaches zero and water bodies have negative values (BOM, 2014). NDVI will give an indication of grazed and ungrazed areas in remotely sensed imagery.

Enhanced Vegetation Index (EVI)

EVI is an index used to enhance the vegetation signal by incorporating the blue band to reduce atmospheric influences and canopy background, giving higher sensitivity in areas with high biomass.

Calculated from:

$$EVI = G \times \frac{(NIR - RED)}{(NIR + C1 \times RED - C2 \times BLUE + L)}$$

where NIR, RED and BLUE are reflectance measures, G is the gain factor, L is the adjustment for canopy background that accounts for NIR and RED radiant transfer through the canopy, and C1 and C2 are the correction for aerosol resistance in the red band. The EVI algorithm uses the following values: L = 1, C1 = 6, C2 = -7.5, G = 2.5 (Huete et al., 2002). EVI can be used to detect vegetation changes through pixel-wise map subtraction. Growing length can also be calculated by the period of time EVI is above a given value.

Appendix 2: Setting up predictor variables for use in SDMs

Notes for extraction of predictor variables

Collect all predictor data from the various sources. Once loaded into R, ensure that they are all in the same projection and reproject if necessary.

Extract values to the AgriBase farm polygons for the attributes required of the predictors

1. *Predictor as a raster with continuous variables* – take mean and standard deviation of the pixels/cells that intersect with farm polygons from AgriBase
2. *Predictor as a raster with categorical variables* – in the case of the LCDB, convert variables to a binomial distribution giving cells of interest a 1 and other cells 0. Sum the cell values that intersect with the farm polygons for AgriBase. For the deprivation variable, take the centroid point of the farm polygon and extract the deprivation index that overlaps with the centroid point
3. *Predictor as a vector* – convert to raster and extract as for relevant raster layers, or extract using the intersect tool in ArcGIS or other GIS program then import into R and append to the AgriBase dataframe.
4. *Measures of distance* – take the centroid point of the farm polygon and measure the distance to the variable of interest

For raster layers that have a small pixel size, for example LENZ layers with 25m x 25m, it is recommended to aggregate them into larger cells to save on computational time with little or no influence on the outputs.

Check the pairwise correlations of the predictors once they have been loaded and appended to the AgriBase dataframe

R can run out of memory quickly with large spatial datasets – the ‘biglm’ is a package in R that can be used to help reduce the memory needed for analysis

For spatial data, R keeps attributes of a SpatialPolygonsDataFrame in slots. To work with just polygons use `data@polygon` and to work with attribute data use `data@data` – the ‘@’ signifies which of the slots to select.

Much of the extraction of predictor values can be done in ArcGIS and for large datasets, particularly polygons; this may be substantially faster. If analysts have capabilities to run extractions in ArcGIS then this may be the preferred approach to obtain farm-level data. For example: to extract the native forest cover and exotic forest cover for each AgriBase polygon by using the Union function in ArcGIS to join the two separate forest layers together and then the clip function to give the area of the forest in each farm polygon.

Annotated R Code

```
### Extracting the predictor variables ###
```

The following code extracts the predictor datasets over each farm polygon for all of NZ, defined by the AgriBase shape file. The following code uses mainly spatial data.

Install libraries using the `install.packages("library1", "library2")` code for the first time using them.

Load the libraries in that you need.

```
library(raster)
library(maptools)
library(rgdal)
library(sp)
library(rgeos)
library(geosphere)
```

Set the working directory if your R-code and data are stored in different folders. This is not necessary if you are drawing your data from and saving your data to the same folder where you have opened your r-code from; R will automatically draw the data from the folder that you opened the code from.

```
setwd("C:/Users/thollings/Desktop/Livestock Demographics Lit Review/SDM Models")
```

Load in the data – in this case AgriBase – this is the shape file.

`#load('ab_loaded.RData')` – use this code with the relevant file name if you have already run and saved some previous work. Otherwise load in the original file:

```
AgriBase<- (readOGR(getwd(), "agb_0814"))
```

Ensure that all files are in the same projection

```
ab_extent<-spTransform(AgriBase, CRS("+proj=longlat +a=6378137.0
+rf=298.257222 +towgs84=0.0,0.0,0.0"))
```

Check the imported file and subset out the relevant data from the shape file – in this case we remove the columns with data on crops and horticulture etc

```
names(ab_extent)
ab<-subset(ab_extent, select = c(1:5,21,24,26,27,36,39,49,51:52,56) )
names(ab)
```

Check the dataframe and notice any inconsistencies or errors.

```
head(ab)
summary(ab)
```

Here R has loaded in AgriBase and replaced the 0's as 'no data', giving them NA's in the livestock columns. These columns should actually be 0's, i.e. they are true zero's, and therefore the NA's in this case need to be replaced with zero's.

```
ab$BEF_NOS[is.na(ab$BEF_NOS)]<-0
ab$GOAT_NOS[is.na(ab$GOAT_NOS)]<-0
```

```

ab$HORS_NOS[is.na(ab$HORS_NOS)]<-0
ab$PIGS_NOS[is.na(ab$PIGS_NOS)]<-0
ab$CAM_NOS[is.na(ab$CAM_NOS)]<-0
ab$DAI_NOS[is.na(ab$DAI_NOS)]<-0
ab$DEE_NOS[is.na(ab$DEE_NOS)]<-0
ab$POU_NOS[is.na(ab$POU_NOS)]<-0
ab$SHP_NOS[is.na(ab$SHP_NOS)]<-0

```

Check what is in the workspace that you no longer need and remove it to conserve memory. Repeat this often when there is data being stored that is no longer needed.

```

ls()
rm(ab_extent)

```

In order to have a single value representing livestock on farms we calculate the livestock units (LSU) or grazing equivalents on each farm. This is calculated by giving each livestock species a value representing their equivalence to one adult dairy cow and multiplying by the total number of these animals on the farm. References: [http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Livestock_unit_\(LSU\)](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Livestock_unit_(LSU)) and <http://www.scotland.gov.uk/Topics/farmingrural/Agriculture/grants/Schemes/LMCMS/Options/Info/LivestockInfo>

```

ab$LSU<-ab$DAI_NOS + (ab$BEF_NOS * 0.8) + (ab$PIGS_NOS * 0.4) + (ab$SHP_NOS * 0.1) + (ab$HORS_NOS * 0.8) + (ab$GOAT_NOS * 0.1) + (ab$POU_NOS * 0.01) + (ab$DEE_NOS * 0.25)

```

For predictor variables that require a distance measure or for deprivation index which requires extraction of a single category for each farm, we calculate the centroid for each farm polygon to use in the predictor extraction.

```

centroids<-gCentroid(ab, byid = TRUE)

```

We load in the relevant predictor shape files and reproject to the same projection

```

# Social deprivation index
depriv<- (readOGR(getwd(), "nz-deprivation-index-2006-meshblock-data"))
dep<- spTransform(depriv, CRS("+proj=longlat +a=6378137.0 +rf=298.257222 +towgs84=0.0,0.0,0.0"))

```

Extract the information from the predictor shape file that overlaps with the centroid of the farm and then append the value to the AgriBase data. Check the data and then save the file.

```

depriv_pts<-over(centroids, dep)
ab$depriv<-depriv_pts$NZDep2006
head(ab)
save(ab, "ab_deprivation.RData")

```

Our predictor variables of farm distance to roads, rivers and urban centre we load the relevant predictor data into R, reproject if necessary to the right projection and then use a distance function to measure the distance between the nearest relevant predictor and the farm centroid. While this approach worked well for the small test dataset containing ~700 farms, processing power was quickly exhausted when the whole dataset was considered.

```

river<- readOGR(getwd(),'nz-mainland-river-polygons-topo-150k')
road<- readOGR(getwd(),'nz-mainland-road-centrelines-topo-1250k')
urban<- readOGR(getwd(),'nz-urban-north')

road<- readShapeSpatial('nz-mainland-road-centrelines-topo-
1250k.shp', proj4string=CRS("+proj=longlat +a=6378137.0
+rf=298.257222 +towgs84=0.0,0.0,0.0"))

```

Always check the data after you have imported it. Check the classification of all variables by R to ensure R has classified them correctly.

```

plot(road)
summary(road)
class(road)
names(road)

```

Calculate the distance to the predictor variable and append to the AgriBase data. disthaversine calculates the shortest distance between two points

```

ab$dist.rd<-dist2Line(centroids, road, distfun = distHaversine) ##
distance function assumes shortest distance; as the crow flies
ab$dist.river<-dist2Line(centroids, river, distfun = distHaversine)
ab$dist.urban<-dist2Line(centroids, urban, distfun = distHaversine )

```

Check the data and the output to make sure they are realistic values and it has done what you expected it to

```

head(ab)
summary(ab)

```

If your data contain many observations you may want to consider the use of an alternative to R. In our case we used an open source spatial database extender for the PostgreSQL Database Management System. This extension (PostGIS) adds spatial functions and geometry data types to the database allowing the use of SQL queries to perform the same operations described above for the R code in a shorter time frame. PostGIS interfaces with QGIS for easy viewing of the data held in the spatial database.

Spatial data are added to the SQL database (we used PostGreSQL) via the command line. Projections must be specified for each layer added. Any self intersections in the loaded shapefiles must be corrected. The following query provides the corrected polygons as a new item in the table.

```

SELECT AddGeometryColumn('fol_1114','geom_fixed',2193,'MULTIPOLYGON',2);

UPDATE fol_1114.geom
SET ST_MakeValid(fol_1114.geom)
WHERE ST_IsValid(fol_1114.geom)='f'
FROM fol_1114 ;

```

```
CREATE TABLE fol_clean AS
SELECT gid, fol_id, ST_MakeValid(geom)
FROM fol_1114;
```

To create a table of minimum distances between farms and rivers (for example, the same process was followed for roads and cities) we first calculated distances for all rivers within 3km of a farm centroid and then in a second step, calculated minimum distances to rivers for those farms that were more than 3km away from the closest river. This allowed the query to run faster than calculating the distance between a single farm in NZ and then finding the closest one. We repeated the same logic for roads and town centres using 3km from a tarred road as our first step for roads and 200km away from a town as our first step for towns. Our queries follow in 2 steps.

Step 1:

```
CREATE TABLE agbase0814_river AS
SELECT farm_id, min(ST_Distance(ST_Centroid(a.geom_fixed), b.geom))/1000 AS distance
FROM agbase0814 a, riversname b
WHERE ST_DWithin(ST_Centroid(a.geom_fixed), b.geom, 30000) GROUP BY farm_id;
```

Step 2:

```
INSERT INTO agbase0814_river
(SELECT farm_id, min(ST_Distance(ST_Centroid(a.geom_fixed), b.geom))/1000 AS distance
FROM
(SELECT farm_id,geom_fixed FROM agbase0814 WHERE NOT EXISTS (SELECT farm_id FROM
agbase0814_river WHERE agbase0814.farm_id = agbase0814_river.farm_id)) AS a,
linz_topo250_river_cl_cur b
WHERE ST_DWithin(ST_Centroid(a.geom_fixed), b.geom, 100000)
GROUP BY farm_id);
```

The previous layers were vector layers – i.e. layers of either lines, points or polygons. These are extracted differently to raster layers, i.e. gridded cells or pixels. The code below is for the extraction of predictor variables in raster format

Extract raster data from a polygon layer (AgriBase). The following are raster layers from Land and Environments NZ (LENZ). The process for all layers is the same. First, load the raster layers into R

Mean annual temperature

```
temp<-raster("lenz-mean-annual-temperature.tif")
```

Note: the temperate datasets have been multiplied by a factor of 10 – e.g. 53 in the is 5.3°C
The cell/pixel sizes for these layers are 25m x 25m. In order to speed up computation time substantially, and due to the fact the predictor data are derived from much larger spatial scales and is unlikely to influence the results at a lower resolution of one hectare, we aggregate the pixels to 4 x

4 pixel cells to give a pixel size of 100m x 100m. The mean of the 16 pixels being aggregated is taken and given as the value of the cell.

```
temp.ag<-aggregate(temp, fact = 4, fun=mean)
```

Extract the data of the new raster layer that overlays with the farm polygon and add to the AgriBase data. The value is calculated by taking the mean value of all the raster cells that overlap with the farm polygon.

```
ab$temp_mean <- extract(temp.ag, ab, fun = function(x) mean(x, na.rm=TRUE))
```

Calculate the standard deviation also and add to the AgriBase data. Replace the NA's that are given with 0's, as these are true 0's.

```
ab$temp_sd <- extract(temp.ag, ab, fun = function(x) sd(x, na.rm=TRUE))
ab$temp_sd[is.na(ab$temp_sd)]<-0
```

Save the data file

```
save(ab, file = "ab_annualtemp.RData")
```

Mean minimum temperature of the coldest month

```
temp.min<-raster("lenz-mean-minimum-temperature-of-the-coldest-month.tif")
temp.min.ag<-aggregate(temp.min, fact = 4, fun=mean)
ab$temp_min_mean <- extract(temp.min.ag, ab, fun = function(x) mean(x,
na.rm=TRUE))
ab$temp_min_sd <- extract(temp.min.ag, ab, fun = function(x) sd(x,
na.rm=TRUE))
ab$temp_min_sd[is.na(ab$temp_min_sd)]<-0
save(ab, file = "ab_mintemp.RData")
```

Mean annual solar radiation

```
solar<-raster("lenz-mean-annual-solar-radiation.tif")
solar.ag<-aggregate(solar, fact = 4, fun=mean)
ab$solar_mean <- extract(solar.ag, ab, fun = function(x) mean(x,
na.rm=TRUE))
ab$solar_sd <- extract(solar.ag, ab, fun = function(x) sd(x, na.rm=TRUE))
ab$solar_sd[is.na(ab$solar_sd)]<-0
save(ab, file = "ab_solarrad.RData")
```

Winter solar radiation

```
solar.winter<-raster("lenz-winter-solar-radiation.tif")
solar.winter.ag<-aggregate(solar.winter, fact = 4, fun=mean)
ab$solar_winter_mean <- extract(solar.winter.ag, ab, fun = function(x)
mean(x, na.rm=TRUE))
ab$solar_winter_sd <- extract(solar.winter.ag, ab, fun = function(x) sd(x,
na.rm=TRUE))
ab$solar_winter_sd[is.na(ab$solar_winter_sd)]<-0
```

```
save(ab, file = "ab_wintersolarrad.RData")
```

Annual water deficit

```
waterdef<-raster("lenz-annual-water-deficit.tif")
waterdef.ag<-aggregate(waterdef, fact = 4, fun=mean)
ab$waterdef_mean <- extract(waterdef.ag, ab, fun = function(x) mean(x,
na.rm=TRUE))
ab$waterdef_sd <- extract(waterdef.ag, ab, fun = function(x) sd(x,
na.rm=TRUE))
ab$waterdef_sd[is.na(ab$waterdef_sd)]<-0
save(ab, file = "ab_waterdeficit.RData")
```

Save all the extracted data into a single dataframe to import later

```
save(ab, file = "ab_rasterlayers.RData")
```

The Land Cover Database (LCDB) is a polygon layer of vegetation and land use. Polygon by polygon extraction is difficult in R and computationally expensive.

Load in the LCDB and reproject if necessary. Due to the size this can take some time

```
lcdb<-readOGR(getwd(), "lcdb-v40-land-cover-database-version-40")
summary(lcdb)
#lcdb_trans<- spTransform(lcdb, CRS("+proj=longlat +a=6378137.0
+rf=298.257222 +towgs84=0.0,0.0,0.0"))
plot(lcdb)
```

Categorise the LCDB classifications in high quality pasture, low quality pasture and non-pasture with binomial columns

```
lcdb_extent$hqpasture<- ifelse(lcdb_extent$Name_2012 == "High Producing
Exotic Grassland",1, 0)

lcdb_extent$lqpasture<- ifelse(lcdb_extent$Name_2012 == "Low Producing
Grassland" , 1 , ifelse (lcdb_extent$Name_2012 == "Depleted Grassland", 1,
ifelse (lcdb_extent$Name_2012 == "Tall Tussock Grassland", 1, 0)))

head(lcdb_extent)
```

Rasterize the LCDB to make extraction of variables easier. To do this get the cell size from a previously loaded raster to overlay onto the LCDB.

```
mask.raster<-temp.rast ## 25m x 25m raster
```

If you are worried about losing information due to the cell size, you can make the cell/pixel size smaller and use this

```
disag.rast<-disaggregate(temp.rast , fact = c(5,5), method = 'bilinear')
```

```
plot(lcdb.rast)
head(lcdb_extent)
```

```
lcdb.hqrast<-rasterize(lcdb_extent, mask.raster, field = "hqpasture")
lcdb.lqrast<-rasterize(lcdb_extent, mask.raster, field = "lqpasture")
plot(lcdb.lqrast)
```

Extract the high quality and low quality pixels which overlay with AgriBase farm polygons. We take the sum of the pixels/cells as the cells have been given a 1 if it is high/low quality pasture and 0 if not, so the sum represents the number of cells/pixels within a farm of high/low quality pasture and can be easily converted to an area measure as we know the pixel size.

```
ab$lcdb.hqpasture<-extract(lcdb.hqrast, ab, fun = sum)
plot(ab$total_biomass~ab$lcdb.hqpasture)
```

```
ab$lcdb.lqpasture<-extract(lcdb.lqrast, ab, fun = sum)
plot(ab$total_biomass~ab$lcdb.lqpasture)
```

For the LCDB we may need to check the outputs with a similar extract in ArcGIS or other GIS program – this can be much faster and is good to check the accuracy of our R extraction. It may be preferable to use ArcGIS or other program for this particular step. In ArcGIS the appropriate tool is Intersect in the Analysis Tools of the ArcToolbox. It provides an output table of the portions of features which overlap between the two layers. This can be downloaded into excel and written into R as a .csv file and appended to the AgriBase dataframe by Farm ID.

The rasters available for slope, aspect and the digital elevation model were very large. Similarly to our experiences when measuring the distance between the centroid of a polygon to a line or point feature, the R approach worked well on small datasets but was not workable for manipulation of very large rasters. We present the R code for reference below and provide an example of our PostGIS query for completeness.

```
## SLOPE, ASPECT AND DEM data

aspect.r <- raster("asp1.tif")
aspect.r
rm(aspect, aspect.r, dat.shp, joined)
aspect_extent<-crop(aspect.r, myclip)
plot(aspect_extent)
save(aspect_extent, file = "aspect_extent.RData")
ab$mean_aspect <- extract(aspect_extent, ab, fun=mean)
ab$sd_aspect <- extract(aspect_extent, ab, fun= sd)
```

```
dem.r <- raster("25mdem_GD.tif")
dem.r
dem_extent<-crop(dem.r, myclip)
plot(dem_extent)
save(dem_extent, file = "dem_extent.RData")
rm(dem, dem.r)
ab$mean_dem <- extract(dem_extent, ab, fun=mean)
ab$sd_dem <- extract(dem_extent, ab, fun= sd)
```

```

slope.r <- raster("slope_NZGD.tif")
slope.r
slope_extent<-crop(slope.r, myclip)
plot(slope_extent)
save(slope_extent, file = "slope_extent.RData")
rm(slope, slope.r)
ab$mean_slope <- extract(slope_extent, ab, fun=mean)
ab$sd_slope <- extract(slope_extent, ab, fun= sd)

par(mfrow=c(2,3))
plot(ab$total_biomass, ab$lcdb.pasture)
plot(ab$total_biomass, ab$road_dist)
plot(ab$total_biomass, ab$urban_dist)
plot(ab$total_biomass, ab$river_dist)
plot(ab$total_biomass, ab$sd_slope)
plot(ab$total_biomass, ab$sd_dem)

```

Post GIS sql query for the extraction of the “slope” raster to each farm polygon.

```

CREATE TABLE agbase0814_slope AS
SELECT farm_id, (ST_SummaryStats(ST_Union(ST_Clip(b.rast, ST_Transform(a.geom_fixed,4167))))).* FROM
agbase0814 a, slope_nzgd b WHERE ST_Intersects(ST_Transform(a.geom_fixed, 4167), b.rast) GROUP BY
farm_id;

```

It was necessary to examine the extent of the correlation of the extracted predictor variables and to add composite predictors to decrease this destabilising effect on the constructed models.

