Ministry for Primary Industries
Manatū Ahu Matua

Development and application of new tools for the analysis of *Salmonella* surveillance data identifying the spatial and temporal determinants of raised notifications in New Zealand

MPI Technical Paper No: 2012/04

Prepared for the Ministry for Primary Industries
by Prof Nigel French, Dr Debbie Leader, Prof Martin Hazelton,
Dr Jonathan Marshall, Dr Jackie Benschop

ISBN No: 978-0-478-38860-2 (online)
ISSN No: 2253-3923 (online)

May 2012

New Zealand Government

Growing and Protecting New Zealand

Final Report: SCIG-MAS-001

Development and application of new tools for the analysis of *Salmonella* surveillance data identifying the spatial and temporal determinants of raised notifications in New Zealand

April 2011

prepared for Dr Donald Campbell
New Zealand Food Safety Authority

by

Prof Nigel French
Dr Debbie Leader
Prof Martin Hazelton
Dr Jonathan Marshall
Dr Jackie Benschop

[m]EpiLab, Hopkirk Research Institute
Institute of Veterinary, Animal and Biomedical Sciences
College of Sciences
Massey University
New Zealand
Email - N.P.French@massey.ac.nz
Phone - +64 (06) 356 9099 extn 81188
Fax - +64 (06) 350 5714

Contents

Section 1

Executive Summary

Salmonellosis is consistently in the top three most notified infectious disease in New Zealand. In this study, over 15,000 salmonellosis notifications (non-typhoidal) were analysed at the serotype level over a ten-year period from 2000 to 2009 across all of New Zealand. Spatial and temporal trends were identified and risk factors associated with these trends were investigated. A number of patterns and relationships were observed that improve our understanding of the epidemiology of the most prevalent serotypes, and provide insight into the relative importance of different transmission pathways - both food and environmental. Although some general patterns were evident across all salmonellosis cases, it is clear that determinants of disease at the serotype level vary considerably.

In addition to exploratory methods for analysing spatial and temporal patterns, we adapted and applied a number of statistical models to make inferences and test hypotheses. These included the following: a model for determining background spatial and temporal trends [20]; an extension of the Knorr-Held and Richardson model [19] to identify outbreaks over the 10 year period [27]; and a number of Zero-Inflated Poisson (ZIP) and Negative Binomial (ZINB) regression models to investigate possible risk factors.

When all cases were considered together, investigation into the spatial and temporal trends showed a higher level of notification of salmonellosis in the lower South Island compared to the rest of New Zealand. The majority of notified cases occurred within the first 2-3 years of the study period when ma-

jor epidemics of *Salmonella* Typhimurium DT160, Typhimurium DT135 and Brandenburg were occurring. However a diverse range of spatial and temporal patterns were observed when the ten most prevalent serotypes were analysed individually (these accounted for over 50% of all cases). For example, *S.* Brandenburg was predomininatly found in lower South Island in the first 2-3 years of the study period, whereas *S.* Typhimurium DT156 showed a higher level of notification in areas of North Island where the notification rate was consistently high over the study period. Apart from the major epidemic in North Canterbury between 2000 and 2002, the risk of *S.* Typhimurium DT160 was spread relatively evenly across both Islands, whereas others such as *S.* Typhimurium DT1 showed areas of consistently elevated risk in particular regions. *S.* Typhimurium DT42 and *S.* Saint Paul both showed higher relative risk in South compared to North Island.

Reasons for the spatial variation in risk could be determined for some of the serotypes examined; for example the strong concentration of *S.* Brandenburg in areas of South Island was associated with high densities of sheep and dairy farms, and the seasonal association with lambing and calving periods indicated a strong environmental component to infection and disease in these areas. Similarly the distribution of *S.* Saint Paul was strongly associated with remote rural areas and high sheep densities, also indicating environmental exposure pathways. For other serotypes, such as *S.* Typhimurium DT156, the overall spatial pattern appeared to be determined by repeated outbreaks in the same region (in this case the Nelson / Marlborough region), for which there was not an obvious seasonal pattern or any association with environmental factors. When considered alongside other information from earlier studies, these indicate the predominance of food pathways as a source of infection and disease for *S.* Typhimurium DT156.

The spatial pattern alone did not provide clues to the determinants of the most prevalent serotype, *S.* Typhimurium DT160, but the seasonal pattern, showing annual peaks in late spring; the evidence of multiple localised outbreaks nested within a large-scale epidemic; and the previously reported concurrent epidemic in wild birds, was consistent with a disease caused predominantly by exposure to wild bird feacal material, either through environmental pathways or the contamination of food. The association with rural areas and sheep densities in North Island also indicate involvement with

other hosts species, which is consistent with spread between wild birds and livestock, and the isolation of this serotype in faecal samples from multiple host species.

An analysis of age and gender showed some variation between serotypes. Generally, most serotypes were more prevalent in young children, particularly pre-school children, and many were significantly associated with males. Exceptions include *S.* Infantis, which showed a generally higher level of disease in adults compared to other serotypes, and *S.* Typhimurium DT135 which displayed more variation in the age distribution over the ten-year period. The strongest association with males compared to females was observed in lower South Island, and this was most evident for the serotype strongly associated with this region: *S.* Brandenburg.

The Knorr-Held and Richardson model [19] was successfully adapted to salmonellosis data and was able to identify most of the pre-identified outbreaks, with some major exceptions. The major outbreak of *S.* Typhimurium DT42 associated with contaminated flour in 2008 was detected with high probability, but the large outbreak of *S.* Typhimurium DT160 associated with an umu function [6] was not detected because most of the cases were not present in the EpiSurv database provided for analysis. This model was also able to identify a large number of cases of multiple serotypes that were not identified as being part of a notified outbreak, but were clustered in space and time and therefore likely to have been associated with a common cause. This suggests that many epidemiologically-important clusters of cases are not being recognised by the current surveillance system as outbreaks, and many cases are not being identified as part of a known outbreak.

Consideration of the top four most prevalent serotypes individually: Brandenburg and Typhimurium DT160, DT1 and DT135, which together accounted for 34% of all cases, underlined the major differences in epidemiology of salmonellosis at the serotype level, indicating very different factors are determining the patterns of disease caused by individual serotypes. Using information provided by this study and other recent reports we conclude:

- The predominance of rural cases of *Salmonella* Brandenburg, particularly during the calving and lambing season, and the relatively low

number of outbreaks associated with this serotype, suggest that food borne exposure is relatively unimportant for transmission of *Salmonella* Brandenburg to humans.

- Although environmental exposure to livestock may play a role in the transmission of *S.* Typhimurium DT160, the available evidence indicates it is not as important as it is for *S.* Brandenburg. The frequency of outbreaks, in both humans and animals, and the epidemiology of sporadic cases, is consistent with a higher level of food borne transmission than *S.* Brandenburg. Infection in wild birds, resulting in contamination of both food and the environment (and transmission to other animal species), may be the most important factor driving the epidemiology of this serotype.

- Given the information provided by this study and other recent reports, it is difficult to determine what is the predominant source of either sporadic or outbreak-related cases of *S.* Typhimurium DT1. The association with rural areas, and the frequent isolation from cattle suggests that environmental pathways resulting from exposure to cattle faeces may be important, but this is not supported by the seasonal pattern or any association with either dairy or beef cattle densities.

- The number of sporadic cases and outbreak associated cases of *S.* Typhimurium DT135 (both notified and non-notified), have declined markedly since 1999. This is therefore a serotype in the post-epidemic phase, that appears to be predominantly associated with food borne exposures.

In summary, we have identified a number of similarities and differences in the spatial and temporal epidemiology of the most common serotypes causing salmonellosis in New Zealand. This information, combined with a detailed analysis of risk factors and other recent reports suggest it is likely that environmental exposure to ruminant livestock and wild birds plays a major role in the epidemiology of several of the most common serotypes, whereas food borne exposure is the predominant pathway for others. The complex dynamics of salmonellosis serotypes are created by multiple clusters of cases in space and time, nested within long-term epidemic behaviour. This makes it difficult to identify, and subsequently determine the cause of outbreaks.

In addition to generic measures aimed at reducing contamination of food, we recommend particular attention be given to the following:

- Reducing exposure of rural children to faecal material from ruminants during the lambing and calving season.
- Reducing wild bird access to food production and retail premises, particularly during late spring and summer.
- Enhancing the surveillance of salmonellosis by the rapid identification and follow-up of clusters of cases in space and time. This may be done through the implementation of new statistical tools, combined with rapid sub-typing and the conduct of case-control studies using standard control sets.

Section 2

Introduction

Salmonella is ranked in the top three most important enteric pathogens in New Zealand, up until recently it was second only to *Campylobacter*, but recently the rates of giardiasis have exceeded those of salmonellosis [1]. In the 2010-2013 *Salmonella* risk management strategy, the New Zealand Food Safety Authority (NZFSA) set a goal of achieving a 30% reduction in the reported annual incidence of food-borne salmonellosis by 2013. In order to reduce the incidence of salmonellosis, a thorough understanding of the epidemiology of this complex disease is required, and thus there is need to understand the spatial and temporal determinants of raised notifications. Because the determinants of disease caused by different *Salmonella* serotypes are likely to vary this understanding must be examined at the serotype level.

There are several risk factors known to be associated with salmonellosis notifications, and many of these are spatially and temporally structured. Social deprivation is a risk factor for many infectious diseases [20], as is gender, urban/rural profile and age. As part of their analysis Adlam et al. [1] investigated the differences between the number of reported cases of salmonellosis in New Zealand for a range of demographic variables. Adlam et al. noted the over representation of males to females in data covering the years 2000 to 2009 and that over this time South Island regions of New Zealand had higher rates of salmonellosis than the rest of the country. They also reported that a higher proportion of people 16 years and younger were notified compared to those 17 years and older and that more than 1 in 10 salmonellosis cases

were observed in rural areas of New Zealand over this same time period [1].

The goal of this project was to analyse salmonellosis notification data spanning the years 2000 to 2009 for all of New Zealand. The aims were four fold:

1. Adapt the model developed for campylobacteriosis under SCIG-MAS-001 and apply to salmonellosis notifications between 2000-2009, to determine background spatial and temporal trends at both the species and serotype level (depending on data availability). The model was a Bayesian statistical model at the meshblock level based on the collection of models described in Diggle et al. [12].

2. Identify potential risk factors that might be associated with these trends by examining the relationship between notification rates and a number of potential explanatory variables. These included: livestock densities, deprivation index, water supply regions, and meteorological variables. We also explored the use of a statistical model to examine variables associated with particular serotypes. For the latter analysis we needed to access variables from the EpiSurv database such as age, gender and particular exposures.

3. Develop the model for identifying anomalous outbreaks described by Knorr-Held and Richardson [19] and applied to campylobacteriosis [20, 27], using appropriate epidemic indicators for salmonellosis. This was applied at different spatial resolutions including meshblock and TA.

4. Consider notification data at both the species (all notified non-typhoidal salmonellosis cases) and serotype level using the three modeling approaches stated above.

An overview of the data is given in the following section. The models used to achieve the above aims are then described in Section 4 and an investigation of risk factors that might be associated with the spatial and temporal trends is given in Section 5.

Section 3

Data overview

The data consist principally of a list of cases with a notification date and an approximate spatial location. The data span the years 2000-2009 for all of New Zealand. The spatial information associated with each notification is the census meshblock, which are small areas of New Zealand that normally contain between 0 and 200 people in their usually resident address. Meshblocks therefore vary in size, with those in urban areas giving a more precise spatial location than those in rural areas. In addition, we were provided with the serotype of the *Salmonella* as well as the age and gender of the case. The 10 most prevalent serotypes were investigated. The most prevalent of these was *S.* Typhimurium DT160 with a total of 2,592 reported cases, representing approximately 17% of all notified cases over the 10 year period. Table 3.1 presents the 10 most prevalent serotypes over the 10 year period along with the total number of notified cases. These 10 serotypes accounted for approximately 55% of all reported cases over the ten year period investigated. Table A.1 gives the total number of reported cases for the 30 most prevalent serotypes along with the number of these that were identified as being part of a known outbreak.