REGIONAL COUNCIL AND NORTH/SOUTH ISLAND CLASSIFICATION

Load in the Regional council classification data and transform. If you have removed AgriBase from the dataframe reread it in

```

RegCouncil_data<-(readOGR(getwd(), "REGC2015_HD_Clipped"))
RC<-spTransform(RegCouncil_data, CRS("+proj=longlat +a=6378137.0
+rf=298.257222 +towgs84=0.0,0.0,0.0"))

```

Create a spatial lines dataset of the AgriBase polygons

```
bound<-gBoundary(ab, byid= TRUE)
```

Run the spatial overlay of the AgriBase boundaries and the regional councils polygons

```
rc.data<-over(bound, RC)
names(rc.data)
```

Attach them to the AgriBase dataframe

```
ab$council<-rc.data$REGC2015
ab$council_name<-rc.data$REGC2015_N
summary(ab$council_name)
```

Classify each AgriBase farm polygon into north or south island based on the regional council classification and classify it as a factor

```
ab$island<-as.factor(iffelse(ab$council_name == "Canterbury Region",  
"south", iffelse(ab$council_name == "Marlborough Region", "south", iffelse  
(ab$council_name == "Nelson Region", "south", iffelse(ab$council_name ==  
"Otago Region" , "south", iffelse(ab$council_name == "Southland Region",  
"south", iffelse(ab$council_name == "Tasman Region", "south", iffelse  
(ab$council_name == "West Coast Region", "south", "north"))))))))
```

There are problems in R with the use of ' in "Hawke's Bay Region" so replace this with "Hawkes Bay Region". The data must be converted to a character string first

```
ab$council_name<-as.character(ab$council_name)  
ab$council_name[which(ab$council_name == "Hawke's Bay Region")]<-"Hawkes  
Bay Region"  
ab$council_name<-as.factor(ab$council_name)  
summary(ab$council_name)
```

CORRELATION OF PREDICTOR VARIABLES

```
library(Hmisc)  
library(RColorBrewer)  
library(gplots)
```

```
load('K:/CEBRA/full_data/ful.RData')
```

Pare down the columns so that we are only seeing the things that are continuous exposure variables

```
names(ful)
```

```
col <- ful[ , c("size_ha", "water_deficit_mean", "water_deficit_stddev",  
"water_balance_mean", "water_balance_stddev", "mean_minimum_temp",  
"stddev_minimum_temp", "temp_mean", "temp_stddev", "solar_winter_mean",  
"solar_winter_stddev", "mean_solar_radiation", "stddev_solar_radiation",  
"lq_shape_area", "hq_shape_area", "oct_vpd_stddev",  
"oct_vpd_mean", "mean_slope", "stddev_slope", "mean_dem", "stddev_dem",  
"mean_asp", "stddev_asp", "deprivation_index", "distance_city",  
"distance_road", "distance_river")]
```

```
col$lq_shape_area[is.na(col$lq_shape_area)] <- 0  
col$hq_shape_area[is.na(col$hq_shape_area)] <- 0  
col <- as.data.frame(sapply(col, as.numeric))
```

Now do the correlation matrix and pearsons correlation coefficient and save a matrix object.

```
mat <- cor(col, use="pairwise.complete.obs")  
cor(col, na.rm=T)
```

```

#+ scale_colour_gradient(limits=c(0, 1), low="red")
rc <- rainbow(nrow(mat), start = 0.3, end = 0)
cc <- rainbow(ncol(mat), start = 0.3, end = 0)
help(heatmap) #default is to reorder by the mean - but it is better to use distance based on (cor)

symnum(mat)

hU <- heatmap.2(mat, RowV=FALSE, symm = TRUE, col =
rev(rainbow(50)),distfun = function(c) as.dist(1 - c), trace="none",
dendrogram="none", margin=c(10,10))

plot(ab$total_biomass~ab$lcdb.pasture)
cor.test(ab$total_biomass, ab$lcdb.pasture, conf.level = 0.95 , method =
"pearson" )
rcorr(ab$total_biomass, ab$lcdb.pasture)
ab$animalsperha<-ab$total_biomass/ab$lcdb.pasture

```

Correlations between predictors

```

predictors<-subset(ab, select = c(19:50) ) ## can change this to have more or less
predictors
cor(predictors@data, use = "pairwise.complete.obs") ## the table of pairwise correlations
panel.cor <- function(x, y, digits=2, prefix="", cex.cor, ...)
{
usr <- par("usr"); on.exit(par(usr))
par(usr = c(0, 1, 0, 1))
r <- abs(cor(x, y, use="pairwise.complete.obs"))
txt <- format(c(r, 0.123456789), digits=digits)[1]
txt <- paste(prefix, txt, sep="")
if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
text(0.5, 0.5, txt, cex = cex.cor * r)
}

```

Plot the pairwise correlations between predictors

```

pairs(predictors, lower.panel=panel.smooth, upper.panel=panel.cor)
dev.off()

```

```

pairs(predictors, lower.panel=panel.smooth, upper.panel=panel.cor)
pdf("pairs.pdf", height=20, width=20)
dev.off()

```

```

predictors$rel.biomass <- predictors$total_biomass /
predictors$lcdb.pasture
predictors$log.rel.biomass <- log(predictors$rel.biomass)
pdf("pairs1.pdf", height=40, width=40)
pairs(predictors, lower.panel=panel.smooth, upper.panel=panel.cor)
dev.off()

```

This generated the following figure which allowed us to identify and address correlation.

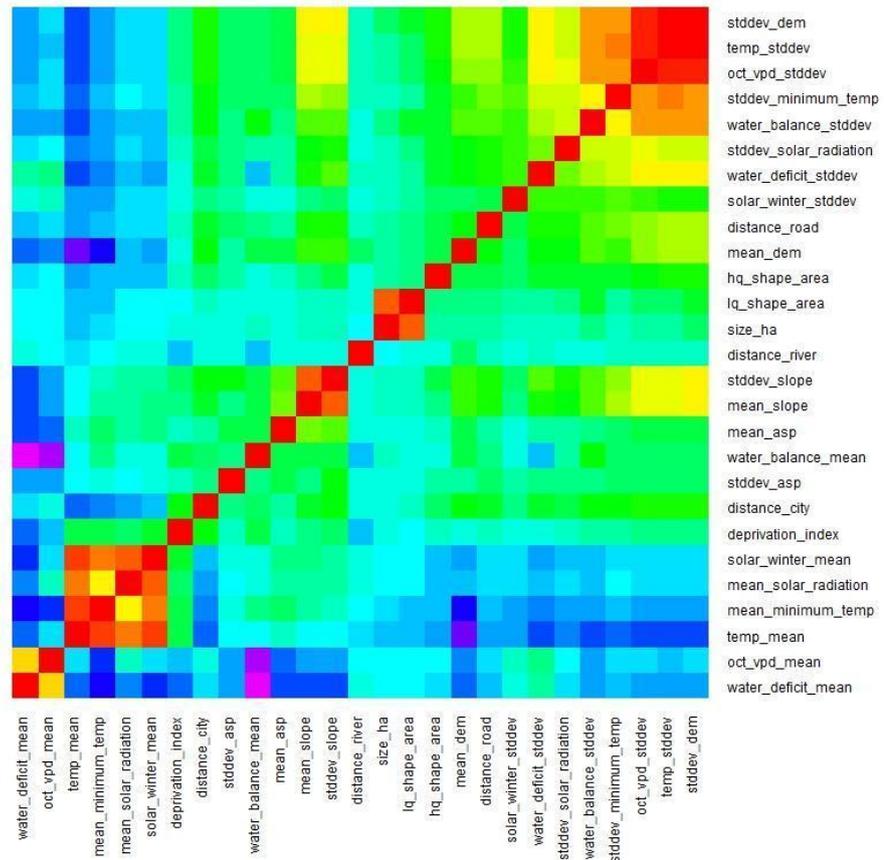
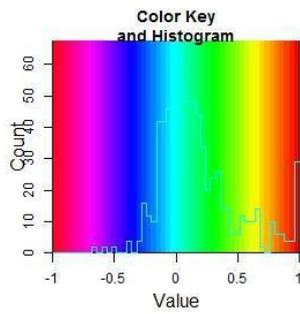


Figure 14.1: Correlation identified amongst predictor variables extracted from GIS and land cover data

To account for the large amount of correlation between slope, aspect and DEM, these were combined as suggested by Stage in 1976 (An expression for aspect, slope and habitat type on tree growth, Forest Science) as the slope multiplied with the cosine of the aspect and the slope multiplied with the sine of the aspect. DEM, slope and aspect could then be removed from the final model. Solar radiation in winter, mean and sd and temperature winter, mean and sd were (as might be

expected) highly correlated. These measures were condensed by finding the difference between solar radiation in the winter/mean winter temperature and mean solar radiation/mean temperature respectively. This significantly reduced the amount of correlation in the dataset.

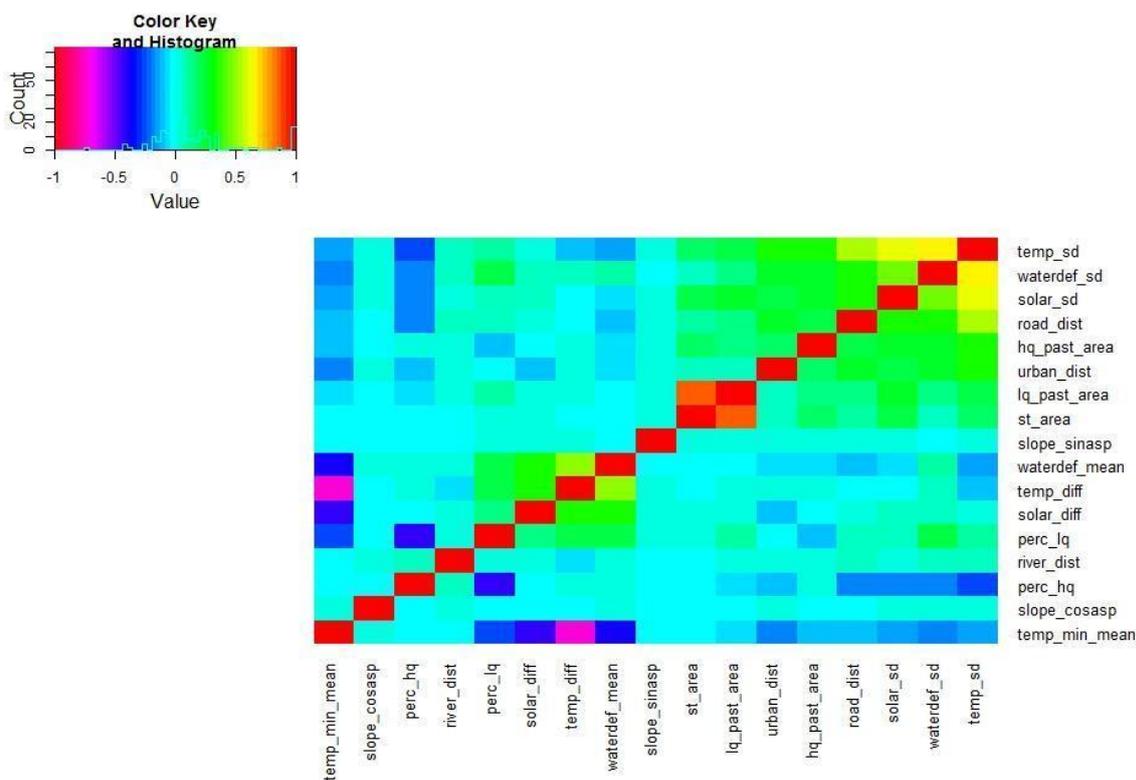


Figure 14.2: remaining correlation once synthesized variables have been added to the dataset representing variables included in the model building process.

Livestock Species Distribution Modelling

Note: check the classification of all variables by R before they are used in the models – e.g. if they are numeric categorical variables such as the decile index of deprivation measured from 1-10, make sure R classifies them as such.

SET UP THE BUFFERS FOR THE 200 RANDOMLY SELECTED FARMS

```
library(raster)
library(maptools)
library(rgdal)
library(sp)
library(rgeos)
library(geosphere)
```

load in AgriBase as spatial data

```
AgriBase<- (readOGR(getwd(), "agb_0814"))
```

extract the random farm ploygons (200 gives an average number of famrs as 20,000 (20% of the dataset once missing data has been removed))

```
target<-sample(1:nrow(AgriBase), size = 200, replace = FALSE)
```

subset these from the whole spatialpolygon - while maintaining spatial information in order to extract farms in a 3km radius

```
ab.subset<- AgriBase[target ,]
```

Outline the boundaries of the target farm

```
ab.bound<-gBoundary(ab.subset, byid = TRUE)
```

Calculate the 3km buffer around the target farm boundary

```
ab.buffer<-gBuffer(ab.bound, width = 3000, byid = TRUE)
```

Run the intersection of the target farm 3km buffer with AgriBase to determine all farms within 3km of target farm

```
ab.buffer.farms<-gIntersection(ab.buffer, AgriBase, byid = TRUE,
drop_lower_td = TRUE)
```

```
ab.buffer.farms
```

Has been converted to a SpatialPolygons, need to add the attribute data back

```
summary(ab.buffer.farms)
```

Convert back to SpatialPolygonsDataframe. Good reference

<http://gis.stackexchange.com/questions/90575/join-grid-or-polygon-layers-in-r>

```
new.attrs <- data.frame(do.call(rbind,
strsplit(row.names(ab.buffer.farms), " ")))
AgriBase$X2 <- row.names(AgriBase)
#ab.buffer$X2 <- row.names(ab.buffer)
ab.target.farms <- merge(new.attrs, AgriBase, by='X2') ## dataframe
for use in models
row.names(ab.target.farms) <- row.names(ab.buffer.farms)
```

Set up the SpatialPolygonDataFrame for later

```
ab.target.poly = SpatialPolygonsDataFrame(ab.buffer.farms, ab.target.farms)
class(ab.target.poly)
plot(ab.target.poly)
```

```
head(ab.target.farms)
length(ab.target.farms$FARM_ID)
```

SET UP TRAINING AND TARGET DATA

Divide into training and target data - target data are buffer farms

```
ab<-read.table("/home/thollings/SDM Models/ab_23122014.csv", header= T, sep
= ",")
```

```
ab.target.farms$farm_id<-ab.target.farms$FARM_ID ## make sure the titles
are the same between the dataframes - AgriBase original is in capitals
ab$target<-(ab$farm_id %in% ab.target.farms$farm_id)
ab.target<-subset(ab, target == "TRUE")
ab.training<-subset(ab, target == "FALSE")
```

Remove farms with suspected missing values from the training data

```
ab_missingdata<-read.table("ab_missingdata.csv", header = T, sep = ",")
ab.training1<-merge(ab.training, ab_missingdata[,c(2,8)], by = "farm_id")
ab.train.narm<-subset(ab.training1,
ab.training1$Is.there.records.for.farm>0)
```

For model validation we also need to remove it from the target data

```
ab.target1<-merge(ab.target, ab_missingdata[,c(2,8)], by = "farm_id")
ab.target.narm<-subset(ab.target1, ab.target1$Is.there.records.for.farm>0)
```

Output target and training to .csv

```
write.csv(ab.train.narm, file = "ab.training.data.csv", row.names = FALSE)
write.csv(ab.target.narm, file = "ab.target.data.csv", row.names = FALSE)
```

ONCE YOU HAVE THE DATA RELOAD IN AND RUN THE FOLLOWING CODE

Replace NA's with 0 where appropriate

```
ab$lq_past_area[is.na(ab$lq_past_area)]<-0
ab$hq_past_area[is.na(ab$hq_past_area)]<-0
```

Change area's into HA rather than square meters

```
ab$hq_ha<-ab$hq_past_area/10000
```

```
ab$lq_ha<-ab$lq_past_area/10000
```

Merge the regional council data

```
ab_council<-read.table("regional_councils.csv",header=T,sep=",")  
ab<-merge(ab, ab_council[,c(1,4:6)], by = "farm_id")
```

```
sapply(ab, class)  
ab$depriv<-as.factor(ab$depriv)  
ab$council<-as.factor(ab$council)
```

```
ab.rmna<-subset(ab, !is.na(temp_mean))  
ab.rmna<-subset(ab.rmna, !is.na(council_name))  
length(ab.rmna$lsu)
```

Remove farms with no enterprise data as suspected missing values

```
ab_missingdata<-read.table("ab_missingdata.csv", header = T, sep = ",")  
  
ab.rmna<-merge(ab.rmna, ab_missingdata[,c(2,8)], by = "farm_id")  
ab.rmna<-subset(ab.rmna, ab.rmna$Is.there.records.for.farm>0)  
summary(ab.rmna)
```

Appendix 3: Bayesian ZIP model for LSUs and animal counts, covariate table

Mean and standard deviation of the posterior estimates of covariates for species distribution models for New Zealand farms using counts of livestock units as the outcome variable (LSU) 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.

LSU model	Cattle model					
Bernoulli component						
Predictor variable	mean	std dev	OR (5%-95%CI)	mean	std dev	OR (5%-95%CI)
intercept	3.353	0.007	28.58 (28.48 - 28.68)	3.074	0.012	21.62 (21.53 - 21.71)
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 defficit	-	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 - 1.18)	-	0.573	0.016	1.77 (1.75 - 1.83)

			1.19)			
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 defficit	-	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 (Manwatu-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)	-	0.091	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)	-	0.054	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)	-	0.057	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)

Appendix 4: External validation of model estimates against external data sources

Appendix 4.1 – methodology for matching Statistics New Zealand mesh blocks to Agribase Farms

When matching the AgriBase properties to mesh blocks, property centroids were buffered by 3km to assess if the properties whose centroids were falling outside of the mesh blocks were an artefact of how the matching was done. Agreement between this buffered selection and the mesh block totals was significantly poorer than when centroids were not buffered before being assigned to mesh blocks.

The following table (Table 14.1) shows the combinations of factors tried when deciding on the best selection criteria for AgriBase farms for inclusion in the validation against APS mesh block data.

1. Each AgriBase farm centroid was buffered by 3km to improve overlap with mesh blocks
2. Each AgriBase farm centroid was overlapped with the mesh blocks
3. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm class = "LIF" were removed
4. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm type ="LIF" were removed
5. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB were removed
6. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, NEW, OTH, UNS, NAT were removed
7. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, NEW, OTH, UNS, NAT, BEF were removed
8. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF were removed
9. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF, NEW, UNS were removed
10. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF, NAT, OTH were removed
11. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF, NAT, NEW, UNS were removed
12. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF, NEW were removed
13. Each AgriBase farm centroid was overlapped with the mesh blocks and all farm types that were *not* DAI, DEE, DRY, GRA, PIG, SHP, SNB, BEF, NEW, OTH, UNS, NAT, BEF were removed

Table 14.1: MAE values for LSUs and individual species for the 13 selection variations described above. The option (option3) that minimised the most MAE values for the individual species and for LSUs was chosen as the most appropriate selection criteria. The final row represents mean error (not absolute error) which was also taken into account.

combination	1	2	3	4	5	6	7	8	9	10	11	12
lsu	692.4	686.9	686.9	709.1	794.1	770.8	740.5	749.4	745.1	742.9	743.6	745.9
beef	283.9	278	275.2	283.4	297.1	289.2	292.9	296.2	294.7	293.7	293.9	294.9
dairy	385.9	389.3	392.7	403.7	458.1	442.9	421	427.3	424.2	423.2	423.4	424.8
deer	117.8	116.8	117.7	121.7	136.7	132.5	127.1	128.3	127.6	127.5	127.4	127.8
pigs	32.9	32.7	32.7	33.9	35.2	33.8	34.6	35.2	34.9	34.9	34.5	34.9
sheep	2121	2097.3	2112	2176	2417	2344	2267	2293	2281	2273.5	2277	2283
mean error	-	22.3	10.6	18.5	-112	-114	9.9	11.7	11.1	10.5	10.94	11.02

The variation between option 2 and 3 was further explored. Option 2 is a selection of all AgriBase centroids that overlap mesh blocks, and option 3 is the same selection with all lifestyle properties (in the farm class column) removed. The following figure results:

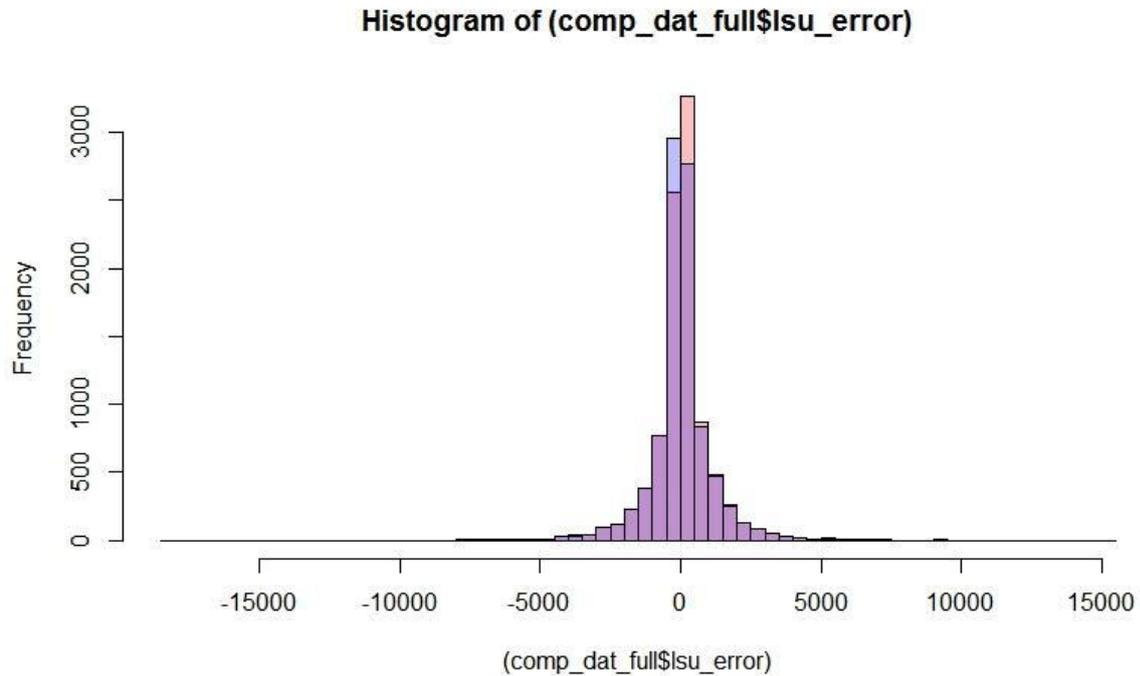


Figure 14.3: The red bars represent the size of error calculated at the mesh block level for strategy 2 (red) and strategy 3 (blue). While removing the lifestyle properties does decrease the MAE (Table 14.1) for some species, and affects the mean error positively, when plotted the variation in errors remains the same (the large amount of overlapping red and blue which shows as purple).

Appendix 4.2 – FarmsOnLine Properties by Landuse Type

The following table shows the counts of properties in FOL with each of the farm types or categories which each farm is assigned to. The farm category identifier is not useful for the purposes that we require it for in this analysis as it does not appear to have much bearing on the population of animals on the farms in question.