Variables considered as potential risk factors include the Social Deprivation Index (SDI), age and gender, water zone, weather information, rurality and ruminant densities. The SDI was provided at the meshblock level from the 2006 census data. Information pertaining to the water zone for each meshblock was provided by the Institute of Environmental Science and Research

Table 3.1: Total number of cases from January 2000 to December 2009 for the ten most prevalent serotypes

Serotype	Number of Cases
S. Typhimurium DT160	2592
S. Typhimurium DT1	1010
S. Typhimurium DT135	844
S. Brandenburg	734
S. Typhimurium DT156	705
S. Infantis	657
S. Typhimurium DT101	570
S. Enteritidis phage type 9a	544
S. Typhimurium DT42	334
S. Saint Paul	310

Ltd (ESR) from the Water Information New Zealand (WINZ) database. Weather data that had been recorded at virtual weather stations located throughout New Zealand were obtained from the NIWA National Climate Database and included average rainfall, temperature and absolute humidity. However only data for the years spanning 2000 to 2007 were available. Figures A.1 to A.3 illustrate the average rainfall, absolute humidity and temperature for each District Health Board (DHB) over this period. For each meshblock eight urban/rural profiles were provided as defined by the 2006 census. Three of the profiles were urban (independent urban areas, main urban areas and satellite urban areas) and four were rural (highly rural/remote areas, rural areas with high urban influence, rural areas with moderate urban influence and rural areas with low urban influence). The remaining profile (outside urban/rural profile) was unable to be categorised and for the purposes of this study was not considered. This was considered reasonable since only 6% of all reported cases of salmonellosis over the ten year period were identified as being outside the urban/rural profile. To examine the influence of rurality on disease incidence, the populations in each of the seven urban/rural profiles were aggregated to a classification of either urban or rural [26].

Data on ruminant densities (per hectare) in each meshblock for poultry, sheep, dairy and beef were acquired from the Agribase^{TM}database. To examine the influence of each of the ruminant densities on disease incidence, the density (per hectare) for each species within each meshblock was classified into one of three groups (None, Medium density and High density). Once all

meshblocks with zero density were separated into the None group the median of the remaining densities was used to classify these remaining meshblocks in to Medium (less than the median) and High (greater than the median) density. For example in 2006 there were 29,727 meshblocks that did not have any sheep in them, for the remaining meshblocks the median sheep density per hectare was 0.743 which when used to classify the remaining meshblocks it Medium and High density meant that these two groups contained 4,586 and 4,587 meshblocks respectively.

All spatial and temporal data were interpolated or aggregated to the meshblock and week level to facilitate comparison with the notification data. Further aggregations included, Territorial Authority and week level and District Health Board and Month. This further aggregation was deemed necessary to reduce the computational running time.

3.1 Visualising the Human Case Data

Figure 3.1 shows the initial mapping of the total number of reported cases of *Salmonella* in each meshblock over the ten year period. All maps produced in this document are high resolution maps and we recommend the use of 'Pan and Zoom' in the Adobe reader 'Tools' for optimum viewing of these images. A higher number of reported cases of salmonellosis occurred in the upper and lower South Island, which is consistent with the results reported by Adlam et al.[1]. The number of reported cases for each of the ten most prevalent serotypes were then mapped at the meshblock level. For illustration the Figures 3.2 and 3.3 show the total number of cases of *S.* Typhimurium DT160 and Brandenburg between 2000 and 2009 at the meshblock level respectively. These two serotypes were chosen to be included here as *S.* Typhimurium DT160 was the most common serotype to be reported over the ten year period and *S.* Brandenburg was chosen to illustrate a serotype with a very different epidemiology to *S.* Typhimurium DT160.

There were 8 meshblocks that had 5 cases of *S.* Typhimurium DT160 over the 10-year period. Three of these meshblocks were in the Canterbury region. There was one meshblock that had 16 cases of *S.* Brandenburg over the 10-year period. As can be seen in the figure for *S.* Brandenburg this meshblock

is located in Southland. Comparing the distribution of these two serotypes we see there is a wider geographical distribution of *Salmonella* Typhimurium DT160 than for *S.* Brandenburg, which appears to be more concentrated in the lower South Island. Further examination of these two figures showed a much higher incidence of reported cases of *S.* Typhimurium DT160 in Auckland, Wellington and Christchurch over the ten year period compared to the incidence of reported cases of *S.* Brandenburg in these cities (see the inserts in Figures 3.2 and 3.3 respectively).

The spatial differences observed in these two plots are further examined by applying the model described in Section 4. The total number of reported cases between 2000 and 2009 for these two serotypes and the eight other most prevalent serotypes can be seen in Table 3.1.

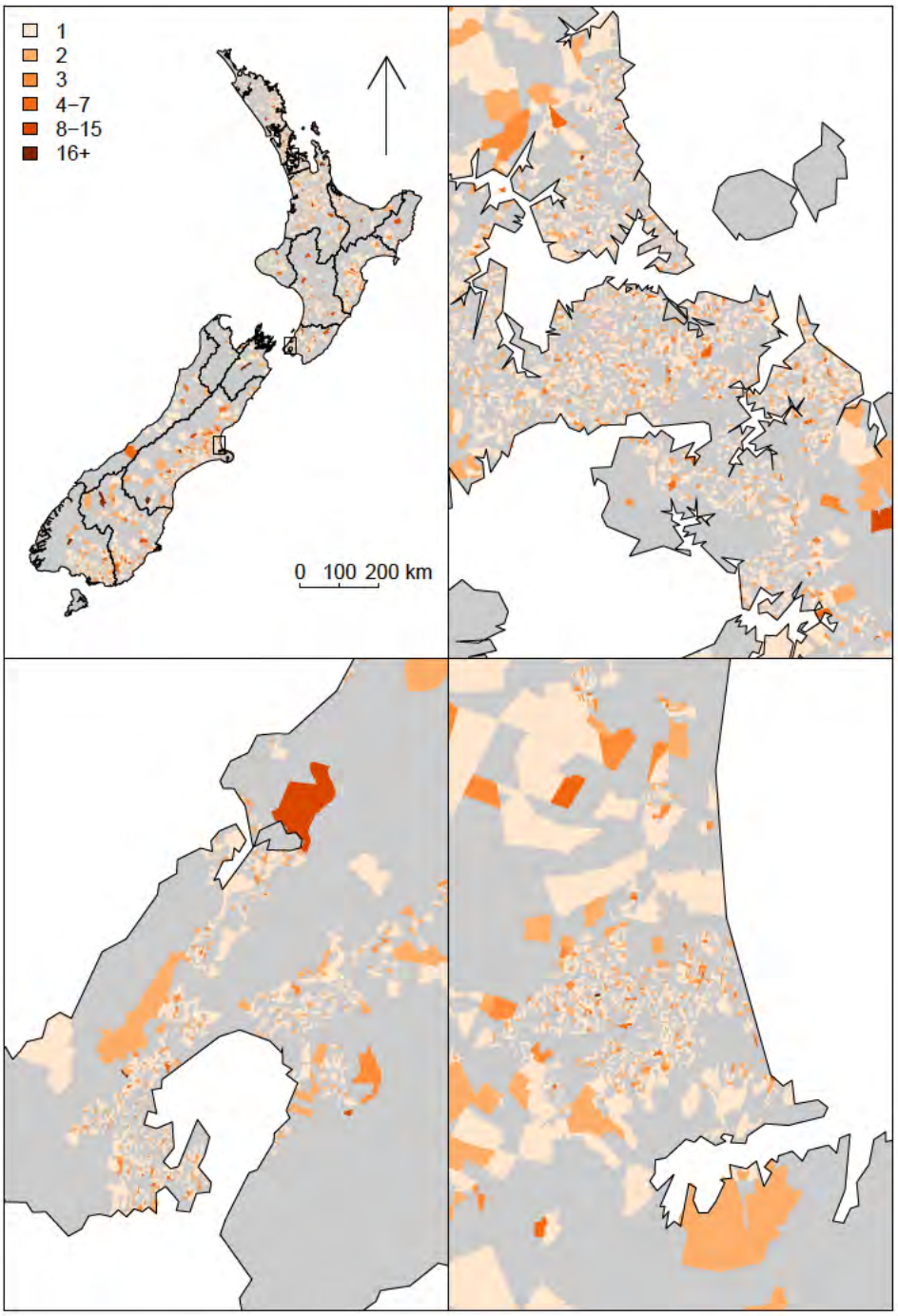

Figure 3.1: Total number of *Salmonella* cases between 2000 and 2009 at Meshblock level for all of New Zealand (top left), with inserts for Auckland (top right), Wellington (bottom left) and Christchurch (bottom right).

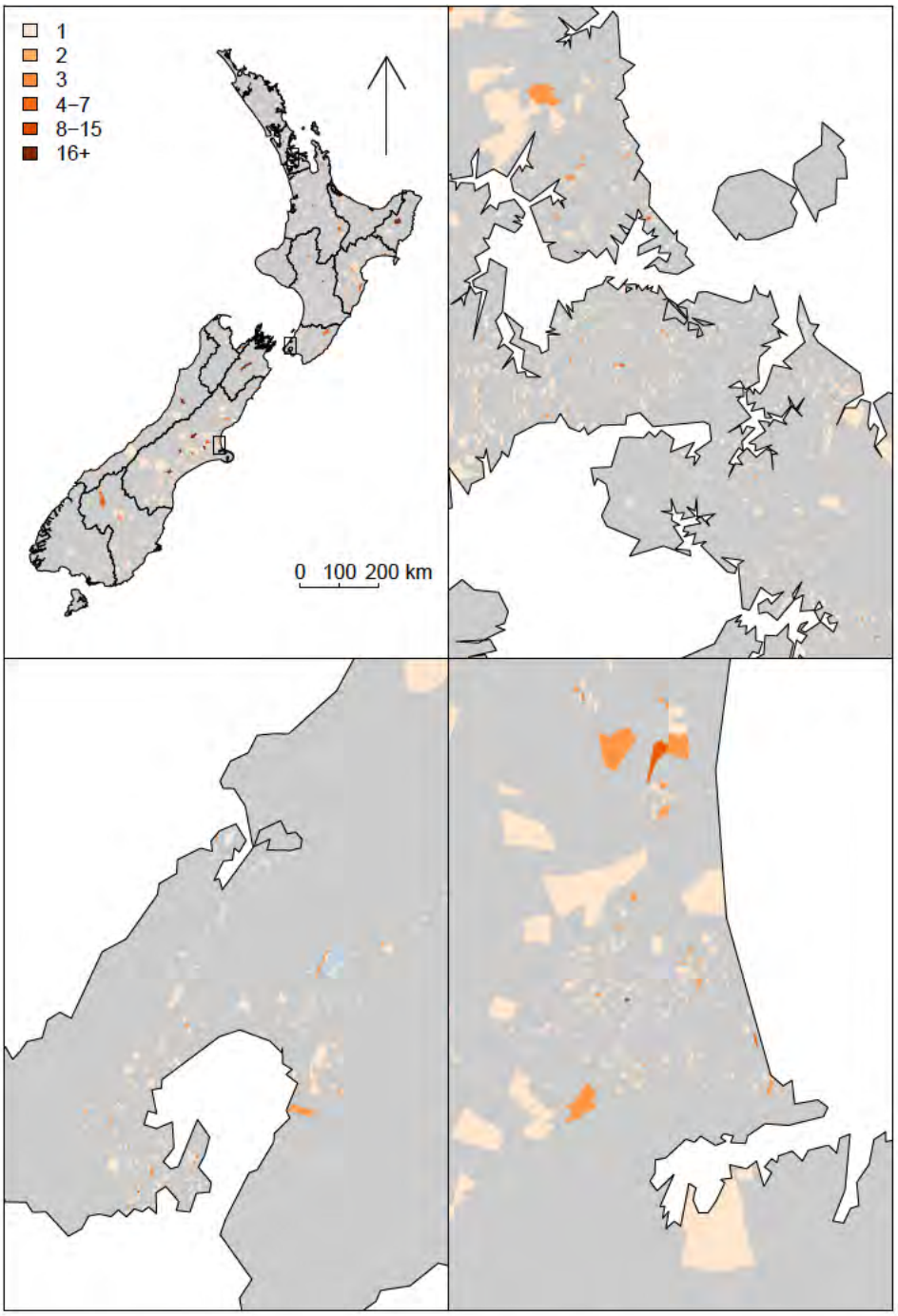

Figure 3.2: Number of *Salmonella* Typhimurium DT 160 Cases between 2000 and 2009 at Meshblock level for all of New Zealand (top left), with inserts for Auckland (top right), Wellington (bottom left) and Christchurch (bottom right).