Category according to FOL	FOL with animals	Total FOL	% with animals	Dairy	Beef	Deer	Goats	Sheep	Pigs	Horses	Camelids	Buffalo
No category	2	14	14.29	0	2	0	0	0	0	1	0	0
Arable	963	1538	62.61	90	556	48	15	725	59	117	9	0

Commercial	149	204	73.04	25	134	28	10	59	11	23	4	1
Dairy	16489	18934	87.09	15110	6086	296	381	2831	1178	1872	40	0
Education	68	119	57.14	18	55	4	5	36	7	22	2	0
Elderly	5	10	50.00	1	5	0	0	3	0	1	0	0
Food Processing	8	13	61.54	1	8	3	0	4	3	2	0	0
Forestry-Indigenous	153	1426	10.73	58	288	27	12	204	31	63	1	0
Forestry-Protected	59	388	15.21	23	115	14	3	63	22	50	1	0
Forestry-Vacant	32	280	11.43	11	26	7	1	22	1	40	0	0
Forestry Exotic	404	3800	10.63	4	27	0	5	19	3	6	0	0
H-Berry	27	95	28.42	4	14	1	0	14	3	5	1	0
H-Citrus	31	119	26.05	0	21	0	0	18	3	8	1	0
H-Flower	43	163	26.38	0	4	0	0	2	1	1	0	0
H-Glasshouse	122	647	18.86	4	30	1	4	12	1	7	1	0
H-Kiwifruit	311	1761	17.66	4	79	3	5	56	2	21	1	0
H-Market Garden	231	918	25.16	32	230	9	17	102	24	42	0	0
H-Pipfruit	115	681	16.89	18	173	4	8	72	13	37	2	0
H-Stonefruit	75	482	15.56	27	392	31	25	265	43	112	8	0
H-Vines	317	1895	16.73	5	73	8	8	59	11	21	1	0
Health	4	7	57.14	10	48	2	4	20	3	12	0	0
Horticulture	615	2553	24.09	16	187	21	13	189	19	74	2	0
Industrial	171	365	46.85	34	154	24	15	72	15	30	3	0
Lifestyle	50520	139373	36.25	5099	36784	1538	2645	20446	3736	9940	582	1
Liquor	7	13	53.85	1	5	2	2	5	1	3	0	0
Maori Sites	18	63	28.57	3	17	1	1	7	1	4	0	0
Mining	16	57	28.07	8	12	2	3	11	3	6	0	0
New Block	2	84	2.38	0	2	0	0	2	0	0	0	0
Other	994	2497	39.81	274	790	110	97	661	143	330	10	0
Parking	0	1	0.00	0	0	0	0	0	0	0	0	0
Pastoral	26225	36699	71.46	2720	23035	2339	928	17628	2070	5531	87	2
Religious	15	32	46.88	4	15	0	1	6	1	2	0	0
Reserve	112	436	25.69	31	82	9	3	58	4	28	2	0
Residential	589	1467	40.15	107	439	40	53	342	80	155	11	0
Retail	10	31	32.26	3	9	2	2	7	1	3	0	0
S-Aquaculture	4	21	19.05	1	3	0	0	0	0	1	0	0
S-Deer	977	1151	84.88	79	577	733	76	457	62	160	8	0
S-Horses	510	720	70.83	25	275	8	16	125	14	417	4	0
S-Other	79	720	10.97	14	50	8	25	30	2	19	1	0
S-Pigs	83	108	76.85	6	55	1	4	38	57	14	0	0
S-Poultry	194	357	54.34	55	196	81	5	109	43	94	1	0
Sporting	234	936	25.00	5	151	5	1	83	4	15	1	0
Service Station	0	2	0.00	0	0	0	0	0	0	0	0	0

Tourism	4	10	40.00	0	4	1	0	1	0	0	0	0
Utility	24	74	32.43	3	19	1	1	10	3	7	0	0

Appendix 4.3 – Matching the Agricultural Production (APS) dataset to FarmsOnLine (FOL) by address

1. Input data

a. Agricultural Production Census 2012 [Classified dataset]

The APS census 2012 was collected jointly by Statistics New Zealand and the Ministry for Primary Industries (MPI). The dataset (58,070 records) was both authorised for use in the CEBRA project and provided in the form of csv files by the MPI Sector Policy Team (contact: Aaron Carson, Team Leader/ Senior Policy Analyst).

The following input fields (Table 14.2) were used in this analysis:

Table 14.2: *Input fields from the Agricultural Production (APS) Census data used in an analysis to match the APS and FarmsOnLine datasets for the purposes of model validation.*

APS Field Name	Alternative Field Name	Description
Entunit		Farming Enterprise Identifier
BF_StreetNameText		House number and street name (BF ²²)
BF_SuburbLocalityDistText		Suburb or Locality (BF) ²³
BF_CityTownText		City or Town (BF)
lc6731	Sheep_total	Total number of sheep
lc6848	Goats_total	Total number of goats
lc6954	Pigs_total	Total number of pigs
lc7077	BeefCattle_total	Total number of beef cattle
lc7193	DairyCattle_total	Total number of dairy cattle
lc7699	Deer_total	Total number of deer

²² Source of information: Statistics New Zealand's Business Frame

²³ Suburbs and localities are geographic subdivisions used for address purposes. Suburbs refer to subdivisions in urban areas, whereas localities represent the equivalent in rural areas (cf. Land Information New Zealand, 2015).

b. **FOL address data [Classified dataset]**

The FOL dataset (comprising spatial data on farm extent and non-spatial information on owner and addresses; 234,645 records) was created by MPI (and its predecessor Ministry for Agriculture and Forestry) for use during biosecurity incursion responses. The address dataset was supplied by Mary van Anandel in form of a csv file, authorised for use by the MPI FarmsOnline Team (contact: Nigel White, System Business Analyst).

The following input fields were used in this analysis (Table 14.3):

Table 14.3: *Input fields from FarmsOnLine (FOL) the Agricultural Production (APS) Census data used in an analysis to match the APS and FOL datasets for the purposes of model validation.*

FOL Field Name	Description
FamrID	Farm Identifier
Number	House number
Road	Road name
Locality	Suburb/Locality or Town/City

2. Method, steps in wording, and quality assurance steps

Address matching was undertaken in RStudio Desktop 0.99 software using for R version 3.2.2 and the external packages stringdist (Van der Loo, 2014), stringr (Wickham, 2015), and reshape2 (Wickham, 2007). The selection of matching physical address strings (i.e. sequences of characters) was based on three string distances: the Jaro distance, the Jaro-Winkler distance, and the Jaccard distance. Information on string distances was primarily sourced from Van der Loo (2014). The choice of string distances, their application order and choice of thresholds was based on an experimental approach, aiming foremost at a high confidence output as opposed to a maximum number of matched records of lower confidence.

String distances allow approximate string matching by providing a quantitative measure on the similarity between two strings in terms of string metrics (see Van der Loo, 2014, for details). Van der Loo (2014) differentiates three types of string metrics: edit-based distances, *q*-gram distances, and heuristic distances. *Edit-based distances* count the number of operations (changes such as

substitution, insertion, deletion, or transposition) necessary to turn one string into another. *Q-gram distances* (e.g. the Jaccard distance) are calculated by comparing the occurrence of character sequences of *q* length between strings.

Heuristic distances have been developed for particular purposes. The Jaro distance and Jaro-Winkler distances were developed for linking records based on slightly inaccurate text fields. Both metrics have successfully been applied for matching short strings such as name and address data. The Jaro distance is based on the assumption that character mismatches and transpositions are caused by typing errors, whereas matches between remote characters within a sequence are unlikely to be caused by a typing error. The Jaro-Winkler distance is an extension to this assumption, applying an additional penalty for mismatches in the first four characters (e.g. for differing house numbers on the same street).

Based on their special suitability for address comparison, the Jaro-Winkler distance (focusing on matching of identical house numbers on identical streets) and the Jaro distance (focusing on matching overall identical address information) were chosen as the principal functions for address matching in this project. The Jaccard distance was introduced in a third step to eliminate remaining mismatches wrongly interpreted by the heuristic distances as misspellings.

The string metrics used in this analysis give a measure of similarity between strings, with values of 0 indicating identical strings and values > 0 indicating some differences in string sequences. To assure that the matches provided in the final result tables are of high confidence, all entries with a final best match Jaro-distance of > 0 or a final best match Jaccard-distance > 0 were manually validated in Excel. Furthermore a random sample of 50 entries where both Jaro- and Jaccard distance were 0 were tested to validate that the matched strings with metrics of 0 were de facto identical.

Erroneous matches identified during the QA processes were removed from the final result table '*matches*' and appended to '*no matches*'.

3. Results

The presented workflow (Table 14.4) allowed to link almost a third (31%) of the APS entries with target livestock present to corresponding FOL farms:

Table 14.4: Workflow linking Agricultural Production (APS) Census data with target livestock present to FOL farms.

Analysis results	Number of records
Original APS dataset	58070
APS with target livestock species present	43430
APS after duplicate addresses removed	41182 = Analysis sample
Preliminary positive matches	13038
Matches after subsection query	12963
Matches after duplicate FOL addresses removed	12933
QA result - erroneous matches	5
Final result - Matches	12928 31.4% of analysis sample
Final result - Unmatched	28254

The final project output was

- 1) a csv table (FOL.APS.AddressMatching_FinTable_Matches.csv) with the 12928 matched records and the fields²⁴ shown in Table 14.5:

Table 14.5: Fields for the matched records for analysis matching FarmsOnLine (FOL) with Agricultural Production (APS) Census data

Output Field Name	Description
FOL_FarmID	FarmsOnLine Farm Identifier
APS_Stock_total	Total number of target species on farm as per APS data
APS_Deer_total	Total number of deer on farm as per APS data
APS_DairyCattle_total	Total number of dairy cattle on farm as per APS data
APS_BeefCattle_total	Total number of beef cattle on farm as per APS data
APS_Pigs_total	Total number of pigs on farm as per APS data
APS_Goats_total	Total number of goats on farm as per APS data

²⁴ For amended version of this result table, including information on livestock grazing on farm but not owned by farmer see Amendment (Section 10)

2) a csv table (FOL.APS.AddressMatching_FinTable_APSLocationPivot_15Jan16.csv) with the summary of matches per APS location (2287 records) and the following fields for information on regional variance in matching successes (see Figure 14.4 and 14.5):

Table 14.6: Information on regional variance in matching success between the FarmsOnLine and Agricultural Production Census information.

Output Field Name	Description
APS_Location	Name of location in APS dataset
Record_N	Number of records for location in APS dataset
Match_N_phy	Number of physical farm addresses matched
Match_Pct_phy	Percent of physical farm addresses matched

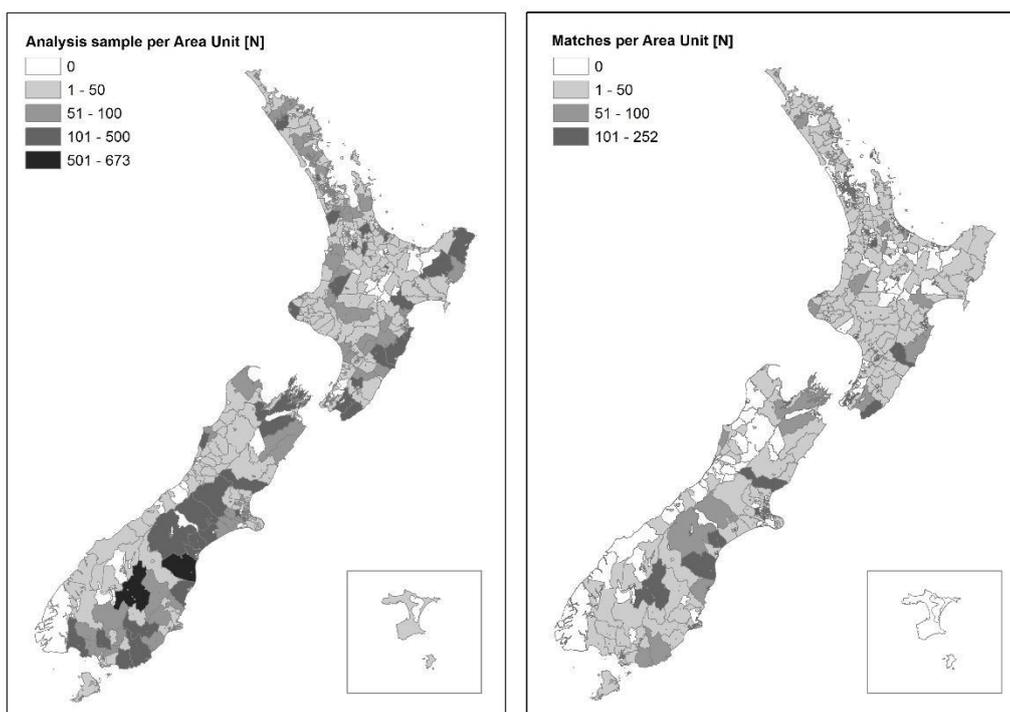


Figure 14.4: Regional distribution of APS analysis sample and achieved matches per area unit (aggregations of meshblocks) as per 2013. 492 entries of the APS analysis sample ($N = 41182$) could not be unambiguously attributed to an area unit due to incorrect/insufficient location information in

the APS dataset and are consequently not represented in the figures. Area unit boundaries were sourced from Statistics New Zealand licensed under CC-BY.

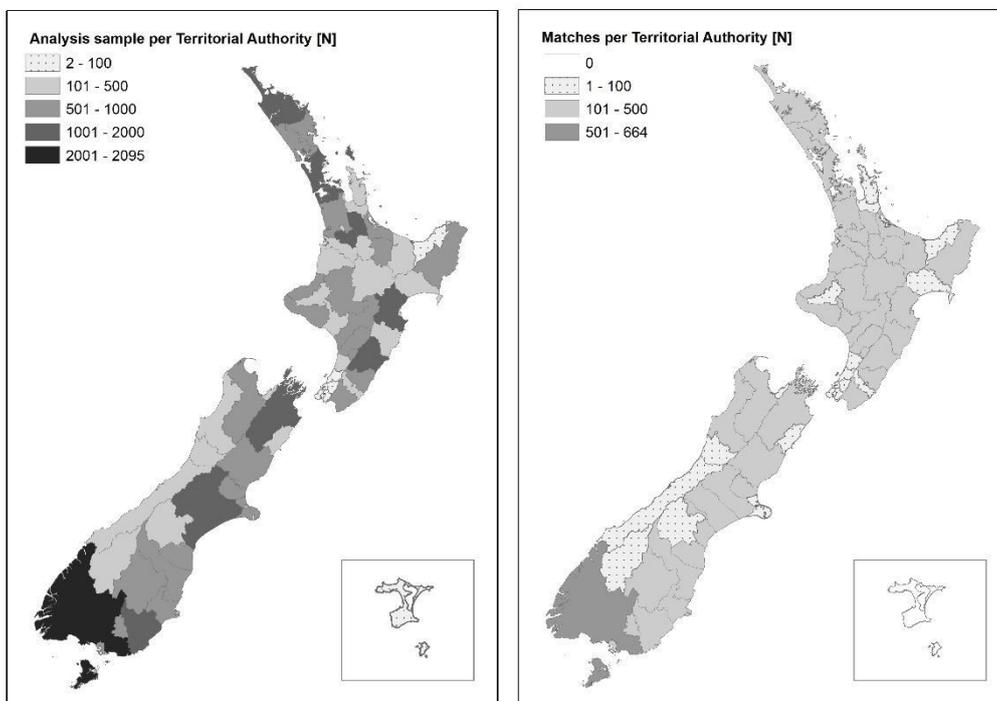


Figure 14.5: Regional distribution of APS analysis sample and achieved matches per territorial authorities as per 2013. 492 entries of the APS analysis sample ($N = 41182$) could not be unambiguously attributed to an area unit due to incorrect/ insufficient location information in the APS dataset and are consequently not represented in the figures. Territorial authority boundaries were sourced from Statistics New Zealand licensed under CC-BY.

4. Discussion Points

The QA review suggested that the project's primary objective of achieving matches of high confidence was met. Less than 0.4% of the 1353 validated address matches were invalid and were consequently removed from the final output. As all of the address matches where string distances suggested some differences in the matched addresses have been validated, high confidence exists in the quality of the output table. Sources of remaining uncertainty are:

- Minor deviations in road names may have been misinterpreted as misspellings (unlikely, addresses with notably different spelling were validated by NZ post coder finder (New Zealand Post, 2016) as part of the QA process).

- Some of the 11625 records for which string distances suggested identical sequences (string distance = 0) may not be identical, as only 50 records had been validated manually (*very unlikely as function-based calculation of string metrics*).
- Farms may be mismatched due to wrong information in APS and FOL data (*probability unknown*).

The matching rate could potentially be improved by:

Extending the workflow to also consider postal addresses, owner/ legal names, farm names or combinations of these for record matching.

Rectifying address information in both APS and FOL data by using an address validator (allows standardised formatting, completion of missing information, correction of misspelling, etc.).

Geolocating (determining coordinates for) APS census and spatially join to FOL dataset.

Improve matching of FOL entries to NZFS suburb layer [Block 7]. A minor issue was identified where farms overlap multiple locations. The current code assigns the first entry's location to the FOL farm, assuming that farms are within one location. The identified issue does not affect the quality of the analysis output, but assigning the location with the largest areal proportion might increase the matching rate.

This project provided a starting point for the MPI Spatial Analysis Solutions Team (SAS) to familiarise with the APS dataset as well as the quality of address information in both APS and FOL dataset. The acquired knowledge and data analysis skills will benefit the SAS Team as well as the MPI Sector Policy Directorate (current custodians of the APS data within MPI) in developing strategies to improve the existing dataset and to establish links between them.

5. Amendment :

To add the additional livestock numbers (animals grazing on properties but not owned by the person filling out the survey) to the existing result table, the following steps were performed in RStudio:

- 1) Read APS parts into RStudio and merge to one data frame
- 2) Select fields of interest
- 3) Add fields and copy data of livestock totals per type (Sheep, Goats, Beef Cattle, etc.) for both grazing (not owned) and owned and sum to livestock total.
- 4) Reduce data frame to only farm records with target livestock species present (to reduce processing time).
- 5) Read result table exported for QA and remove identified errors
- 6) Create new fields in result table for both grazing and owned target livestock species totals, respectively, and overall total.
- 7) Link APS livestock information to result table via row names:

Row names act in R/RStudio as unique identifier for data frame entries. Since the result table was queried from the initial APS table, the row names have carried through to the result table and can now be used to link back to the original entries for transferring the livestock numbers.

- 8) Reduce fields in final output table and rename fields
- 9) Export final output table

Appendix 4.4 – Survey of veterinary practitioners to provide validation data focussed on lifestyle farms

This work was performed to assess the external validity of model outputs predicting the numbers and species of livestock present on the farm-level. To acquire such information in a time-effective manner, elicitation of knowledge from practicing large animal veterinarians was chosen. This choice was motivated by technical considerations (perceived integrity, reliability and level of knowledge) as well as practical considerations (existing contacts). While this would result in a lower degree of precision and accuracy than direct counting of animals or interviewing farmers, it provided a larger and more representative dataset, as well as incorporating a range of production areas and characteristics.

As achieving population representativeness (in terms of geographical coverage and numbers of vets interviewed) was beyond the scope of this work, the survey was stratified to include a range of farming environments and systems on the North Island which were classified according to three descriptors: rural areas with high urban influence; rural areas with moderate urban influence; and rural areas with low urban influence (Statistics New Zealand; New Zealand: an Urban / Rural Profile, 2001) (Figure 14.6). A specific consideration was the inclusion of non-commercial smallholder blocks or 'lifestyle' holdings. Suitable veterinary practices within these areas were identified and were approached by e-mail, including a fact sheet providing background and information on the survey. This was followed up by phone calls arrange the visits. The vets to be interviewed were requested to define a geographical area for which they felt confident of being able to provide information on livestock numbers and species. Topographic maps covering the specified areas were generated on the basis of this; these were overlaid with the spatial FOL polygon layer, colour-coded by farm type, set at a level of transparency (Figures 14.7 to 14.9).

The interviews were conducted face-to-face at the practice location during June and July 2015; they were semi-structured and participatory. The maps provided a useful visual frame of reference for the vets to identify and discuss holdings about which they felt confident of being able to estimate

the approximate numbers and species of animals present. Additional clarification questions were asked by the interviewer (such as the production type of cattle, whether they were grazed on contract basis etc.) if required. In practice, discussion about one farm rapidly led to another as the vets identified nearby properties. This continued until the point at which the vets indicated they could no longer add meaningful information. On average, one farm per minute was discussed. The duration per interview was approximately 90 minutes.

A summary table showing the number of farm-level records per location is given in Table 14.7.