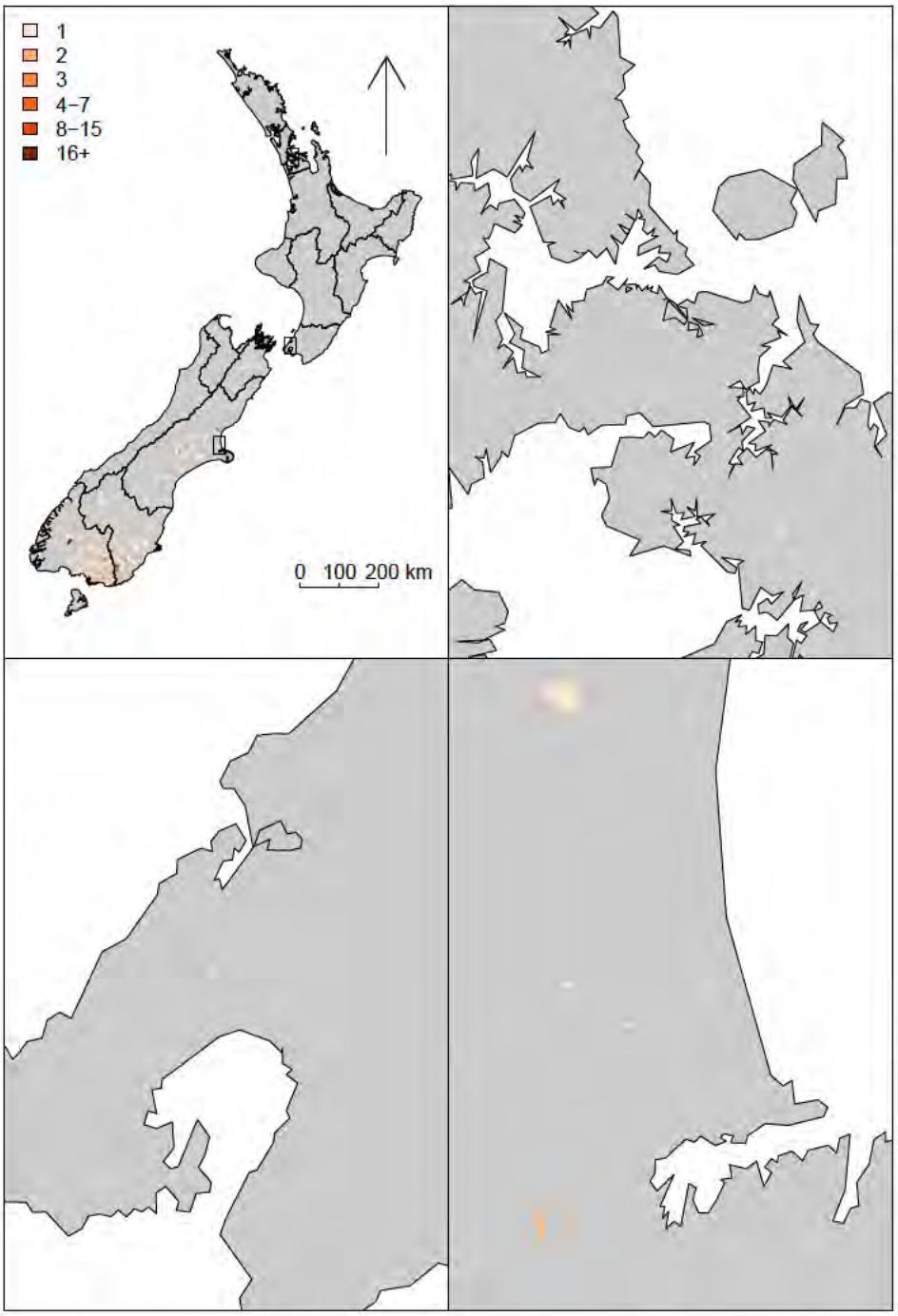

Figure 3.3: Number of *Salmonella* Brandenburg Cases between 2000 and 2009 at Meshblock level for all of New Zealand (top left), with inserts for Auckland (top right), Wellington (bottom left) and Christchurch (bottom right).

The spatial patterns in observed case numbers over time was examined by mapping reported case numbers per year in each meshblock for the ten most prevalent serotypes. For illustration, Figures 3.4 and 3.5 show the total number of cases of *S.* Brandenburg and Typhimurium DT160 for the years 2000, 2003, 2006 and 2009 respectively at the meshblock level. In 2000 there were 151 reported cases of *S.* Typhimurium DT160 and the majority of these were located in North Canterbury. Over the next 2 years a large increase in the number of reported cases of *S.* Typhimurium DT160 with a much higher dispersion across the country was observed. The following years show a steady decrease in the number of reported cases of *S.* Typhimurium DT160 whilst maintaining a fairly wide geographical distribution across the country.

The spatial pattern observed for *S.* Typhimurium DT160 is in stark contrast to that for *S.* Brandenburg, for which there were 168 reported cases in 2000 with the majority of these located in the lower South Island. The following years show a steady decrease in the number of reported cases of *S.* Brandenburg, the majority of these were again located in the lower South Island. Reported case numbers, by year, for these two serotypes and the eight other most prevalent serotypes can be seen in Table A.2. These patterns were also observed when the number of reported cases per 1000 population in each DHB were plotted against month for each of the ten most prevalent serotypes (Figures 3.6 to 3.8 and Figures A.4 to A.10).