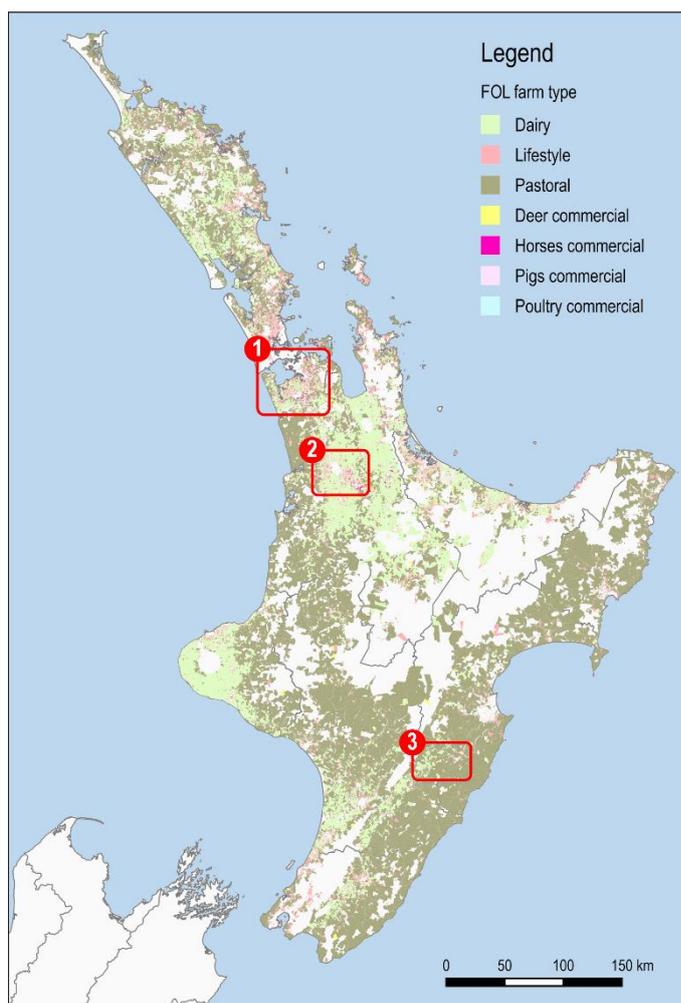


Figure 14.6: *Regions covered by the validation survey. 1: Peri-urban area / rural with strong urban influence; 2: Semi-rural area / rural with moderate urban influence; 3: Rural area / rural with low urban influence*

Table 14.7: Summary of survey results

Area category	Vet practice	Region	Sub-region	Vet	Date of interview	No. records	
Peri-urban	Franklin Vets	South Auckland	Ararimu	John Prangley	4/06/2015	4	
			Ararimu	Ilyse Jennens / Emily Roose	4/06/2015	14	
			Ardmore	John Prangley	4/06/2015	52	
			Karaka	John Prangley	4/06/2015	22	
			Karaka	Ilyse Jennens / Emily Roose	4/06/2015	27	
			Tuakau	Ilyse Jennens / Emily Roose	4/06/2015	16	
			Waiuku	David Hawkins	4/06/2015	66	
						<i>Subtotal</i>	201
Semi-rural	VetFocus	Hamilton	Whatawhata	Michelle Dixon	12/06/2015	60	
						<i>Subtotal</i>	60
Rural	Vet Services Hawkes Bay	Waipukurau	Takapau	Richard Hilson	22/06/2015	72	
						<i>Subtotal</i>	72
						Total	333

Methodology discussion:

The approach of using semi-structured interviews with veterinarians was effective to elicit information on the numbers and species of livestock present on farms of different types in the target areas. The use of maps as visual aids was intuitive to the interviewed vets, facilitated discussion, and allowed them to 'visualise' the target area and orient themselves. The maps also facilitated explanation of the purpose and objectives of the project, and gave vets a visual snapshot of FOL. Lastly, colour-coding the polygons by farm type and including the owner names enabled the vets to identify changes in ownership and purpose (which would not have been possible if a printed register had been used).

There are clear limitations to the accuracy of the information provided, which cannot easily be mitigated. These include the following:

- The level of confidence of the vets is likely to be variable and cannot be tested. This leads to a degree of uncertainty.
- The numbers of animals present are highly variable over time; it is difficult to factor this in. For example, dairy calving is highly seasonal. Vets and farmers tend to give numbers of adult stock but find it much harder to be specific about the numbers of young stock.
- Animal movements are frequent. Underreporting of animal movements between farms in NAIT has been identified by compliance activities carried out by OSPRI. In addition, the vets confirmed that there comparatively short-range movements are frequent, e.g. for seasonal grazing of young stock (to and from runoff blocks etc.). As a consequence, the number of animals on any given property is likely to fluctuate substantially.
- Practices were selected based on convenience in the identified areas based on which vets were able to participate and spend time discussing their client base.

The interviews revealed inaccuracies in FOL regarding names of owners, farm types and parcel size. This is inevitable due to the dynamic nature of buying and selling land (including subdividing), conversions to different production systems and so on. It was difficult to quantify the degree of these inaccuracies.

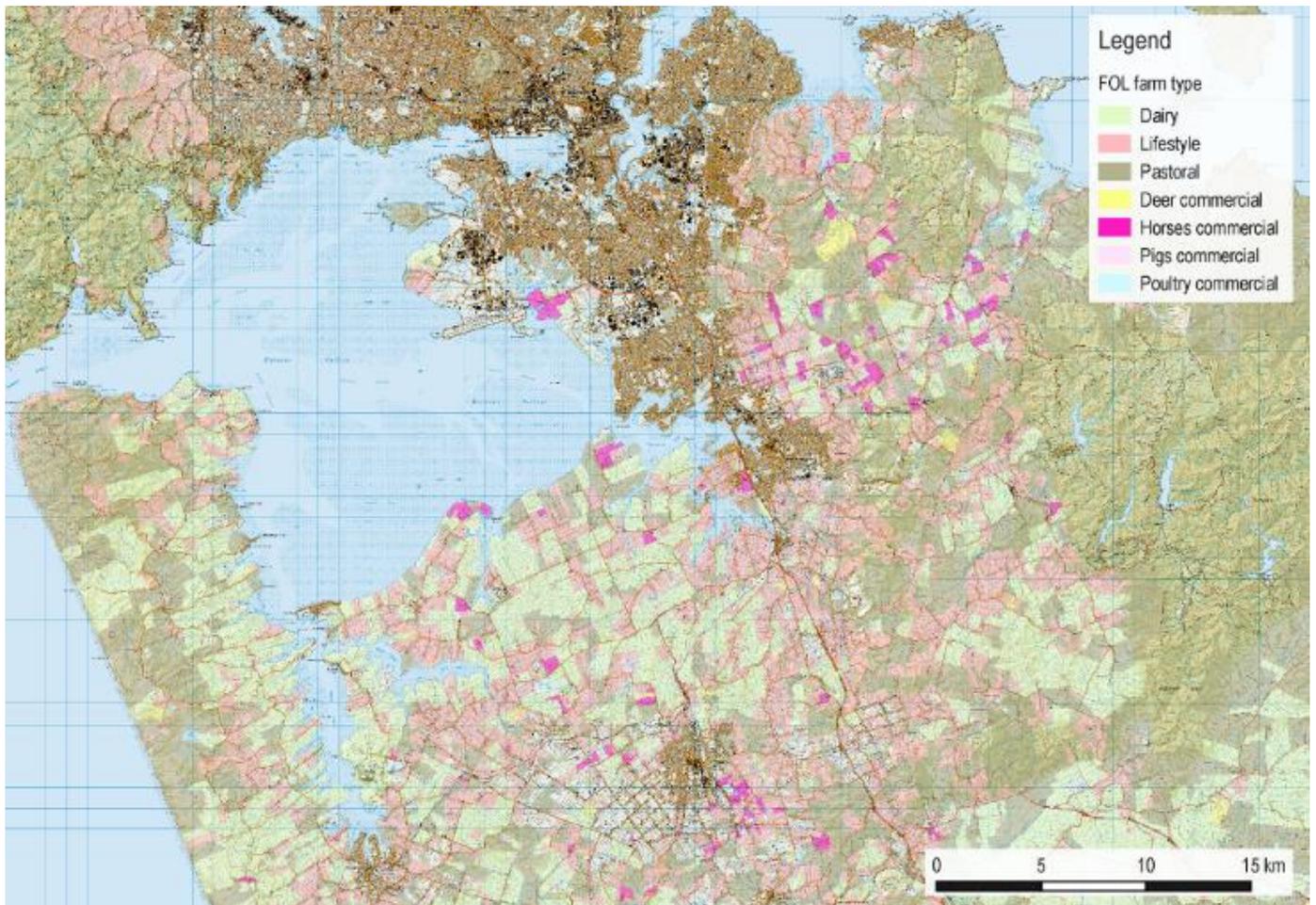


Figure 14.7: Peri-urban survey region (South Auckland)

Figure 14.8: Semi-rural survey region (Hamilton area)

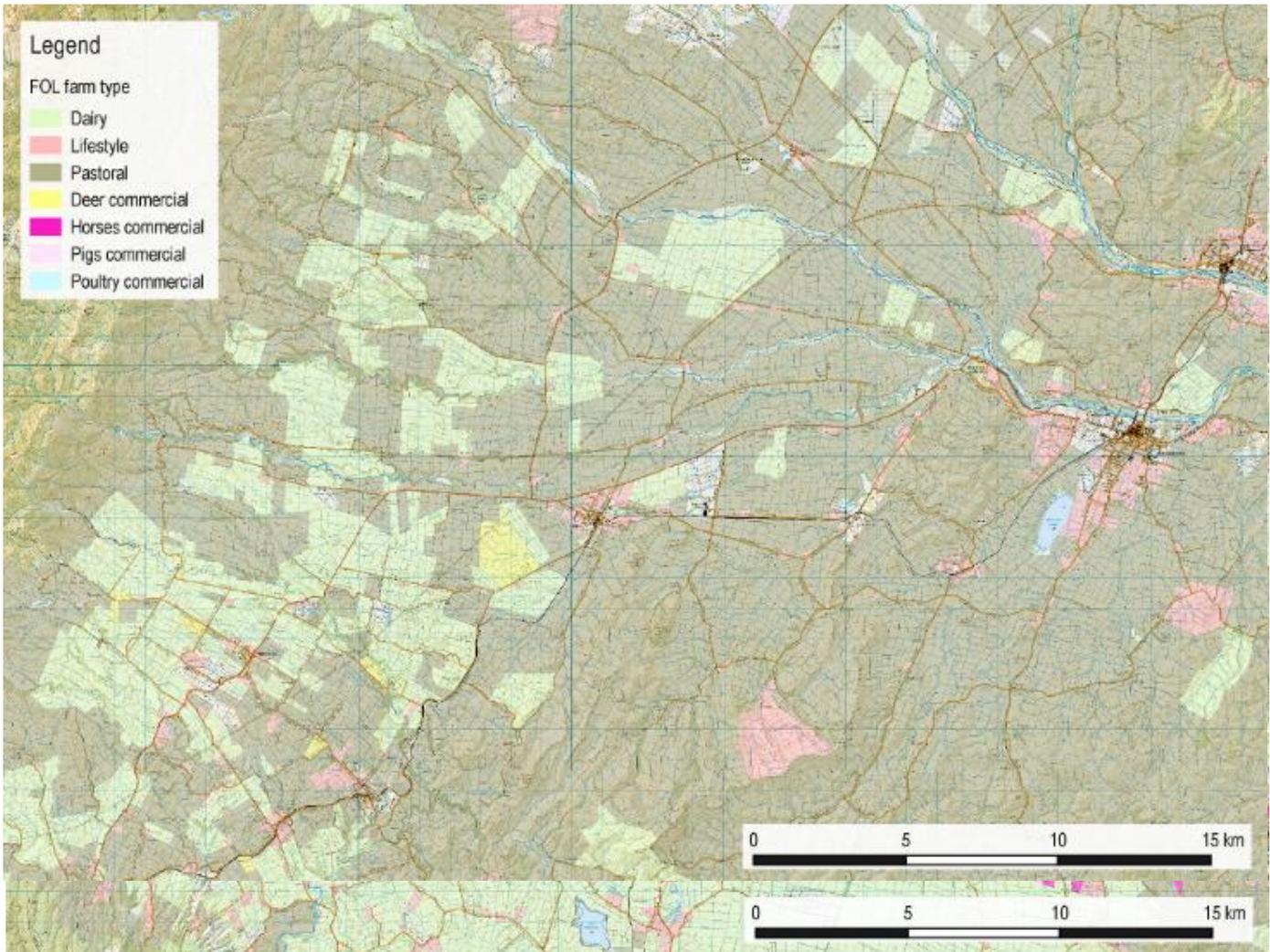


Figure 14.9: Rural survey region (Central Hawkes Bay)

References for Appendix 4:

- Anonymous, 2013. Eurostat: Statistics Explained [WWW Document]. URL <http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:LSU>
- Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688. doi:10.1111/j.1365-2699.2006.01584.x
- Catley, A., Alders, R.G., Wood, J.L.N., 2012. Participatory epidemiology: Approaches, methods, experiences. *Vet. J.* doi:10.1016/j.tvjl.2011.03.010
- Cutler, D.R., 2007. Random Forests for Classification in Ecology. *Ecology* 88, 2783–2792. doi:10.1890/07-0539.1
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography (Cop.)*. 29, 129–151. doi:10.1111/j.2006.0906-7590.04596.x
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009. doi:10.1111/j.1461-0248.2005.00792.x
- 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.
- 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.* 0169. doi:10.1080/00480169.2015.1117955
- Jost, C.C., Mariner, J.C., Roeder, P.L., Sawitri, E., Macgregor-Skinner, G.J., 2007. Participatory epidemiology in disease surveillance and research. *Rev. Sci. Tech.* 26, 537–549.
- MacPherson, L., 2015. Agricultural Production Statistics June 2014. Wellington, New Zealand.
- Mariner, J.C., Hendrickx, S., Pfeiffer, D.U., Costard, S., Knopf, L., Okuthe, S., Chibeu, D., Parmley, J., Musenero, M., Pisang, C., Zingesser, J., Jones, B. a, Ali, S.N., Bett, B., McLaws, M., Unger, F., Araba, a, Mehta, P., Jost, C.C., 2011. Integration of participatory approaches into surveillance systems. *Rev. Sci. Tech. Int. Off. Epizoot.* 30, 653–659.
- Quantum GIS Development Team, 2016. Quantum GIS Geographic Information System.
- Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aiotti, L., Hay, S.I., Gilbert, M., 2014. Mapping the Global Distribution of Livestock. *PLoS One* 9, e96084.
- Sanson, R., Pearson, A., 1997. Agribase - a national spatial farm database, in: Eighth IVSEE, Paris. pp.

12.12.1 – 12.16.3.

Appendix 5: Overview of a workshop held at University of Melbourne on 29 September 2015 regarding Lifestyle Sector estimation

Attendees:

Mary Van Andel (MPI)
Daan Vink (MPI)
Chris Jewell (Massey University)
Marta Hernandez-Jover (Charles Sturt University)
Ruth Beilin (University of Melbourne)
Mark Stevenson (University of Melbourne)
Rodrigo Estevez (Previously CEBRA, via skype)
Tom Kompass (CEBRA)
Susie Hester (CEBRA)
Andrew Robinson (CEBRA)
Mark Burgman (CEBRA)
Tracey Hollings (CEBRA)

Workshop summary

The workshop opened with three short presentations to provide project context and background information to participants. The first, by Daan Vink (MPI), outlined the aims of the project, the project components and necessary details for participants including an overview of the available livestock databases. Participants raised concerns about the lack of a clearly defined definition of “lifestylers” and the difficulty in assigning one given the diversity in the types and characteristics of small farm holdings.

Surveys of small farm holders have been conducted in Australia and the results of the survey findings were presented by Marta Hernandez-Jover (Charles Sturt University). The survey questions were centred around landholders biosecurity procedures, where they seek information, their response to animal health problems as well as demographic information of the farm owners. The survey response was between 10-20% on ~5,000 sent out via post or email.

A short summary of the remote-sensing and GIS project component for the development of stochastic spatial models was also given (Tracey Hollings, CEBRA). The predictor data available and spatial modelling approach were discussed.

Following these initial presentations participants broke into two working groups; one group to discuss the spatial models, and the second group to discuss survey design.

Outcomes of the spatial working group:

A small subset of data in at least two separate regions will be used to investigate the efficacy of several statistical modelling approaches. The approaches are the K Nearest Neighbour, GLM, Random Forests and Bayes nets. A measure of biomass obtained from AgriBase will be used to fit the models and then predictions disaggregated to the species level. Models for the number of animals for each species at the farm-level will also be compared. The desired output is a seven species dimensional distribution of each farm; initially data will just be run for sheep and cattle due to the large amount of data available. Regional stratifications will be incorporated based on Mary's work with the livestock databases. Spatial correlation of the predictor variables will be tested for. The accuracy will be measured on the 3km buffer that would be used in the quarantine response in the event of an FMD incursion.

Outcomes of the survey working group:

Surveys will be used to validate spatial models in the 3km buffer zones. The favoured approach was to visit farms in person and this was considered feasible given the size of the target area. Approaching vets, homekill businesses, shearers, Trademe and other individuals/organisations were considered good options. For example, taking maps into local vets and asking them questions about specific areas of interest could provide valuable information and remove issues of privacy. Some assessment of the value to government of livestock estimation information and/or cost-benefit analysis should be given in the final report.

Appendix 6: NAIT Analysis to inform the parameterisation of AADIS

Data:

2 000 455 movements recorded in the NAIT data between July 1st 2012 and Feb 1st 2016. Each movement is structured as a “to and from” NAIT number (along with associated FOL ids) and a “to and from” production type. People in charge of animals (PICA’s) select their production type based on their primary production. These production types include 30 production types (which include as 3 separate items NULL, none and “blank”) to allow PICAs to describe their primary production type. Sale yards and meat processing plants are able to leave their production type blank accounting for the column on the far right of the movements to plot in Figure 14.10 and the movements from plot in Figure 14.11.

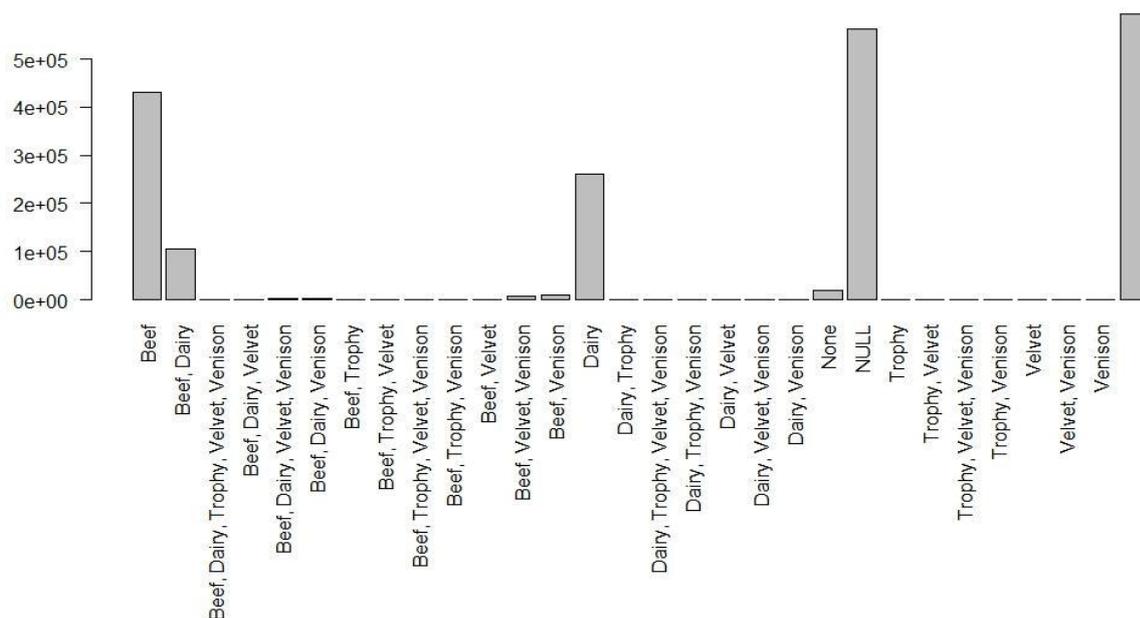


Figure 14.10: Counts of observations of movements to production types with counts of NAIT numbers on the Y axis and production description on the X axis

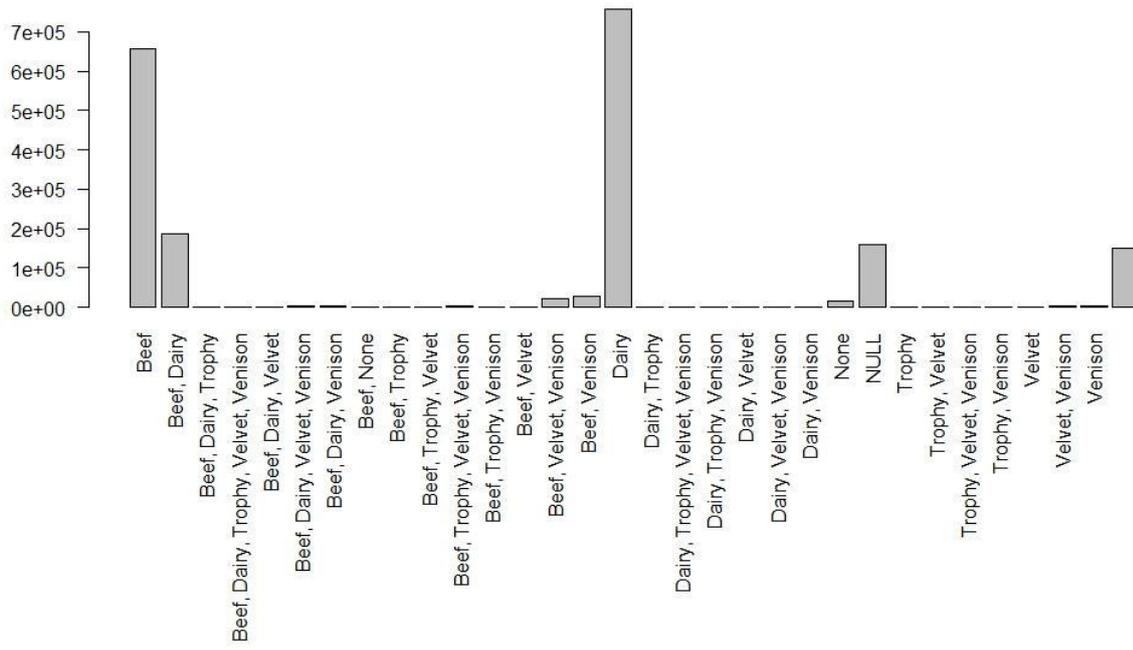


Figure 14.11: Movements *from* production types with counts of NAIT numbers on the Y axis and production description on the X axis

Fortunately, large numbers of PICAs regard their primary production type as dairy or beef. This makes it possible to examine movement patterns between and within production types, however both of these categories are secondary to the large numbers of blank, NULL and none choices allowed by NAIT.

When examined by year counts of movements are as follows:

Table 14.8: Counts of animal movements by year as present in National Animal Identification and Tracing (NAIT) Scheme database from its start to the time of analysis.

Year	Count of movements	Notes
2012	256678	From July 2012 to end Dec 2012
2013	538005	
2014	603753	
2015	565813	

2016	36206	From 1 Jan 2016 to end Feb 2016
------	-------	---------------------------------

Movement numbers do not show an obvious increase over time.

Null, blank and none production types:

The percentage of picas recording NULL, blank or none as their production type were examined by year. It appears that the option for the users of leaving this field blank has been allowed since 2014 as prior to this they were allowed to either record “None” or “NULL” but leaving it totally blank was not an option. Saleyards and Meat processors understandably have no production type, however it is not clear why a mix of having a blank field, None and NULL is desirable and why this has been allowed to change since 2014 (Table 14.9). It appears that NULL is no longer used as of 2014 but the use of None vs a blank field amongst PICA types cannot be made sense of as in 2015 it seems that farms and lifestyle blocks were allowed the option of either of these and other PICA types used blank.

Table 14.9: *Counts of production types by year which are unspecified as present in National Animal Identification and Tracing (NAIT) Scheme database from its start to the time of analysis.*

From_prod	2012	2013	2014	2015	2016
None	1272	3644	5183	5059	287
NULL	46967	86502	25858	0	0
Blank	0	0	63372	81103	4479

to_prod	2012	2013	2014	2015	2016
None	2037	4000	6288	6446	263
NULL	139933	299689	122395	0	0
Blank	0	0	217249	350246	26870

Table 14.10: *Break down of those National Animal Identification and Tracing (NAIT) Scheme numbers that had “None” or blank in 2015 as their production types*

	None	Blank
Farm or lifestyle block	6446	28
Game estate/safari park	0	21
Import/export transitional facility	0	94
Meat processing plant	0	262773
Other	0	1325
Quarantine area	0	212
Sale yard	0	85333
Show/rodeo ground	0	460

Beef, dairy and deer production types:

For the purposes of disease modelling, movement patterns within a production sector are of interest. From the available data production types “Beef”, “Dairy” and an aggregate of all deer based production types that did not include any other species were examined.

As each movement originates with a production type and goes to a production type, when movements onto a particular type were calculated they were different to movements off of a particular type. Counts of movements by year, by direction and by production type in the NAIT data can be seen in Table 14.17.

Table 14.11: *Counts of movements to and from production types of interest by year as present in National Animal Identification and Tracing (NAIT) Scheme database from it’s start to the time of analysis.*

From	2012	2013	2014	2015	2016
dairy	94302	205032	233121	213043	13169
beef	84947	173318	195322	189239	12744
deer	48	2672	3894	3480	408
to					
dairy	27830	79427	98011	55624	1229
beef	67755	119351	120911	116091	6094

Of interest for disease modelling is how movements go from each production type between farms and sale yards and meat processing facilities (a dead end for disease transmission) and how these movements change over a year in terms of distance and frequency. Regional and monthly differences are relevant, however a balance must be struck between detail and parsimony in modelling. Currently the New Zealand Standard Model in ISP uses an average over all regions and does not include seasonality – it is possible that both of these could be incorporated in time.

As a starting point for AADIS parameterisation, region is excluded and time varying covariates are only included for season although both of these could be modified in time.

Movement frequency:

The simple sale yard mechanism in AADIS requires movements between farms and movements between farms and sale yards to be estimated together. For reasons unrelated to modelling, it is helpful to have an understanding of movements between each of the place types, these are shown in table 9 (farm, sale yard and processing plant). Estimates of frequency are by farm by day.

Table 14.12: Values used in the Australian Animal Disease Spread Model (AADIS) for movements onto dairy, beef and deer farms. Values are frequencies of animal movements onto farms per day.

	Overall	Spring	Summer	Autumn	Winter
Dairy	0.028	0.029	0.018	0.026	0.037
Beef	0.005	0.007	0.004	0.005	0.004
Deer	0.003	0.001	0.005	0.003	0.002

Table 14.13: Values used in the Australian Animal Disease Spread Model (AADIS) for movements off dairy, beef and deer farms. Values are frequencies of animal movements off farms per day.

	Overall	Spring	Summer	Autumn	Winter
Dairy	0.03	0.032	0.019	0.028	0.040

Beef	0.008	0.01	0.006	0.008	0.008
Deer	0.004	0.003	0.006	0.005	0.004

Table 14.14: Movement frequencies per day from farms to each of the Person In Charge of Animals (PICA) types and production types by season

To farm	Overall	Spring	Summer	Autumn	Winter
Dairy	0.02	0.02	0.012	0.02	0.07
Beef	0.003	0.004	0.002	0.003	0.003
Deer	0.003	0.0009	0.005	0.003	0.002
To sale yard					
Dairy	0.008	0.01	0.006	0.007	0.009
Beef	0.002	0.004	0.002	0.003	0.003
Deer	0.0002	0.0009	0.005	0.003	0.002
To slaughter					
Dairy	0.018	0.015	0.022	0.027	0.011
Beef	0.006	0.006	0.008	0.007	0.005
Deer	0.009	0.011	0.011	0.011	0.006

Table 14.15: Movement frequencies per day to farms from farms and sale yards by each of the production types by season as present in National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis.

From farm	Overall	Spring	Summer	Autumn	Winter
Dairy	0.02	0.022	0.013	0.021	0.03

Beef	0.005	0.006	0.003	0.004	0.005
Deer	0.004	0.002	0.005	0.004	0.004
From sale yard					
Dairy	0.002	0.002	0.0007	0.002	0.002
Beef	0.004	0.006	0.003	0.003	0.005
Deer	0.0004	0.002	0.005	0.004	0.003

Movement distance:

The simple sale yard mechanism in AADIS requires movement distances between farms and movement distances between farms and sale yards to be estimated together (results used can be seen in table 11). For reasons unrelated to modelling, it is helpful to have an understanding of movement distances between each of the place types, these are shown in table 12(farm, sale yard and processing plant). Estimates of distance are shown as a mode, and maximum suitable for a beta pert distribution. The beta pert distribution is an acceptable approximation of the distribution as shown in Figure 14.12.

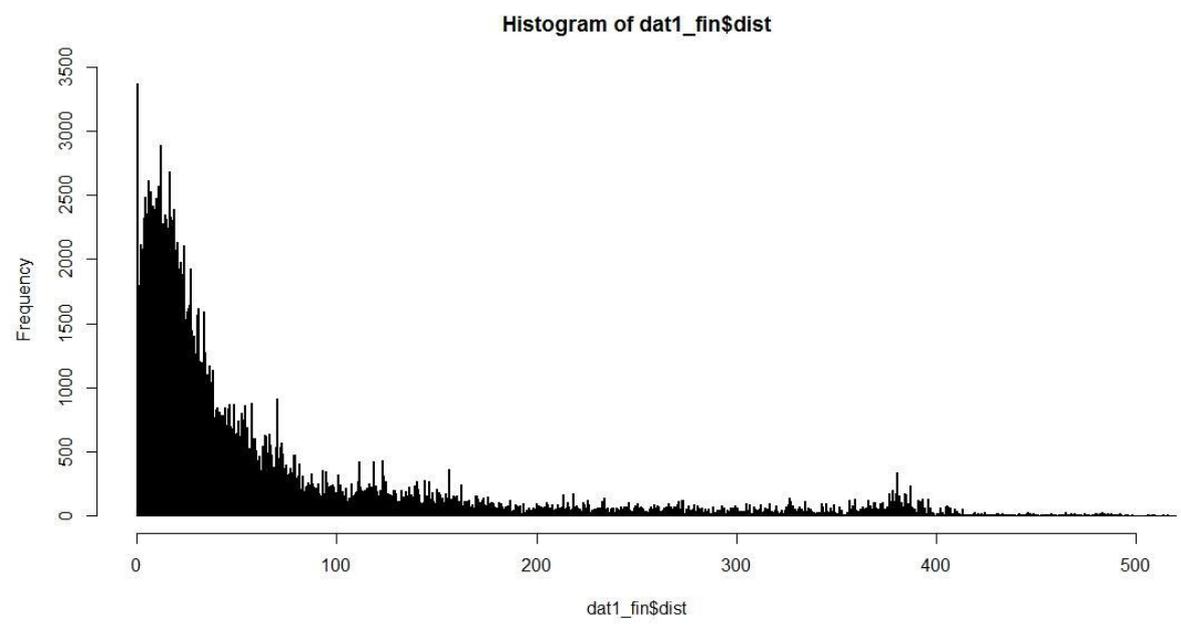


Figure 14.12: *Movements off dairy production properties. Distance is shown on the x axis and frequency on the y axis. Aside from the peak at 0 which are associated with registration, the distribution adequately approximates a beta pert.*

Due to the many to many relationships in the NAIT and FOL datasets, and assumption was made to simplify the relationships. In the event of multiple FOL numbers being associated to a single NAIT number, the first FOL number listed was regarded as the point to measure distance from. According to the NAIT act, multiple FOL numbers associated with a NAIT number must be within 20km of each other. Any bias associated with this work around is expected to be non-differential, affecting all strata equally.

Table 14.16: *Australian Animal Disease Spread Model (AADIS) distance parameters as present in National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis. Distances shown in km for the three species included in the NAIT database.*

	Summer mode	Summer max	Autumn mode	Autumn max	Winter mode	Winter max	Spring mode	Spring max
deer	49	1052	33	704	36	779	22	766
dairy	14	506	13	657	17	492	17	317
beef	15	400	15	428	13	469	17	338

Table 14.17: *Australian Animal Disease Spread Model (AADIS) distance parameters as present in National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis. Distances shown in km for the three species included in the NAIT database for each type of movement required to parameterise the model.*

	all year		Spring		summer		autumn		v
	mode	max	mode	Max	mode	max	mode	max	
movements off dairy farms	14.43	434.5	12.72	382.7	21.98	404.6	19.45	473.89	13.45
movements off dairy farms to other farms	14.27	653.77	13.38	421.81	6.91	701.09	17.49	759.24	12.51
movements off dairy farms to sale	19.62	91.47	18.62	89.48	17.66	96.08	19.06	96.79	13.29
movements off dairy farms to processing plants	19.32	397.03	22.34	389.32	29.21	386.72	25.89	437.8	30.56
movements onto dairy farms	12.8	718.75	12.19	497.99	16.78	719.51	19.46	749.98	8.79

movements onto dairy farms from other farms	11.38	729.53	10.68	527.93	14.41	721.88	17.42	759.1	6.95
movements onto dairy farms from sales	18.03	271.48	16.18	240.93	21.72	264.27	13.67	315.26	16.98
movements off beef farms	21.55	428.28	18.45	391.56	15.38	415.73	16.07	462.19	15.51
movements off beef farms to other farms	10.54	545.49	21.45	453.31	17.99	522.1	16.69	641.54	14.87
movements off beef farms to sale	19.31	118.95	11.85	114.52	17.24	115.07	17.11	122.37	17.07
movements off beef farms to processing plants	27.94	439.72	33.63	402.05	34.92	417.12	33.21	472.02	31.25
movements onto beef farms	17.04	389.24	15.57	342.13	13.33	470.8	12.73	486.12	15.11
movements onto beef farms from other farms	13.21	508.42	11.27	437.92	7.88	585.63	17.43	638.06	10.69
movements onto beef farms from sales	17.7	257.36	16.29	242.53	14.47	244.26	13.94	321.75	17.25
movements off deer farms	29.85	651.48	28.5	360.29	64.03	916.49	32.06	347.31	43.34
movements off deer farms to other farms	41.07	989.67	30.68	795.79	51.88	1055.97	35.65	703.86	40.59
movements off deer farms to sale	14.9	111.92	14.97	169.47	12	65.67	26.07	59.84	14.65
movements off deer farms to processing plants	32.18	330.41	30.96	332.89	67.19	330.41	32.21	293.96	40.89
movements onto deer farms	25.41	900.94	24.69	711.72	36.5	1042.7	25.88	386.19	41.29
movements onto deer farms from other farms	26.16	911.9	26.59	729.5	39.92	1045.73	24.71	490.75	41.69
movements onto deer farms from sales	39.94	206.28	41.52	189.15	19.54	172.37	110.82	223.63	46.22

Gypsy day:

While movements do seem to increase in frequency in winter for with dairy production type, the magnitude and extent of gypsy day movements is not immediately visible from the NAIT data.

An analysis based on consignment size was performed to isolate the effect and examine it by region. Further data limitations were identified at this point as there are a large number of movements that have only a single animal recorded in them. One of the characteristics of gypsy day is that a whole herd is moved to a new farm at the end of May, start of June. These large numbers of single animal “consignments” are skewing the analysis are excluded from this point onwards.

Cut points of 150 and 250 animals per consignment were considered for whole herd movements of dairy cattle. Results are normalised by the number of months present in the data set for each season (assuming 3 months per season as is the Australia/NZ convention).

Table 14.18: *Counts of animal movements stratified by season when more than 150 animals and more than 250 animals were moved at one event as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis.*

	3 months each				
	spring	summer	autumn	winter	
150 cut	0	75922	35504	43045	93231
point	1	676	776	1431	2078
250 cut	0	76350	36086	43914	94315
point	1	248	194	562	994
months in dataset		12	11	9	11
normalised to a year (cp 150)		169	212	477	567
normalised to a year (cp 250)		62	53	187	271

The median size of a dairy herd according to AgriBase is 300+ animals, so it would appear that there are less than 270 short range movements per year at the worst case scenario associated with gypsy day. While the whole herd moves, the disease risk might actually be less from direct contact as herds are not mixing although risk of infected conveyors i.e. trucks would appear to increase as frequency increases. Distances of these movements (cut point for number of animals moved 250 head) are as follows by season.

Table 14.19: *Descriptive statistics of movement distances for consignments of more than 250 animals in each season as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis.*

Season	1 st Quar	Median	Mean	Mode	3 rd Quar	Max
Spring	0.399	11.74	29.18	2.66	41.96	722
Summer	0	20.85	50.49	5.5	53.32	413.10
Autumn	9.18	29.26	64.25	8.5	65.35	873.2
Winter	2.036	16.92	48.67	3.12	48.39	1189

While an increase in distance travelled by large consignments (>250 animals) is visible in the data, the number of movements does not seem to be large. As a point of clarification – NAIT numbers do not go with farmers unless the movement happens within 20km – so if on GD the farmer moved less than 20km he could just add his FOL number to his existing NAIT number – showing no movement in the NAIT data. If the new FOL is >20km away then a new NAIT number is created.

Furthermore, consignments in NAIT have nothing to do with what animals travelled on what trucks, in this case the capacity of the trucks has no bearing on the number of consignments – cut points do better to reflect the size of the herds moved (hence 150 and 250 as conservative estimates of dairy farms).

To examine the GD effect by region and by month, a cut point was created at 100 animals per consignment to be included as a whole herd movement. This is conservative and creates a larger number of data points. There are 4943 movements in this reduced dataset (Figure 14.10). Of these, 3988 are within their region of origin and 955 of these move out of their region of origin. Movement distances within region can be seen in figure 3. Note that the y axis differs between figure 2 and 3 and figure 4/5

Simply put, one of the intuitive ways to interpret figures 4 and 5 is to see how many regions are involved in long range, large consignment movements in each month – it is easy to see that in May and June, more regions are involved in long range movements than in other times of the year. This indicates the presence of GD activity in NAIT.

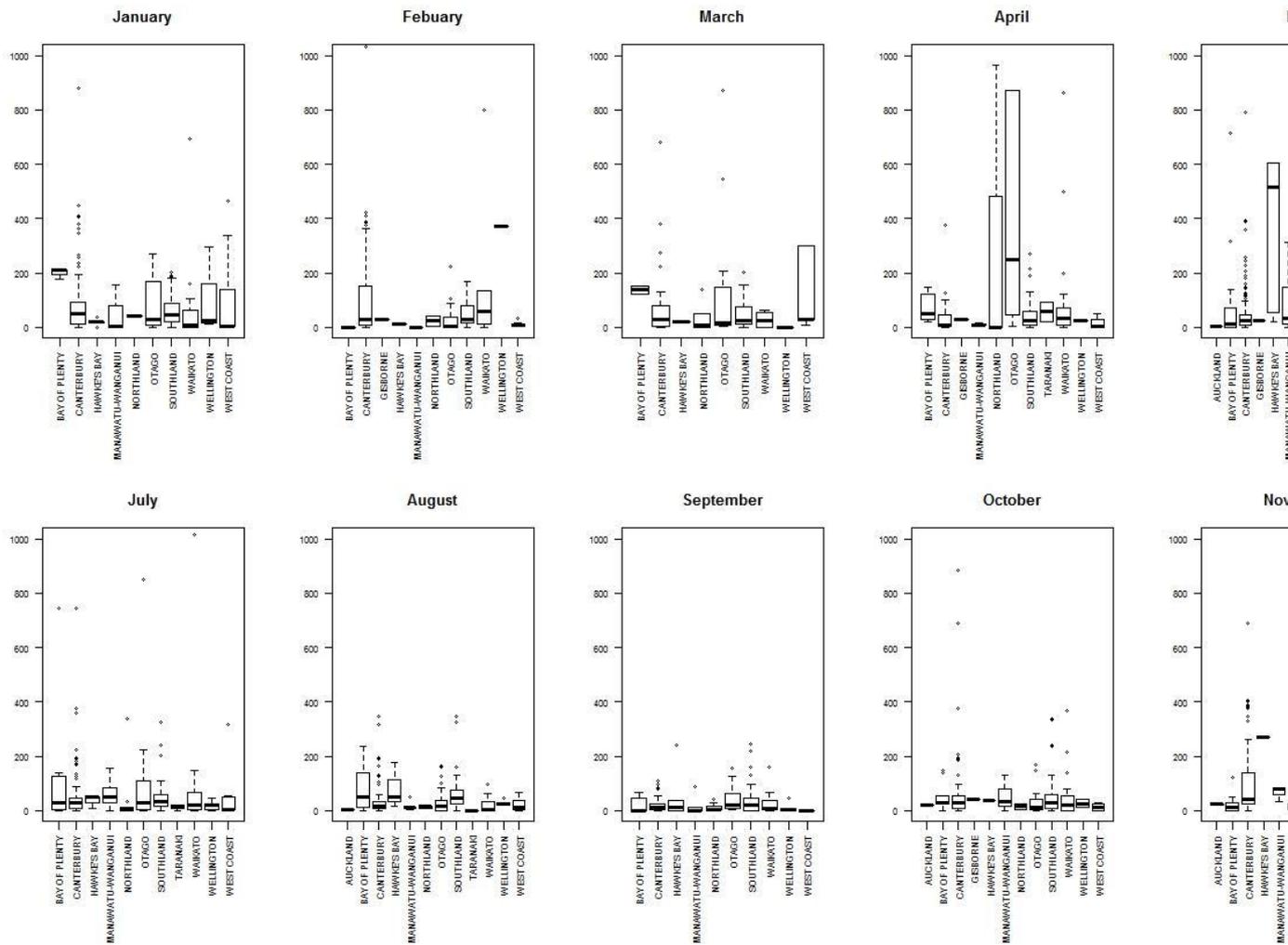


Figure 14.13: Distance travelled during gypsy day movements (in this case defined as a consignment of more than 100 animals moved together) by month and region in km. Note the scale of the y axis is 1000 km. This graph shows all ~5000 movements present in the data as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis..

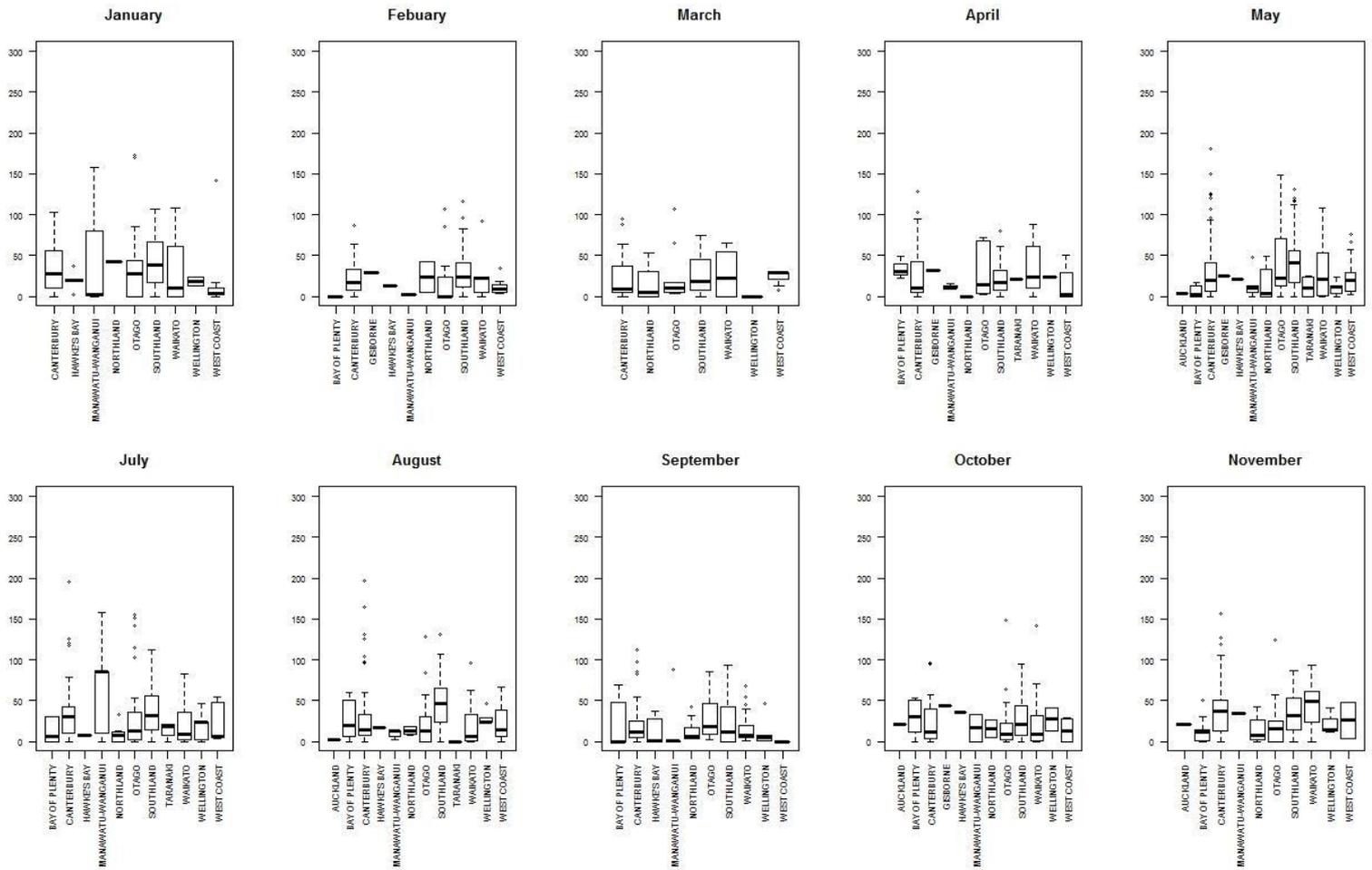


Figure 14.14: Distance travelled during gypsy day movements (in this case defined as a consignment of more than 100 animals moved together) by month and region in km. Note the scale of the y axis is 300 km. These movements are those that go within region only as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis.