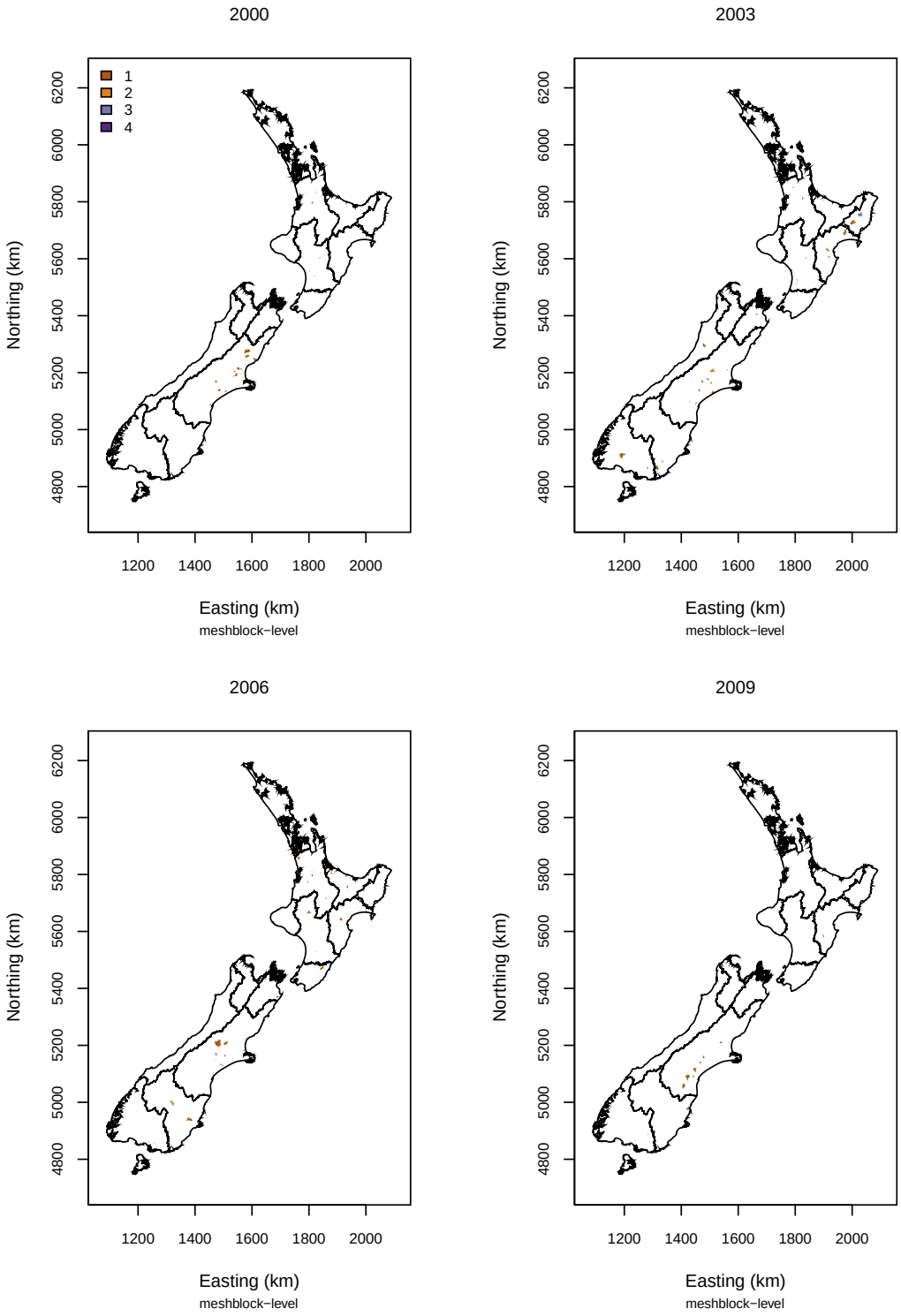

Figure 3.4: Number of *Salmonella* Typhimurium DT 160 Cases in 2000, 2003, 2006 and 2009 at Meshblock level. The use of 'Pan and Zoom' in the 'Tools' section of Adobe Reader is recommended for viewing details of urban areas.

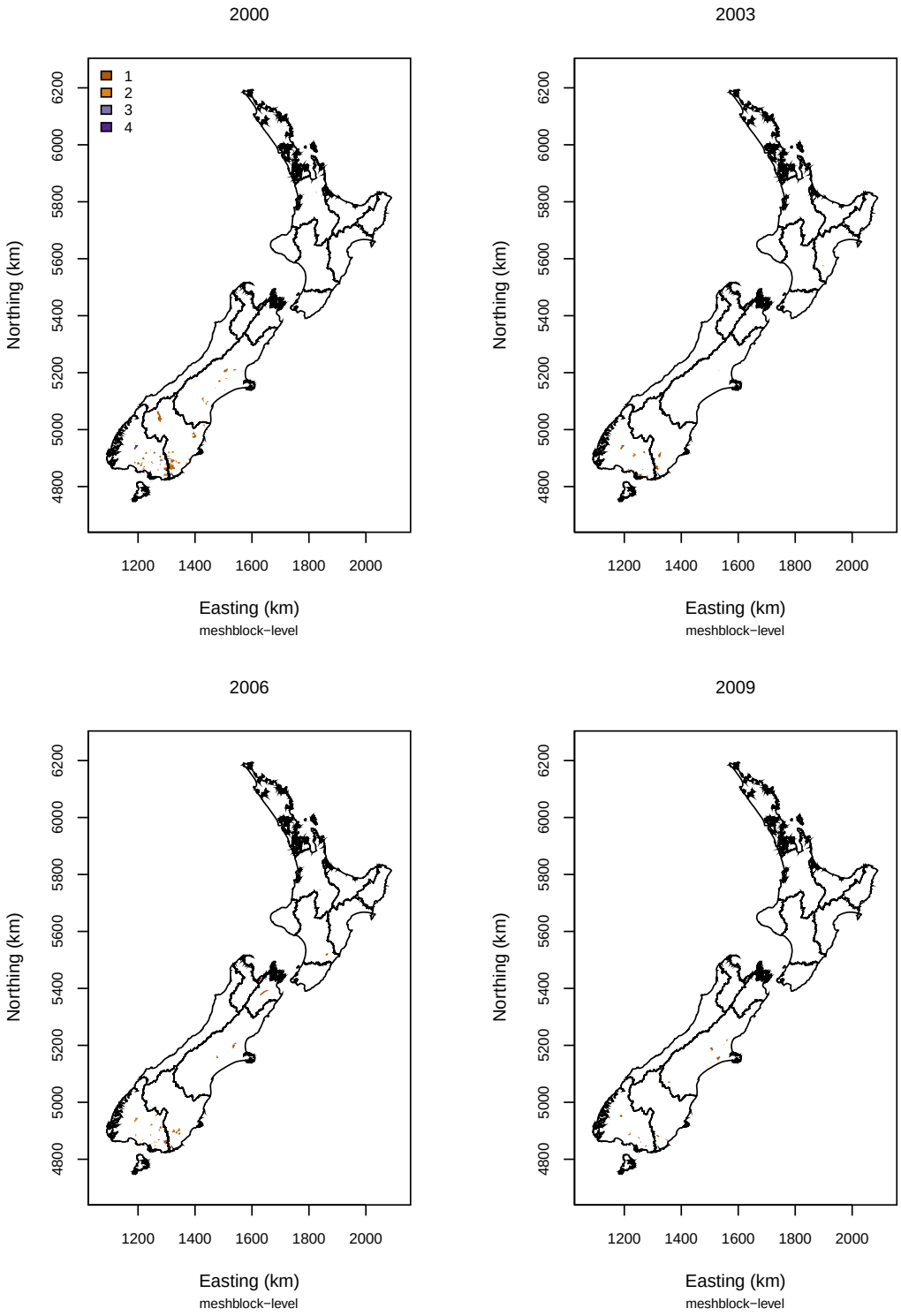

Figure 3.5: Number of *Salmonella* Brandenburg Cases in 2000, 2003, 2006 and 2009 at Meshblock level. The use of 'Pan and Zoom' in the 'Tools' section of Adobe Reader is recommended for viewing details of urban areas.