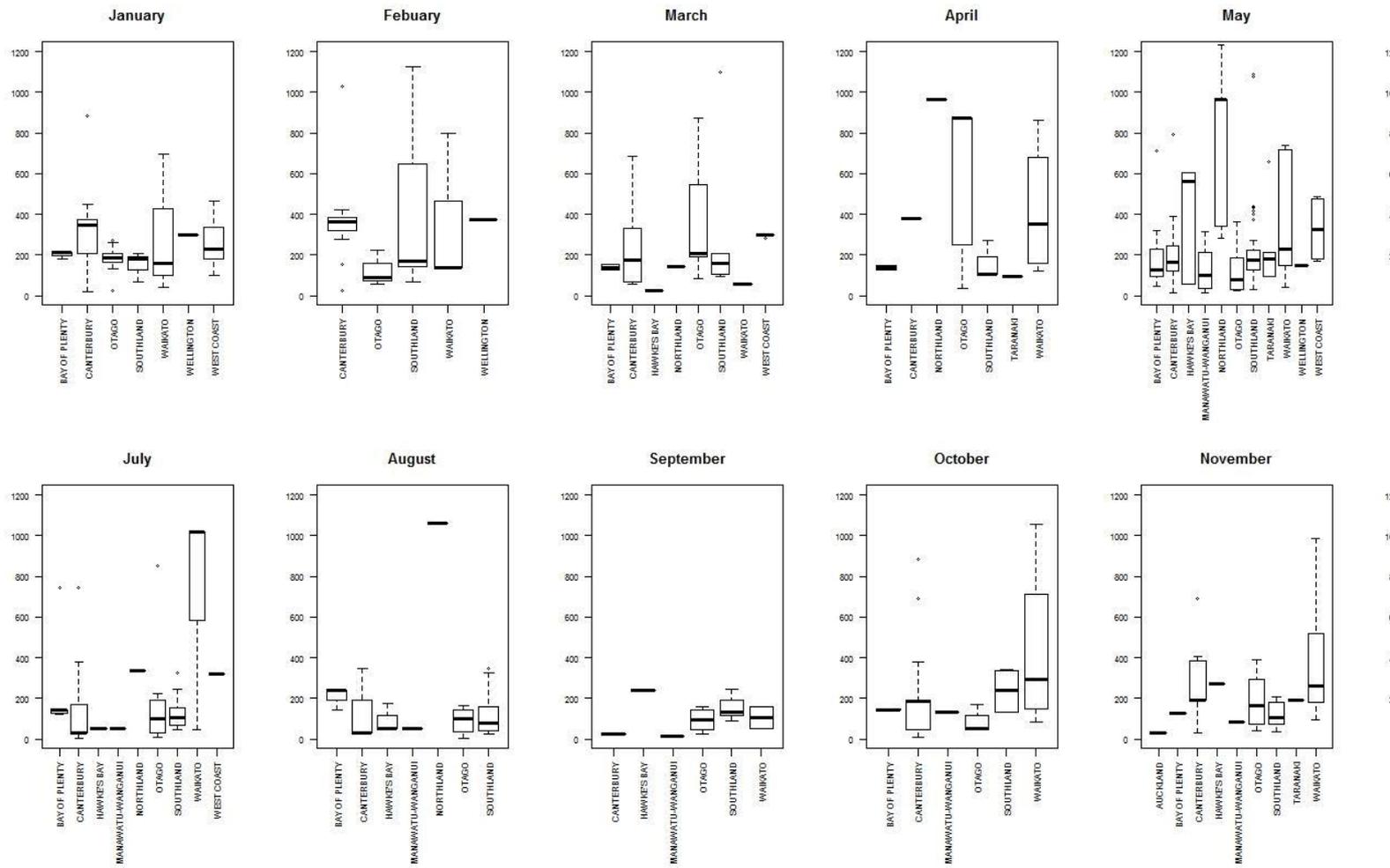


Figure 14.15: Distance travelled during gypsy day movements (in this case defined as a consignment of more than 100 animals moved together) by month and region in km. Note the scale of the y axis is 1200 km. These movements are those that leave their region of origin for another region as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis..Movements are plotted by the region that they originate in. For instance in June, many movements leave the Waikato and some move as far as 1100km.

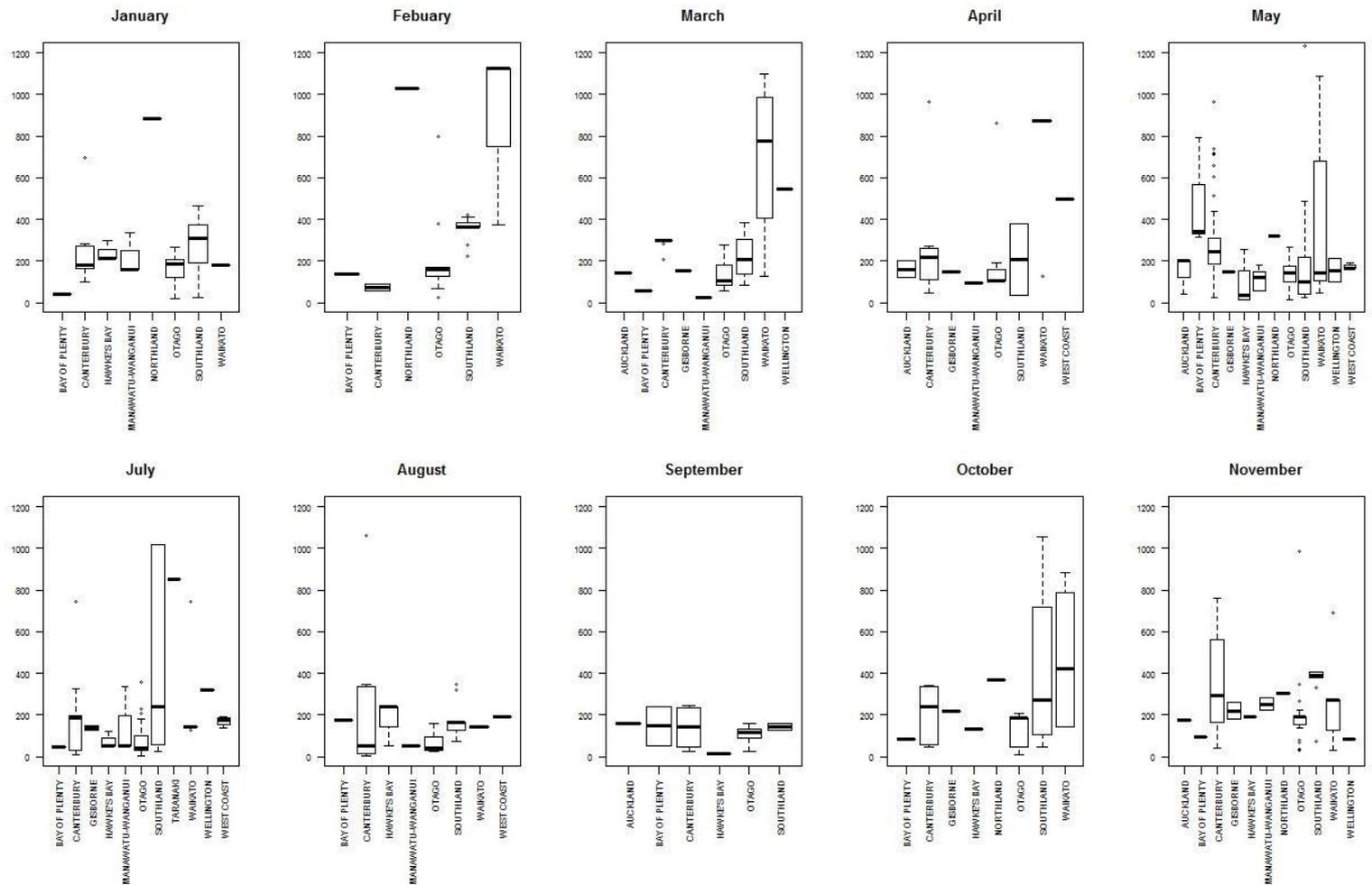


Figure 14.16: Distance travelled during gypsy day movements (in this case defined as a consignment of more than 100 animals moved together) by month and region in km. Note the scale of the y axis is 1200 km. These movements are those that leave their region of origin for another region. Movements are plotted by the region that they go to. For instance in June, many movements are made to Southland and some move from as far away as 1000km. These graphs represent the data as recorded in the National Animal Identification and Tracing (NAIT) Scheme database from it's start to the time of analysis.

Appendix 7: Code snippet used to calculate the mode used in creation of AADIS database table

Calculation of the mode of number of animals of farms

```
modefunc <- function(x){ tabresult <- tabulate(x) themode <-  
which(tabresult == max(tabresult))  
  if(sum(tabresult == max(tabresult))>1) themode <- NA  
  return(themode)  
}  
#rows  
apply(mat1, 1, modefunc)  
#columns  
  
apply(mat1, 2, modefunc)
```

Appendix 8: Livestock database comparison

Table 14.20: Mean animal counts per premises by species, farm type, and district, based on AgriBase data

Region	District	Enterprise type	Dairy	Beef	Deer	Sheep	Pigs
Auckland	Auckland	PLVSTCK	133	0	0	641	0
		LIF	3	0	0	7	0
		PIGB	36	0	0	15	40
	Franklin	PLVSTCK	67	2	4	136	0
		LIF	6	0	0	4	0
		DAIRY	14	277	1	9	17
		GRADRY	8	9	0	2	0
		PIGF	20	0	0	4	18
		PIGB	24	3	0	1	806
	Manukau	PLVSTCK	52	2	3	142	0
		LIF	4	0	0	7	0
		DAIRY	36	216	0	5	0
		GRADRY	7	7	0	15	0
		PIGF	25	0	0	0	10
	Rodney	PLVSTCK	58	1	7	111	0
		LIF	5	0	0	5	0
		DAIRY	29	271	0	21	0
		GRADRY	6	6	0	4	0
		PIGF	4	0	0	0	478
		PIGB	53	0	0	0	332
	Waitakere	PLVSTCK	18	0	1	21	0
		LIF	4	0	0	5	0
		DAIRY	0	94	0	0	0
		GRADRY	1	0	0	5	0
PIGF		6	0	0	0	2	
PIGB		36	0	0	192	100	
Bay of Plenty	Opotiki	PLVSTCK	70	3	5	122	1
		LIF	7	0	0	3	1
		DAIRY	29	400	0	42	1
		GRADRY	11	8	5	1	0
		PIGF	2	0	0	0	6
		PIGB	18	0	0	2	28
	Rotorua	PLVSTCK	73	7	65	409	0

		LIF	6	0	0	6	0
		DAIRY	7	449	4	18	0
		GRADRY	30	125	2	102	0
		PIGF	5	3	0	9	8
		PIGB	2	0	0	5	23
	Western Bay of Plenty	PLVSTCK	55	1	19	111	0
		LIF	6	0	0	4	0
		DAIRY	15	425	0	12	2
		GRADRY	6	18	1	15	0
		PIGF	5	0	1	21	525
		PIGB	1	0	0	8	346
	Whakatane	PLVSTCK	82	11	9	127	2
		LIF	7	1	0	2	1
		DAIRY	25	380	1	28	1
		GRADRY	11	71	0	1	0
		PIGF	8	39	0	11	578
		PIGB	80	0	0	7	364
Canterbury	Ashburton	PLVSTCK	116	1	87	1149	7
		LIF	5	0	0	6	0
		DAIRY	23	951	2	103	0
		GRADRY	11	72	0	10	3
		PIGF	113	0	0	13	867
		PIGB	31	0	0	164	799
	Christchurch	PLVSTCK	74	0	8	589	14
		LIF	4	0	0	6	0
		DAIRY	59	223	0	447	6
		GRADRY	12	11	3	23	0
		PIGF	4	0	0	0	780
		PIGB	5	0	0	319	2484
	Hurunui	PLVSTCK	129	7	50	1776	2
		LIF	5	0	0	8	0
		DAIRY	18	860	0	12	0
		GRADRY	5	70	6	51	0
		PIGF	0	0	0	0	28720
		PIGB	0	0	0	11	138
	Kaikoura	PLVSTCK	139	0	39	1104	0
		LIF	4	0	0	3	0
		DAIRY	15	469	5	11	16
		GRADRY	16	7	4	9	0
		PIGB	147	34	0	130	296
	Mackenzie	PLVSTCK	166	8	155	2617	0
		LIF	3	0	0	10	0

		DAIRY	1	745	0	73	1
		GRADRY	0	0	0	27	0
Selwyn		PLVSTCK	41	5	44	705	5
		LIF	4	0	0	7	0
		DAIRY	13	845	5	11	14
		GRADRY	10	3	0	31	0
		PIGF	5	0	0	18	1274
		PIGB	17	0	0	237	486
Timaru		PLVSTCK	60	3	110	649	2
		LIF	4	0	0	8	0
		DAIRY	11	763	0	6	2
		GRADRY	10	6	6	107	0
		PIGF	202	0	0	73	1343
		PIGB	31	0	0	16	420
Waimakariri		PLVSTCK	35	2	30	271	6
		LIF	4	0	0	7	0
		DAIRY	15	656	2	124	47
		GRADRY	13	20	1	39	0
		PIGF	45	0	0	4	513
		PIGB	10	0	0	7	1070
Waimate		PLVSTCK	107	12	98	1488	1
		LIF	5	0	0	6	0
		DAIRY	19	933	0	1	0
		GRADRY	12	54	0	87	0
		PIGF	46	0	0	21	582
		PIGB	37	0	0	31	733
Waitaki		PLVSTCK	104	3	41	1209	1
		LIF	4	0	0	9	0
		DAIRY	20	830	0	77	0
		GRADRY	28	44	0	59	0
		PIGF	53	0	0	120	745
		PIGB	40	0	0	10	599
Gisborne	Gisborne	PLVSTCK	304	2	25	1954	1
		LIF	4	0	0	8	1
		DAIRY	37	258	0	216	8
		GRADRY	43	14	0	11	0
		PIGF	2	0	0	18	10
		PIGB	4	0	0	20	44
Hawke's Bay	Central Hawke's Bay	PLVSTCK	223	0	48	1764	0
		LIF	5	0	0	8	0
		DAIRY	12	765	1	58	0
		GRADRY	22	74	0	12	0

		PIGF	34	1	0	7	1287
		PIGB	0	0	0	0	30
Hastings		PLVSTCK	160	2	29	1233	0
		LIF	5	0	0	6	0
		DAIRY	41	771	0	79	0
		GRADRY	12	33	0	8	0
		PIGF	7	0	0	7	763
		PIGB	5	0	0	10	251
	Hastings city		PLVSTCK	20	0	4	151
		LIF	4	0	0	7	0
		GRADRY	1	0	0	13	0
Napier		PLVSTCK	110	0	0	1915	0
		LIF	5	0	0	5	0
Wairoa		PLVSTCK	271	0	33	1835	1
		LIF	8	0	0	6	1
		DAIRY	4	360	4	50	1
		GRADRY	0	16	0	0	0
		PIGB	3	1	0	26	23
Manawatu-Wanganui	Horowhenua	PLVSTCK	56	1	9	173	11
		LIF	6	0	0	3	0
		DAIRY	19	493	0	7	5
		GRADRY	11	33	0	8	0
		PIGF	17	0	0	0	1113
		PIGB	13	0	0	2	784
	Manawatu	PLVSTCK	87	2	27	698	0
		LIF	6	1	0	7	0
		DAIRY	24	372	8	83	7
		GRADRY	14	28	4	23	0
		PIGF	63	12	0	3126	2135
		PIGB	14	0	0	13	488
	Palmerston North	PLVSTCK	31	0	7	204	0
		LIF	4	0	0	6	0
		DAIRY	3	296	0	5	1
		GRADRY	3	26	0	0	0
		PIGB	10	0	0	0	1936
	Rangitikei	PLVSTCK	178	5	36	2055	7
LIF		6	2	0	6	1	
DAIRY		31	384	2	111	0	
GRADRY		14	55	0	154	0	
PIGF		12	0	0	40	12	
PIGB		0	0	0	70	380	
Ruapehu	PLVSTCK	196	3	37	1858	1	

		LIF	4	0	0	10	1
		DAIRY	75	479	0	115	1
		GRADRY	63	8	0	165	0
		PIGF	0	0	6	0	26
		PIGB	18	0	0	2	152
	Tararua	PLVSTCK	147	5	13	1948	0
		LIF	6	0	0	8	0
		DAIRY	15	375	0	81	0
		GRADRY	10	61	0	43	0
		PIGF	10	0	0	15	13
		PIGB	10	0	0	22	56
	Wanganui	PLVSTCK	109	1	21	1193	1
		LIF	5	0	0	7	0
		DAIRY	44	374	1	46	0
		GRADRY	14	49	0	37	0
		PIGF	140	308	0	2	1169
		PIGB	25	0	0	15	34
Marlborough	Marlborough	PLVSTCK	92	2	18	927	1
		LIF	3	0	0	7	0
		DAIRY	14	329	1	37	23
		GRADRY	8	12	0	46	0
		PIGF	0	105	0	50	1040
		PIGB	0	0	0	14	41
Nelson	Nelson	PLVSTCK	38	0	4	273	0
		LIF	3	0	0	6	0
		DAIRY	9	148	0	62	0
		GRADRY	8	0	0	2	0
Northland	Far North	PLVSTCK	126	5	1	116	1
		LIF	7	1	0	2	0
		DAIRY	44	295	0	11	1
		GRADRY	10	9	0	4	0
		PIGF	11	0	0	0	6
		PIGB	19	0	0	0	25
	Kaipara	PLVSTCK	173	8	6	246	1
		LIF	7	0	0	3	0
		DAIRY	40	337	0	24	0
		GRADRY	14	11	0	1	0
		PIGF	1	0	0	2	696
		PIGB	60	0	0	0	131
	Whangarei	PLVSTCK	84	3	2	69	0
		LIF	7	0	0	3	0
		DAIRY	21	393	1	2	0

		GRADRY	5	12	0	2	0	
		PIGB	14	0	0	11	34	
Otago	Central Otago	PLVSTCK	105	1	75	3032	0	
		LIF	3	0	0	8	0	
		DAIRY	4	771	0	1	0	
		GRADRY	0	21	0	68	0	
		PIGF	44	0	0	450	0	
	Clutha	PLVSTCK	91	8	51	2354	0	
		LIF	3	0	0	10	0	
		DAIRY	10	618	0	47	1	
		GRADRY	69	84	0	77	0	
	Dunedin	PLVSTCK	58	1	38	1005	0	
		LIF	2	0	0	11	0	
		DAIRY	11	361	8	15	0	
		GRADRY	2	12	0	68	0	
		PIGF	8	1	1	84	972	
	Queenstown-Lakes	PLVSTCK	115	0	217	1993	0	
		LIF	3	0	1	5	0	
		GRADRY	0	8	0	4	0	
	Southland	Gore	PLVSTCK	34	10	45	1552	0
LIF			2	0	0	12	0	
DAIRY			4	520	0	109	1	
GRADRY			12	131	0	291	0	
PIGB			7	0	0	146	280	
Invercargill		PLVSTCK	11	0	14	141	0	
		LIF	4	0	0	12	0	
		DAIRY	9	439	3	1	0	
		GRADRY	9	39	2	165	0	
		PIGF	2	0	10	48	9	
Southland		PLVSTCK	69	5	103	1733	0	
		LIF	3	0	1	11	0	
		DAIRY	7	664	5	65	0	
		GRADRY	22	98	4	130	0	
		PIGF	7	0	0	6	7	
		PIGB	0	0	0	102	44	
Taranaki		New Plymouth	PLVSTCK	64	5	6	265	1
			LIF	4	1	0	4	0
	DAIRY		7	249	0	11	2	
	GRADRY		10	25	0	13	0	
	PIGF		2	0	0	16	500	
	PIGB		15	0	0	2	418	
	South Taranaki	PLVSTCK	92	19	6	788	0	

		LIF	5	1	0	3	0
		DAIRY	6	304	0	12	8
		GRADRY	18	34	0	16	0
		PIGF	2	1	0	2	8
		PIGB	6	0	0	0	1260
	Stratford	PLVSTCK	136	8	9	961	0
		LIF	5	1	0	3	0
		DAIRY	3	260	0	15	0
		GRADRY	17	46	0	68	0
		PIGF	30	0	0	4	773
		PIGB	0	0	0	1	38
Tasman	Tasman	PLVSTCK	44	3	25	329	0
		LIF	4	0	0	6	0
		DAIRY	25	447	5	32	0
		GRADRY	15	18	2	5	0
		PIGF	4	0	0	32	9
		PIGB	1	3	0	5	30
Waikato	Hamilton	PLVSTCK	18	0	30	10	0
		LIF	7	1	0	5	0
		DAIRY	13	231	0	0	0
		GRADRY	2	5	0	0	0
	Hauraki	PLVSTCK	78	9	4	226	0
		LIF	7	1	0	2	0
		DAIRY	7	335	0	5	0
		GRADRY	7	47	0	12	0
		PIGF	8	0	0	8	4
	Matamata-Piako	PLVSTCK	66	12	10	158	0
		LIF	7	2	0	2	0
		DAIRY	4	348	1	2	2
		GRADRY	13	76	0	23	0
		PIGF	11	2	0	12	29
		PIGB	7	0	0	6	26
	Otorohanga	PLVSTCK	194	20	22	903	0
		LIF	6	4	0	3	0
		DAIRY	8	420	1	12	26
		GRADRY	29	111	0	15	0
		PIGF	2	0	0	0	0
		PIGB	7	163	0	20	368
	South Waikato	PLVSTCK	124	24	25	219	0
		LIF	6	1	0	2	0
		DAIRY	2	438	0	2	0
		GRADRY	5	70	0	35	0

		PIGF	11	0	0	0	19
		PIGB	4	0	0	0	41
Taupo		PLVSTCK	191	43	163	1422	0
		LIF	5	0	0	6	0
		DAIRY	20	732	0	27	0
		GRADRY	28	207	0	45	0
		PIGF	4	1	2	9	8
		PIGB	6	10	0	2	53
Thames-Coromandel		PLVSTCK	83	7	1	155	1
		LIF	6	1	0	3	0
		DAIRY	16	261	0	14	33
		GRADRY	31	24	0	12	0
		PIGF	19	0	0	89	240
		PIGB	0	0	0	50	38
Waikato		PLVSTCK	120	7	6	313	2
		LIF	7	1	0	3	0
		DAIRY	6	376	0	2	4
		GRADRY	20	22	0	25	0
		PIGF	10	0	0	2	1263
		PIGB	21	0	0	6	1223
Waipa		PLVSTCK	57	11	24	134	0
		LIF	7	1	0	3	0
		DAIRY	4	368	0	3	4
		GRADRY	12	54	0	17	0
		PIGF	5	3	0	8	304
		PIGB	12	0	0	14	221
Waitomo		PLVSTCK	202	10	19	1352	1
		LIF	7	0	0	9	0
		DAIRY	22	501	1	63	0
		GRADRY	30	119	0	187	0
Wellington	Carterton	PLVSTCK	83	5	16	1012	0
		LIF	5	0	0	6	0
		DAIRY	10	475	0	12	158
		GRADRY	3	7	0	23	0
		PIGB	4	0	0	14	43
	Hutt city	PLVSTCK	11	0	13	261	1
		LIF	1	0	0	11	0
		PIGF	1	0	0	43	33
	Kapiti Coast	PLVSTCK	29	0	22	103	0
		LIF	4	0	0	4	0
		DAIRY	19	398	0	3	0
		GRADRY	2	0	0	0	0

		PIGB	9	0	0	16	62
Masterton		PLVSTCK	108	2	7	1672	0
		LIF	5	0	0	7	0
		DAIRY	6	464	0	303	0
		GRADRY	9	5	0	120	0
		PIGF	13	0	0	0	7
		PIGB	0	0	0	20	28
Porirua		PLVSTCK	28	0	7	418	0
		LIF	3	0	0	11	0
		GRADRY	0	0	0	2	0
South Wairarapa		PLVSTCK	158	7	5	1621	0
		LIF	7	0	0	5	0
		DAIRY	7	496	0	23	63
		GRADRY	19	54	0	39	0
		PIGF	34	0	0	11	7
		PIGB	0	0	0	0	40
Upper Hutt		PLVSTCK	25	0	0	100	0
		LIF	4	0	0	6	0
		DAIRY	65	141	0	1	0
		GRADRY	5	0	0	32	7
		PIGF	2	0	8	12	13
Wellington		PLVSTCK	49	0	4	519	1
		LIF	3	0	0	8	0
		GRADRY	0	0	0	4	0
West Coast	Buller	PLVSTCK	50	30	145	122	0
		LIF	3	0	0	5	0
		DAIRY	15	387	1	5	0
		GRADRY	12	28	0	4	0
		PIGF	0	0	0	0	43
	Grey	PLVSTCK	53	9	93	186	0
		LIF	3	0	0	8	0
		DAIRY	40	571	26	29	0
		GRADRY	8	38	0	7	0
		PIGB	5	30	0	6	54
	Westland	PLVSTCK	76	4	31	76	0
		LIF	3	0	0	6	0
		DAIRY	13	419	0	9	0
		GRADRY	7	28	0	2	0
		PIGB	24	0	0	0	45

Table 14.21: New Zealand Standard Model (NZSM): foot-and-mouth disease (FMD)

simulation parameters

	Value
[MovementType1]	
MovementName	DairyTanker
MovementTypeSourceData	probabilities
TimePeriodStart	1.00
SourceAnimalTypes	dairy
SourceFarmStates	!in_carea & !detected
NumberPerTimePeriod	Constant 0
NumberOfDirectContacts	Constant 0
DestinationType	farm
RestrictOnFarmClass	Y
RestrictOnAnimals	Y
DestinationMustBePopulated	Y
MovementDistance	1,0,1;0,0,40000
RepeatThisMovement	N
POTOffsetRelativeTo	clinical_signs
ProbabilityOfTransmission	Table 1,1,2,3;-4,0.0013,0.0013,0.0026
NumberPerTimePeriod[DAIRY]	Constant 0
DestinationProbability[DAIRY]	1,1,0,0,0,0,0;0,DAIRY,PLVSTCK,GRADRY,PIGB,PIGF,LIF
NumberPerTimePeriod[GRADRY LIF PIGB PIGF PLVSTCK]	Constant 0
DestinationProbability[GRADRY LIF PIGB PIGF PLVSTCK]	1,1,0,0,0,0,0;0,DAIRY,PLVSTCK,GRADRY,PIGB,PIGF,LIF
[MovementType2]	
MovementName	PLVSTCK_HRtoFarm
MovementTypeSourceData	probabilities
TimePeriodStart	1.00
SourceFarmStates	!detected
NumberOfDirectContacts	Constant 1
DestinationType	farm
RestrictOnFarmClass	Y
DestinationMustBePopulated	Y
MovementDistance	1,0,0.7059,0.1806,0.0336,0.042,0.0127,0.021,0.0042;0,0,20000,40000,60000,80000,100000,200000,1000000
POTOffsetRelativeTo	infection
ProbabilityOfTransmission	Table 1,1,11,16;0,0.525,0.8,1
DelayToInfection	Constant 0

DestinationProbability[PLVSTCK]	1,0.742,0.233,0.005,0.01,0.005,0.005;0,PLVSTCK,DAIRY,PIGB	,GRADRY,PIGF,LIF
NumberPerTimePeriod[PLVSTCK]		Poisson 0.03
NumberPerTimePeriod[DAIRY GRADRY PIGB PIGF LIF]		Constant 0
DestinationProbability[DAIRY GRADRY PIGB PIGF LIF]	1,1,0,0,0,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
[MovementType3]		
MovementName		LIF_HRtoFarm
MovementTypeSourceData		probabilities
TimePeriodStart		1.00
SourceFarmStates		!detected
NumberOfDirectContacts		Constant 1
DestinationType		farm
RestrictOnFarmClass		Y
DestinationMustBePopulated		Y
MovementDistance	1,0,0.7059,0.1806,0.0336,0.042,0.0127,0.021,0.0042;0,0,200	00,40000,60000,80000,100000,200000,1000000
POTOffsetRelativeTo		infection
ProbabilityOfTransmission		Table 1,1,6,16;0,0.525,0.9,1
DelayToInfection		Constant 0
NumberPerTimePeriod[DAIRY GRADRY PIGB PIGF PLVSTCK]		Constant 0
DestinationProbability[DAIRY GRADRY PIGB PIGF PLVSTCK]	1,1,0,0,0,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
NumberPerTimePeriod[LIF]		Poisson 0.003
DestinationProbability[LIF]	1,0.0167,0.0166,0.0167,0,0,0.95;0,PLVSTCK,DAIRY,GRADRY,PIGB,PIGF,LIF	
[MovementType4]		
MovementName		DAIRY_HRtoFarm
MovementTypeSourceData		probabilities
TimePeriodStart		1.00
SourceFarmStates		!detected
NumberOfDirectContacts		Constant 1
DestinationType		farm
RestrictOnFarmClass		Y
MovementDistance	1,0,0.7059,0.1806,0.0336,0.042,0.0127,0.021,0.0042;0,0,200	00,40000,60000,80000,100000,200000,1000000

RepeatThisMovement		N
POTOffsetRelativeTo		infection
ProbabilityOfTransmission	Table 1,1,11,16;0,0.62,0.8,1	
DelayToInfection		Constant 0
NumberPerTimePeriod[DAIRY]		Poisson 0.042
DestinationProbability[DAIRY]	1,0.207,0.57,0,0.2,0,0.023;0,PLVSTCK,DAIRY,PIGB,GRADRY,PI	GF,LIF
NumberPerTimePeriod[GRADRY LIF PIGB PIGF PLVSTCK]		Constant 0
DestinationProbability[GRADRY LIF PIGB PIGF PLVSTCK]	1,0,1,0,0,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
[MovementType5]		
MovementName		GRADRY_HRToFarm
MovementTypeSourceData		probabilities
TimePeriodStart		1.00
SourceFarmStates		!detected
NumberOfDirectContacts		Constant 1
DestinationType		farm
RestrictOnFarmClass		Y
MovementDistance	1,0,0.7059,0.1806,0.0336,0.042,0.0127,0.021,0.0042;0,0,20000,40000,60000,80000,100000,200000,1000000	
RepeatThisMovement		N
POTOffsetRelativeTo		infection
ProbabilityOfTransmission	Table 1,1,11,16;0,0.673,0.8,1	
DelayToInfection		Constant 0
NumberPerTimePeriod[GRADRY]		Poisson 0.1125
DestinationProbability[GRADRY]	1,0.081,0.82,0,0.09,0,0.009;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
NumberPerTimePeriod[DAIRY LIF PIGB PIGF PLVSTCK]		Constant 0
DestinationProbability[DAIRY LIF PIGB PIGF PLVSTCK]	1,0,0,0,1,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
[MovementType6]		
MovementName		PIGB_HRToFarm
MovementTypeSourceData		probabilities
TimePeriodStart		1.00
SourceFarmStates		!detected
NumberOfDirectContacts		Constant 1
DestinationType		farm

RestrictOnFarmClass		Y
MovementDistance	1,0,0.7059,0.1806,0.0336,0.042,0.0127,0.021,0.0042;0,0,20000,40000,60000,80000,100000,200000,1000000	
RepeatThisMovement		N
POTOffsetRelativeTo		infection
ProbabilityOfTransmission	Table 1,1,11,16;0,0.458,0.8,1	
DelayToInfection		Constant 0
DestinationProbability[PIGB]	1,0.0044,0.0123,0.1494,0.0113,0.8156,0.007;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
NumberPerTimePeriod[PIGB]		Poisson 0.1096
NumberPerTimePeriod[DAIRY GRADRY PIGF PLVSTCK]		Constant 0
DestinationProbability[DAIRY GRADRY PIGF PLVSTCK]	1,0,0,1,0,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
NumberPerTimePeriod[LIF]		Poisson 0.008
DestinationProbability[LIF]	1,0.0167,0.0166,0.0167,0,0,0.95;0,PLVSTCK,DAIRY,GRADRY,PIGB,PIGF,LIF	
;Comment:DestinationProbability[PIGB]	Incorporating figures from Pig Movement Study 2008	
;Comment:MovementName	Incorporated data from Pig Movement Study 2008	
[MovementType7]		
MovementName		FromSaleyard
MovementTypeSourceData		probabilities
TimePeriodStart		1.00
NumberPerTimePeriod		Constant 0
NumberOfDirectContacts		Constant 0
DestinationType		farm
RestrictOnFarmClass		Y
DestinationMustBePopulated		Y
MovementDistance	1,0,0.15,0.19,0.17,0.2,0.16,0.08,0.04,0.01;0,0,10000,20000,30000,50000,100000,200000,400000,1241000	
RepeatThisMovement		N
POTOffsetRelativeTo		infection
ProbabilityOfTransmission	Table 1,1,11,16;0,0.458,0.776,1	
DelayToInfection		Constant 0
DestinationProbability[PLVSTCK]	1,0.6723,0.233,0.005,0.01,0.005,0.0747;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	
DestinationProbability[DAIRY]	1,0.207,0.57,0,0.2,0,0.023;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF	

DestinationProbability[GRADRY]	1,0.09,0.82,0,0.09,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
DestinationProbability[PIGB]	1,0.0924,0.0924,0.076,0.0924,0.5544,0.0924;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
DestinationProbability[PIGF]	1,0,0,0,0,1,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
DestinationProbability[LIF]	1,0.0166,0.0167,0.95,0,0,0.0167;0,DAIRY,GRADRY,LIF,PIGB,PIGF,PLVSTCK
[MovementType8]	
MovementName	PLVSTCK_ToSaleyard
MovementTypeSourceData	probabilities
TimePeriodStart	1.00
SourceFarmStates	!in_carea & !detected
DestinationType	contact_location
ContactLocationControlName	saleyards
MovementDistance	1,0,0.95,0.03,0.02;0,0,80000,120000,904000
SourceOfSecondaryContacts	source
NumberOfSecondaryContacts[FromSaleyard]	Poisson 1.942
NumberPerTimePeriod[PLVSTCK]	Poisson 0.0135
NumberPerTimePeriod[DAIRY]	Poisson 0.005
NumberPerTimePeriod[GRADRY]	Poisson 0.003
NumberPerTimePeriod[PIGB PIGF]	Constant 0
NumberPerTimePeriod[LIF]	Poisson 0.0014
[MovementType9]	
MovementName	PIGB_ToSaleyard
MovementTypeSourceData	probabilities
TimePeriodStart	1.00
SourceFarmStates	!in_carea & !detected
DestinationType	contact_location
ContactLocationControlName	saleyards
MovementDistance	1,0,0.6904,0.1412,0.0712,0.0609,0.0143,0.022;0,0,50000,10000,150000,200000,250000,400000
SourceOfSecondaryContacts	source
NumberPerTimePeriod[PIGB]	Poisson 0.0089
NumberOfSecondaryContacts[FromSaleyard]	Poisson 2.6
NumberPerTimePeriod[DAIRY GRADRY PIGF PLVSTCK]	Constant 0
NumberPerTimePeriod[LIF]	Poisson 0.0018

;Comment:NumberPerTimePeriod[PIGB]
;Comment:MovementDistance

Incorporated data from Pig Movement Study 2008
 Added in distance distribution from PGGW saleyard dataset

[MovementType10]

MovementName MRTToFarm
MovementTypeSourceData probabilities
TimePeriodStart 1.00
SourceFarmStates !detected
NumberOfDirectContacts Constant 1
DestinationType farm
RestrictOnFarmClass Y
DestinationMustBePopulated Y

MovementDistance 1,0,0.8184,0.1205,0.0183,0.0231,0.0034,0.0093,0.007;0,0,20000,40000,60000,80000,100000,200000,1000000
RepeatThisMovement N
POTOffsetRelativeTo clinical_signs
ProbabilityOfTransmission Table 1,1,6,11,16;-1,0.1,0.2,0.4,0.5
DelayToInfection Constant 0
NumberPerTimePeriod[PLVSTCK] Poisson 0.4743

DestinationProbability[PLVSTCK] 1,0.6723,0.233,0.005,0.01,0.005,0.0747;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF

DestinationProbability[DAIRY] 1,0.207,0.57,0,0.2,0,0.023;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
NumberPerTimePeriod[DAIRY] Poisson 0.8785
NumberPerTimePeriod[GRADRY] Poisson 0.909

DestinationProbability[GRADRY] 1,0.09,0.82,0,0.09,0,0;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
NumberPerTimePeriod[PIGB] Poisson 0.326
NumberPerTimePeriod[PIGF] Poisson 0.293

DestinationProbability[PIGB | PIGF] 1,0.0125,0.0125,0.181,0.0125,0.769,0.0125;0,PLVSTCK,DAIRY,PIGB,GRADRY,PIGF,LIF
NumberPerTimePeriod[LIF] Poisson 0.1429

DestinationProbability[LIF] 1,0,0.025,0.95,0,0,0.025;0,DAIRY,GRADRY,LIF,PIGB,PIGF,PLVSTCK

[MovementType11]

MovementName LRToFarm
MovementTypeSourceData probabilities
TimePeriodStart

		1.00
SourceFarmStates		!detected
NumberPerTimePeriod		Poisson 0.264
NumberOfDirectContacts		Constant 1
DestinationType		farm
RestrictOnFarmClass		N
	1,0,0.9079,0.053,0.023,0.0115,0,0.0023,0.0023;0,0,20000,40	
MovementDistance		000,60000,80000,100000,200000,1000000
RepeatThisMovement		N
POTOffsetRelativeTo		clinical_signs
ProbabilityOfTransmission	Table 1,1,6,11,16;0,0.02,0.04,0.09,0.1	
DelayToInfection		Constant 1
[FixedRoute1]		
FixedRouteName		DairyTankerFR
TimePeriodStart		1.00
FarmClasses		DAIRY
FarmStates		lin_carea & !detected
StateChangeAction		destroy_route
NumberOfFarmsOnRoute		Poisson 5.39
SkipFarmProb		0.08
MaximumLength		40,000.00
TimeToNextTraversal		1 1
AssociatedMovementType		DairyTanker
[LocalSpread1]		
TimePeriodStart		1.00
SourceFarmStates		!depopulated & !detected
POTOffsetRelativeTo		clinical_signs
	5,1000,2000,3000,4000;-	
ProbabilityOfTransmission	1,0.,0.,0.,0.;0,0.007,0.002,0.,0.;1,0.012,0.003,0.001,0.;2,0.01	
	2,0.004,0.001,0.;3,0.009,0.004,0.001,0.	
RelativeSusceptibility[beef]		1.00
RelativeSusceptibility[dairy]		1.00
RelativeSusceptibility[deer]		0.90
RelativeSusceptibility[pigs]		0.80
RelativeSusceptibility[sheep]		

	0.90
RelativeSusceptibility[goats]	0.90
[LocalSpread2]	
TimePeriodStart	1.00
SourceFarmStates	!depopulated & detected
POTOffsetRelativeTo	clinical_signs 5,1000,2000,3000,4000;-
ProbabilityOfTransmission	1,0.,0.,0.,0.;0,0.0035,0.001,0.,0.;1,0.006,0.0015,0.0005,0.;2, 0.006,0.002,0.0005,0.;3,0.0045,0.002,0.0005,0.
RelativeSusceptibility[beef]	1.00
RelativeSusceptibility[dairy]	1.00
RelativeSusceptibility[deer]	0.90
RelativeSusceptibility[pigs]	0.80
RelativeSusceptibility[sheep]	0.90
RelativeSusceptibility[goats]	0.90
[Infectivity1]	
TimePeriodStart	1.00 Lookup
TimeToClinicalSigns	1,0,0.035087719,0.157894737,0.333333333,0.771929825,0. 789473684,0.824561404,0.877192982,0.912280702,0.9473 68421,0.964912281,0.98245614,1;0,1,2,3,4,5,6,7,8,9,11,12,1 6,17
InfectivityRelativeTo	infection
WithinFarmSpreadProb	1.00 Table
Infectivity[][][]	1,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33;0,1 ,0.941,0.882,0.823,0.764,0.705,0.646,0.587,0.528,0.469,0.41 ,0.351,0.292,0.233,0.174,0.115,0.056,0
Infectivity[][vaccimmune][]	Constant 0.5
[Zone1]	
ZoneName	NZ
ZoneType	explicit
UseFarmCentroid	Y
IncludeEntireFarm	Y
TimePeriodStartReference	simulation_start

TimePeriodStopReference	simulation_start
ChangeCount	1.00
TimePeriodStart1	1.00
TimePeriodStop1	1,000.00
[Zone2]	
ZoneName	NZ_standstill
ZoneType	explicit
FarmState	initial_standstill
UseFarmCentroid	Y
IncludeEntireFarm	Y
TimePeriodStartReference	first_detection
TimePeriodStopReference	first_detection
ChangeCount	1.00
TimePeriodStart1	1.00
TimePeriodStop1	14.00
[Zone3]	
ZoneName	Control_Area
ZoneType	radial
FarmState	in_carea
IncludeEntireFarm	N
ChangeCount	1.00
TimePeriodStart1	1.00
TimePeriodStop1	1,000.00
InsideRadius1	-
OutsideRadius1	50,000.00
Duration1	1,000.00
[Zone4]	
ZoneName	ProtectionZone
ZoneType	radial
FarmState	pz
IncludeEntireFarm	N

ChangeCount	1.00
TimePeriodStart1	1.00
TimePeriodStop1	1,000.00
InsideRadius1	-
OutsideRadius1	3,000.00
Duration1	17.00
[Zone5]	
ZoneName	SurveillanceZone
ZoneType	radial
FarmState	in_survzone
IncludeEntireFarm	N
ChangeCount	1.00
TimePeriodStart1	1.00
TimePeriodStop1	1,000.00
InsideRadius1	-
OutsideRadius1	10,000.00
Duration1	90.00
[Resource1]	
ResourceName	ResourceDepop_initial
ActionOption	depopulation
FarmListOption	single_list
FarmProcessingOption	farms_per_time_period
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	4.00
TimePeriodStart1	1.00
TimePeriodStop1	1.00
FarmsPerTimePeriod1	1.00
TimePeriodStart2	2.00
TimePeriodStop2	2.00