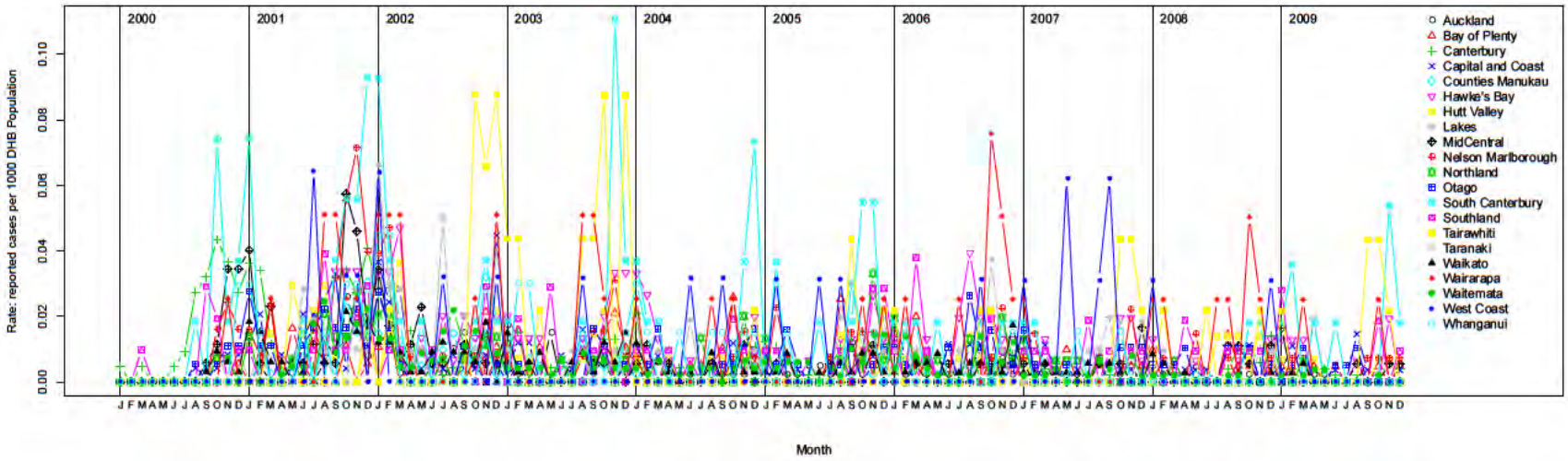

Figure 3.6: Number of reported cases of *Salmonella* Typhimurium DT160 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

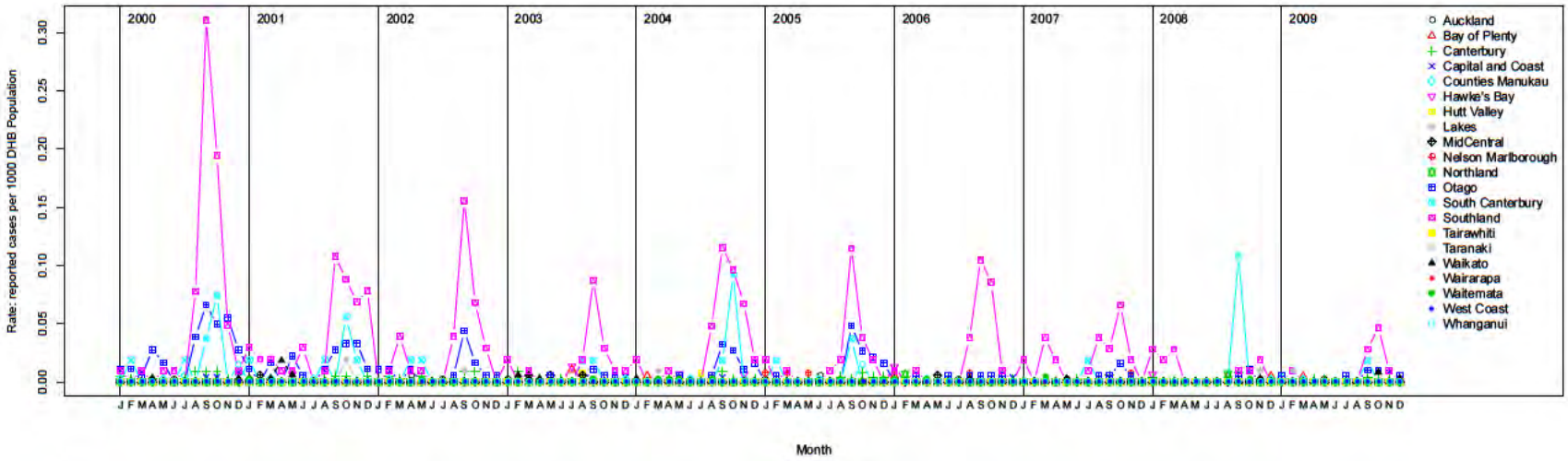

Figure 3.7: Number of reported cases of *Salmonella* Brandenburg per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