FarmsPerTimePeriod2	2.00
TimePeriodStart3	3.00
TimePeriodStop3	3.00
FarmsPerTimePeriod3	10.00
TimePeriodStart4	4.00
TimePeriodStop4	4.00
FarmsPerTimePeriod4	20.00
[Resource2]	
ResourceName	depop_team_plvstck
ActionOption	depopulation
FarmListOption	multiple_list
FarmProcessingOption	time_period_by_farms
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00
TimePeriodStart1	5.00
TimePeriodStop1	1,000.00
TimePeriodByFarms1	Constant Triangular 0 1 5
[Resource3]	
ResourceName	depop_team_dairy
ActionOption	depopulation
FarmListOption	multiple_list
FarmProcessingOption	time_period_by_farms
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00
TimePeriodStart1	5.00
TimePeriodStop1	1,000.00
TimePeriodByFarms1	Constant Triangular 1 1 3
[Resource4]	
ResourceName	depop_team_gradry
ActionOption	depopulation
FarmListOption	multiple_list

FarmProcessingOption	time_period_by_farms
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00
TimePeriodStart1	5.00
TimePeriodStop1	1,000.00
TimePeriodByFarms1	Constant Triangular 0 1 4
[Resource5]	
ResourceName	depop_team_pigs
ActionOption	depopulation
FarmListOption	multiple_list
FarmProcessingOption	time_period_by_farms
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00
TimePeriodStart1	5.00
TimePeriodStop1	1,000.00
TimePeriodByFarms1	Constant Triangular 0 1 3
[Resource6]	
ResourceName	depop_team_lif
ActionOption	depopulation
FarmListOption	multiple_list
FarmProcessingOption	time_period_by_farms
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00
TimePeriodStart1	5.00
TimePeriodStop1	1,000.00
TimePeriodByFarms1	Constant Triangular 0 0 1
[Resource7]	
ResourceName	ResourceVaccination
ActionOption	vaccination
FarmListOption	single_list
FarmProcessingOption	farms_per_time_period
ChangeCountTimePeriodStartReference	first_detection
ChangeCount	1.00

TimePeriodStart1	1.00
TimePeriodStop1	1,000.00
FarmsPerTimePeriod1	200.00
[Depopulation1]	
ControlName	depop_initial
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	4.00
FarmSelectionOption	detected_farm
ActionResource	ResourceDepop_initial
SharedResourcePriority	1.00
WaitingFarmState	waiting
ProcessingFarmState	processing
CompletedFarmState	depopulated
[Depopulation2]	
ControlName	IP_depop_plvstck
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	5.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	detected_farm
FarmClasses	PLVSTCK
ActionResource	depop_team_plvstck
WaitingFarmState	waiting
ProcessingFarmState	processing
CompletedFarmState	depopulated
RemoveDetectedFarms	N
[Depopulation3]	
ControlName	IP_depop_dairy
ActivationOption	detected_farm
TimePeriodStartReference	first_detection

TimePeriodStart	5.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	detected_farm
FarmClasses	DAIRY
ActionResource	depop_team_dairy
WaitingFarmState	waiting
ProcessingFarmState	processing
CompletedFarmState	depopulated
RemoveDetectedFarms	N
[Depopulation4]	
ControlName	IP_depop_gradry
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	5.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	detected_farm
FarmClasses	GRADRY
ActionResource	depop_team_gradry
WaitingFarmState	waiting
ProcessingFarmState	processing
CompletedFarmState	depopulated
RemoveDetectedFarms	N
[Depopulation5]	
ControlName	IP_depop_pigs
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	5.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	detected_farm
FarmClasses	PIGB PIGF
ActionResource	depop_team_pigs
WaitingFarmState	waiting
ProcessingFarmState	processing

CompletedFarmState	depopulated
RemoveDetectedFarms	N
[Depopulation6]	
ControlName	IP_depop_lif
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	5.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	detected_farm
FarmClasses	LIF
ActionResource	depop_team_lif
WaitingFarmState	waiting
ProcessingFarmState	processing
CompletedFarmState	depopulated
RemoveDetectedFarms	N
[Vaccination1]	
ControlName	VaccinationEarly1
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	1,000.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	zone
SelectionZone	Vaccn_zone
SelectionZoneFarmSortOrder	outer_to_inner
SelectionProb	1.00
FarmClasses	DAIRY GRADRY PLVSTCK
FarmStates	!infected & !vaccinated
AnimalTypes	beef dairy
ActionResource	ResourceVaccination
SharedResourcePriority	1.00
CompletedFarmState	vaccinated
DelayedFarmState	vaccimmune
TimePeriodToDelayedState	4.00

SurveillanceControl	HR_contact_visit
RemoveDetectedFarms	Y
ImmunityFunction	Table 1,1,2,3,4,5,6;0,1,1,1,0.5,0.25,0
[Vaccination2]	
ControlName	VaccinationEarly2
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	1,000.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
FarmSelectionOption	zone
SelectionZone	Vaccn_zone
SelectionZoneFarmSortOrder	outer_to_inner
SelectionProb	1.00
FarmClasses	DAIRY GRADRY PLVSTCK
FarmStates	infected & !detected & !vaccinated
AnimalTypes	beef dairy
ActionResource	ResourceVaccination
SharedResourcePriority	1.00
CompletedFarmState	vaccinated
SurveillanceControl	HR_contact_visit
RemoveDetectedFarms	Y
ImmunityFunction	Constant 1
[Surveillance1]	
ControlName	HR_contact_visit
ActivationOption	tracing
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
SurveillanceFarmState	hr_surv
SelectionProbability	1.00
VisitDelay	Lookup 1,0.9,1;0,0,1
VisitFrequency	Constant 1
VisitDuration	Constant 17

DelayToDetection	Lookup 1,0.9,1;0,0,1
DetectionRelativeTo	Clinical_signs
DetectionProbability[][][beef]	Constant 1
DetectionProbability[][][dairy]	Constant 1
DetectionProbability[][][deer]	Constant 1
DetectionProbability[][][goats]	Constant 1
DetectionProbability[][][pigs]	Constant 1
DetectionProbability[][][sheep]	Constant 1

[Surveillance2]

ControlName	MR_contact_visit
ActivationOption	tracing
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
SurveillanceFarmState	mr_surv

SelectionProbability	0.90
VisitDelay	Triangular 0 1 2
VisitFrequency	Constant 2
VisitDuration	Constant 17
DelayToDetection	Triangular 0 1 2
DetectionRelativeTo	Clinical_signs
DetectionProbability[][][beef]	Constant 1
DetectionProbability[][][dairy]	Constant 1
DetectionProbability[][][deer]	Constant 1
DetectionProbability[][][goats]	Logistic 0.25 0.8 0.74 1.7
DetectionProbability[][][pigs]	Constant 1
DetectionProbability[][][sheep]	Logistic 0.25 0.8 0.74 1.7

[Surveillance3]

ControlName	LR_contact_visit
ActivationOption	tracing
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
SurveillanceFarmState	lr_surv
SelectionProbability	

	0.50
VisitDelay	Triangular 1 2 3
VisitFrequency	Constant 3
VisitDuration	Constant 17
DelayToDetection	Triangular 0 1 2
DetectionRelativeTo	Clinical_signs
DetectionProbability[][][beef]	Constant 1
DetectionProbability[][][dairy]	Constant 1
DetectionProbability[][][deer]	Constant 1
DetectionProbability[][][goats]	Logistic 0.25 0.8 0.74 1.7
DetectionProbability[][][pigs]	Constant 1
DetectionProbability[][][sheep]	Logistic 0.25 0.8 0.74 1.7
[Surveillance4]	
ControlName	patrol_visit
ActivationOption	detected_farm
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
SelectionZone	ProtectionZone
SelectionProbability	1.00
VisitDelay	Constant 0
VisitFrequency	Constant 2
VisitDuration	Constant 14
DelayToDetection	Triangular 0 1 2
DetectionRelativeTo	Clinical_signs
DetectionProbability[][][beef]	Constant 1
DetectionProbability[][][dairy]	Constant 1
DetectionProbability[][][deer]	Constant 1
DetectionProbability[][][goats]	Logistic 0.25 0.8 0.74 1.7
DetectionProbability[][][pigs]	Constant 1
DetectionProbability[][][sheep]	Logistic 0.25 0.8 0.74 1.7
[Surveillance5]	
ControlName	post_detection_self_reporting
ActivationOption	time_period
TimePeriodStartReference	first_detection
TimePeriodStart	1.00

TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
SelectionZone	NZ
SelectionProbability	1.00
VisitDelay	Constant 0
VisitFrequency	Constant 1
VisitDuration	Constant 1000
DelayToDetection	Constant 0
DetectionRelativeTo	Clinical_signs Table
	1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,19,20,23;0,0.072, 0.2016,0.2399,0.1737,0.1124,0.072,0.0454,0.0275,0.0193,0. 0128,0.009,0.0053,0.0041,0.0027,0.0011,0.0006,0.0002,0.00 02,0.0001,0.0001
DetectionProbability[DAIRY PIGB PIGF][[]]	
DetectionProbability[GRADRY PLVSTCK LIF][[]]	##### #####
[Surveillance6]	
ControlName	background_surveillance
ActivationOption	time_period
TimePeriodStartReference	simulation_start
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1.00
SelectionZone	NZ
SelectionProbability	1.00
VisitDelay	Constant 0
VisitFrequency	Constant 1
VisitDuration	Constant 1000
DelayToDetection	Constant 0
DetectionRelativeTo	Clinical_signs Table
	1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,22;0,0.058 4,0.1597,0.2076,0.1701,0.1325,0.0904,0.0617,0.035,0.0251, 0.0189,0.0132,0.0104,0.0063,0.0043,0.0031,0.0016,0.0008,0 .0007,0.0001,0.0001
DetectionProbability[DAIRY PIGB PIGF][[]]	
DetectionProbability[GRADRY PLVSTCK LIF][[]]	##### #####
[Tracing1]	
ControlName	tracing_activity

TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
TracingRequired[DAIRY_HRtoFarm][off]	Y
TracingRequired[GRADRY_HRtoFarm][off]	Y
TracingRequired[PIGB_HRtoFarm][off]	Y
TracingRequired[PLVSTCK_HRtoFarm][off]	Y
TracingRequired[LIF_HRtoFarm][off]	Y
TracingRequired[DAIRY_HRtoFarm][on]	Y
TracingRequired[GRADRY_HRtoFarm][on]	Y
TracingRequired[PIGB_HRtoFarm][on]	Y
TracingRequired[PLVSTCK_HRtoFarm][on]	Y
TracingRequired[LIF_HRtoFarm][on]	Y
TracingRequired[MRTtoFarm][off]	Y
TracingRequired[MRTtoFarm][on]	Y
TracingRequired[LRTtoFarm][off]	Y
TracingRequired[LRTtoFarm][on]	N
TracingRequired[DairyTanker][off]	Y
TracingRequired[DairyTanker][on]	Y
TracingRequired[FromSaleyard][off]	Y
TracingRequired[FromSaleyard][on]	Y
ProbMovementForgotten[DAIRY_HRtoFarm][off]	0.11
ProbMovementForgotten[GRADRY_HRtoFarm][off]	0.11
ProbMovementForgotten[PIGB_HRtoFarm][off]	0.11
ProbMovementForgotten[PLVSTCK_HRtoFarm][off]	0.11
ProbMovementForgotten[LIF_HRtoFarm][off]	0.01
ProbMovementForgotten[DAIRY_HRtoFarm][on]	0.08
ProbMovementForgotten[GRADRY_HRtoFarm][on]	0.08
ProbMovementForgotten[PIGB_HRtoFarm][on]	0.08
ProbMovementForgotten[LIF_HRtoFarm][on]	0.01
ProbMovementForgotten[MRTtoFarm][off]	0.21
ProbMovementForgotten[MRTtoFarm][on]	0.19
ProbMovementForgotten[DairyTanker][off]	

	0.01
ProbMovementForgotten[DairyTanker][on]	0.01
ProbMovementForgotten[FromSaleyard][off]	0.06
ProbMovementForgotten[FromSaleyard][on]	0.06
ProbMovementForgotten[LRTtoFarm][off]	0.36
TracingDelay[DAIRY_HRtoFarm][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[GRADRY_HRtoFarm][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[PIGB_HRtoFarm][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[PLVSTCK_HRtoFarm][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[LIF_HRtoFarm][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[DAIRY_HRtoFarm][on]	Lookup 1,0.5,1;0,0,1
TracingDelay[GRADRY_HRtoFarm][on]	Lookup 1,0.5,1;0,0,1
TracingDelay[PIGB_HRtoFarm][on]	Lookup 1,0.5,1;0,0,1
TracingDelay[PLVSTCK_HRtoFarm][on]	Lookup 1,0.5,1;0,0,1
TracingDelay[LIF_HRtoFarm][on]	Lookup 1,0.5,1;0,0,1
TracingDelay[MRTtoFarm][off]	BetaPert 1 2 3
TracingDelay[MRTtoFarm][on]	BetaPert 1 2 3
TracingDelay[LRTtoFarm][off]	BetaPert 2 3 4
TracingDelay[DairyTanker][off]	Lookup 1,0.9,1;0,0,1
TracingDelay[DairyTanker][on]	Lookup 1,0.9,1;0,0,1
TracingDelay[FromSaleyard][off]	Lookup 1,0.5,1;0,0,1
TracingDelay[FromSaleyard][on]	Lookup 1,0.5,1;0,0,1
SurveillanceControls[PLVSTCK_HRtoFarm][off]	HR_contact_visit
SurveillanceControls[LIF_HRtoFarm][off]	HR_contact_visit
SurveillanceControls[DAIRY_HRtoFarm][off]	HR_contact_visit
SurveillanceControls[PIGB_HRtoFarm][off]	HR_contact_visit
SurveillanceControls[GRADRY_HRtoFarm][off]	HR_contact_visit
SurveillanceControls[FromSaleyard][off]	HR_contact_visit
SurveillanceControls[MRTtoFarm][off]	MR_contact_visit
SurveillanceControls[LRTtoFarm][off]	LR_contact_visit
SurveillanceControls[DairyTanker][off]	LR_contact_visit
SurveillanceControls[PLVSTCK_HRtoFarm][on]	HR_contact_visit
SurveillanceControls[LIF_HRtoFarm][on]	HR_contact_visit
SurveillanceControls[DAIRY_HRtoFarm][on]	HR_contact_visit
SurveillanceControls[GRADRY_HRtoFarm][on]	HR_contact_visit
SurveillanceControls[PIGB_HRtoFarm][on]	HR_contact_visit
SurveillanceControls[FromSaleyard][on]	HR_contact_visit

SurveillanceControls[MRTtoFarm][on]	MR_contact_visit
SurveillanceControls[DairyTanker][on]	LR_contact_visit
ProbMovementForgotten[PLVSTCK_HRtoFarm][on]	0.08
[MovementRestriction1]	
ControlName	HRMC_InitStandstill
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	14.00
MovementTypes	DAIRY_HRtoFarm GRADRY_HRtoFarm LIF_HRtoFarm PIGB_HRtoFarm PIGB_ToSaleyard PLVSTCK_HRtoFarm PLVSTCK_ToSaleyard
SourceFarmStates	initial_standstill
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.91
[MovementRestriction2]	
ControlName	MRMC_InitStandstill
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	14.00
MovementTypes	MRTtoFarm
SourceFarmStates	initial_standstill
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.60
[MovementRestriction3]	
ControlName	LRMC_InitStandstill
TimePeriodStartReference	first_detection
TimePeriodStart	1.00
TimePeriodStopReference	first_detection
TimePeriodStop	14.00
MovementTypes	LRTtoFarm
SourceFarmStates	initial_standstill
DestinationFarmClasses	DAIRY !DAIRY

ProbMovementRestricted	0.24
[MovementRestriction4]	
ControlName	HRMC_InfZone
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	DAIRY_HRToFarm GRADRY_HRToFarm LIF_HRtoFarm PIGB_HRToFarm PIGB_ToSaleyard PLVSTCK_HRtoFarm PLVSTCK_ToSaleyard
SourceFarmStates	in_carea & !(in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.94
[MovementRestriction5]	
ControlName	MRMC_InfZone
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	MRTtoFarm
SourceFarmStates	in_carea & !(in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.80
[MovementRestriction6]	
ControlName	LRMC_InfZone
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	LRTtoFarm
SourceFarmStates	in_carea & !(in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.39

[MovementRestriction7]

ControlName HRMC_SurvZone
TimePeriodStartReference first_detection
TimePeriodStart 14.00
TimePeriodStopReference first_detection
TimePeriodStop 1,000.00
DAIRY_HRtoFarm GRADRY_HRtoFarm LIF_HRtoFarm
PIGB_HRtoFarm PIGB_ToSaleyard PLVSTCK_HRtoFarm
PLVSTCK_ToSaleyard
MovementTypes
SourceFarmStates in_carea & (in_survzone | hr_surv | mr_surv | lr_surv)
DestinationFarmClasses DAIRY | !DAIRY
ProbMovementRestricted 0.95

[MovementRestriction8]

ControlName MRMC_SurvZone
TimePeriodStartReference first_detection
TimePeriodStart 14.00
TimePeriodStopReference first_detection
TimePeriodStop 1,000.00
MovementTypes MRtoFarm
SourceFarmStates in_carea & (in_survzone | hr_surv | mr_surv | lr_surv)
DestinationFarmClasses DAIRY | !DAIRY
ProbMovementRestricted 0.85

[MovementRestriction9]

ControlName LRMC_SurvZone
TimePeriodStartReference first_detection
TimePeriodStart 14.00
TimePeriodStopReference first_detection
TimePeriodStop 1,000.00
MovementTypes LRtoFarm
SourceFarmStates in_carea & (in_survzone | hr_surv | mr_surv | lr_surv)
DestinationFarmClasses DAIRY | !DAIRY
ProbMovementRestricted 0.52

[MovementRestriction10]

ControlName	HRMC_OutCA
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	DAIRY_HRToFarm GRADRY_HRToFarm LIF_HRtoFarm PIGB_HRToFarm PIGB_ToSaleyard PLVSTCK_HRtoFarm PLVSTCK_ToSaleyard
SourceFarmStates	!in_carea & (in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.95
[MovementRestriction11]	
ControlName	MRMC_OutCA
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	MRTtoFarm
SourceFarmStates	!in_carea & (in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.85
[MovementRestriction12]	
ControlName	LRMC_OutCA
TimePeriodStartReference	first_detection
TimePeriodStart	14.00
TimePeriodStopReference	first_detection
TimePeriodStop	1,000.00
MovementTypes	LRTtoFarm
SourceFarmStates	!in_carea & (in_survzone hr_surv mr_surv lr_surv)
DestinationFarmClasses	DAIRY !DAIRY
ProbMovementRestricted	0.52

Table 14.22: *Estimated economic parameters for estimation of the economic impacts of a foot-and-mouth disease (FMD) epidemic*

Parameters	Unit cost	Source
Premises-level cost (NZD per premises)		
Cost of depopulation	92,780	i
Cost of vaccination for vaccinate-to-die	5,421	i
Cost of vaccination for vaccinate-to-live (including post-epidemic serological testing)	21,372	i
Cost of active surveillance for high-risk premises (incl. 17 visits)	17,767	i
Cost of active surveillance for medium-risk premises (incl. 9 visits)	9,406	i
Cost of active surveillance for low-risk premises (incl. 6 visits)	6,271	i
Cost of active surveillance within 3 km protection zone (incl. 7 visits + 1 post-epidemic visit)	8,361	i
Cost of post-epidemic surveillance in 10 km surveillance zone (incl. 1 post-epidemic visit)	1,045	i
Cost of diagnostics for suspected premises by veterinarian visit	1,100	i
Cost of diagnostics for suspected by livestock owners, including misdiagnosis	6,600	i
Daily cost (NZD per day)		
Fixed daily cost for crisis centre	252,146	i
Increment in daily cost for crisis centre for larger epidemic (#IPs > 280)	1,821	i
Animal-level cost (NZD per animal)		
Depopulated beef	1,055	ii, iii
Depopulated dairy	2,134	ii, iv
Depopulated deer	461	ii, iii
Depopulated sheep	500	ii, iii
Depopulated pig	301	ii, v
Depopulated goat	209	ii, vi
Beef under movement restriction	4	iii
Dairy under movement restriction	10	iv
Deer under movement restriction	1	iii
Sheep under movement restriction	1	iii
Pig under movement restriction	2	v
Goat under movement restriction	4	vi
Vaccinated beef	998	ii, iii
Vaccinated dairy	1,939	ii, iv
Vaccinated deer	442	ii, iii
Vaccinated sheep	487	ii, iii
Vaccinated pig	273	ii, v
Vaccinated goat	158	ii, vi
Fixed cost (NZD)		
Post-epidemic fixed skeleton cost	8,680,730	i

Increment in post-epidemic skeleton cost for larger epidemic (#IPs > 280)	59,180	i
Post-epidemic surveillance in disease-free zone	4,307,911	i
Other expenses	3,200,500	i

Source i (Ansell, Unpublished results; Bingham, Unpublished results), ii (Inland Revenue, 2011, 2012, 2013, 2014), iii (Beef + Lamb New Zealand Economic Service, 2014), iv (DairyNZ, 2014), v (Askin and Askin, 2012), and vi (Solis-ramirez et al., 2012).

Table 14.23: *Estimated slaughtering capacity by area for subsequent culling of vaccinated animals for control of a foot-and-mouth disease outbreak by vaccinate-to-die*

Area	Region	Slaughtering capacity (number of animals per day per region)
Northern North Island	Northland, Auckland, Waikato, Bay of Plenty	4,305
Eastern North Island	Hawke's Bay, Gisborne	9,053
Western North Island	Taranaki, Wanganui - Manawatu, Wellington	8,997
Northern North Island	Nelson, Tasman, Marlborough, West Coast	712
Eastern South Island	Canterbury	25,875
Southern South Island	Otago, Southland	21,787