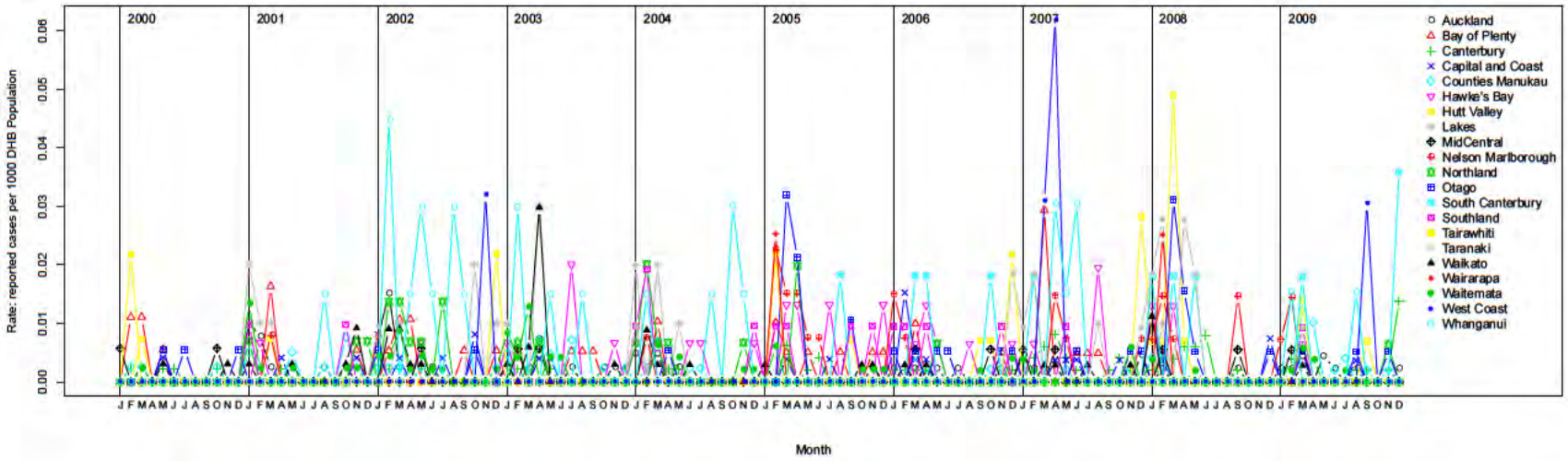

Figure 3.8: Number of reported cases of *Salmonella* Infantis per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

To investigate seasonal patterns in notification rates associated with ruminant densities in urban and rural areas we plotted box plots for six subpopulations (defined by urban/rural status and ruminant density per hectare for sheep, dairy and beef), showing the estimated rates in each month of the ten year time series at both the species level and for each of the ten most common serotypes. Figure 3.9 shows the relationship between ruminant densities and the seasonal pattern of all salmonellosis cases. Figure 3.10 shows the equivalent plots for *Salmonella* Typhimurium DT160 and Brandenburg (see Figures A.11 to A.14 in Appendix A.1 for the plots pertaining to other serotypes).

For all salmonellosis cases (all serotypes) these plots (Figure 3.9) showed marked differences between the areas with low and high ruminant densities and areas with no livestock (although there were few urban areas with livestock, there were sufficient numbers of meshblocks to be included in this analysis). Peak incidence of salmonellosis in the rural areas with low and high ruminant densities was observed in the spring months associated with the New Zealand calving and lambing period (August to October). In contrast, the peak months were in the summer in other areas (November to February). These plots also show a number of other features: the higher rates in rural areas for most years; the greater variability in rural areas and evidence of a difference in seasonality in rural areas with higher ruminant densities compared to rural areas with no livestock. However when these seasonal patterns were investigated at the serotype level some important, but subtle differences were observed (Figure 3.10). For *S.* Brandenburg notification rates in rural areas were seen to peak in spring, with the highest rates observed in September, most notably in areas with high sheep densities. For *S.* Typhimurium DT160 there was also evidence of a peak in spring in rural areas, but this was approximately one month after the peak observed for *S.* Brandenburg, particularly in areas with high ruminant densities.

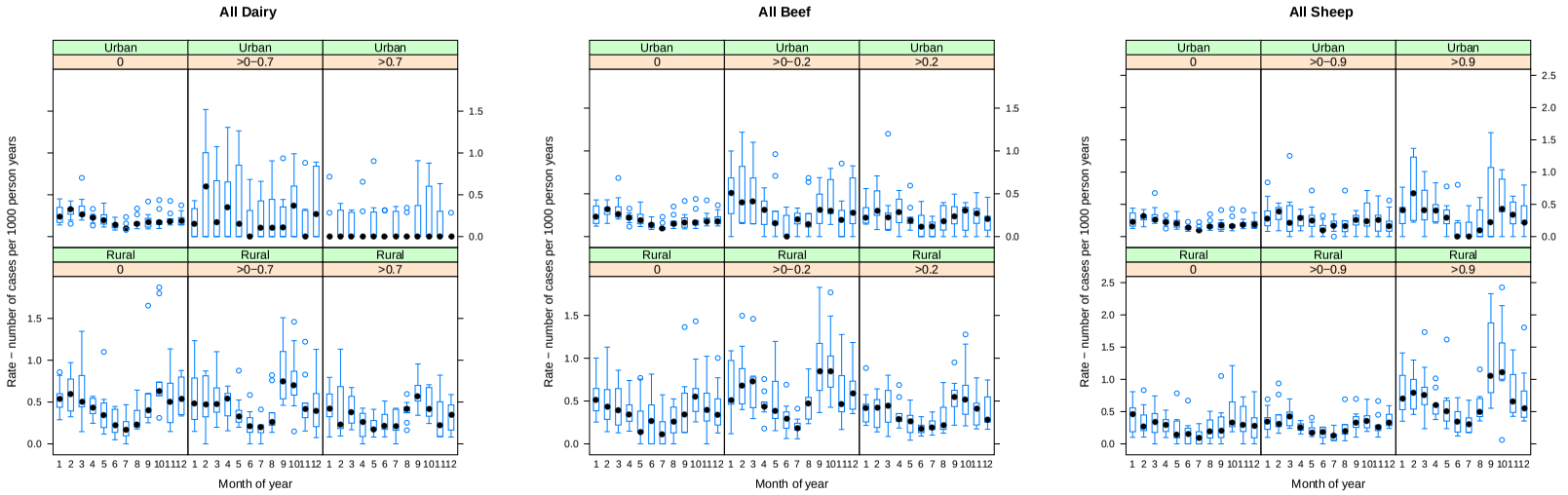

Figure 3.9: Box plots showing the seasonal pattern of rates of human salmonellosis (all serotypes) stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep (right). Each box summarises the rates for each month between 2000 and 2009.

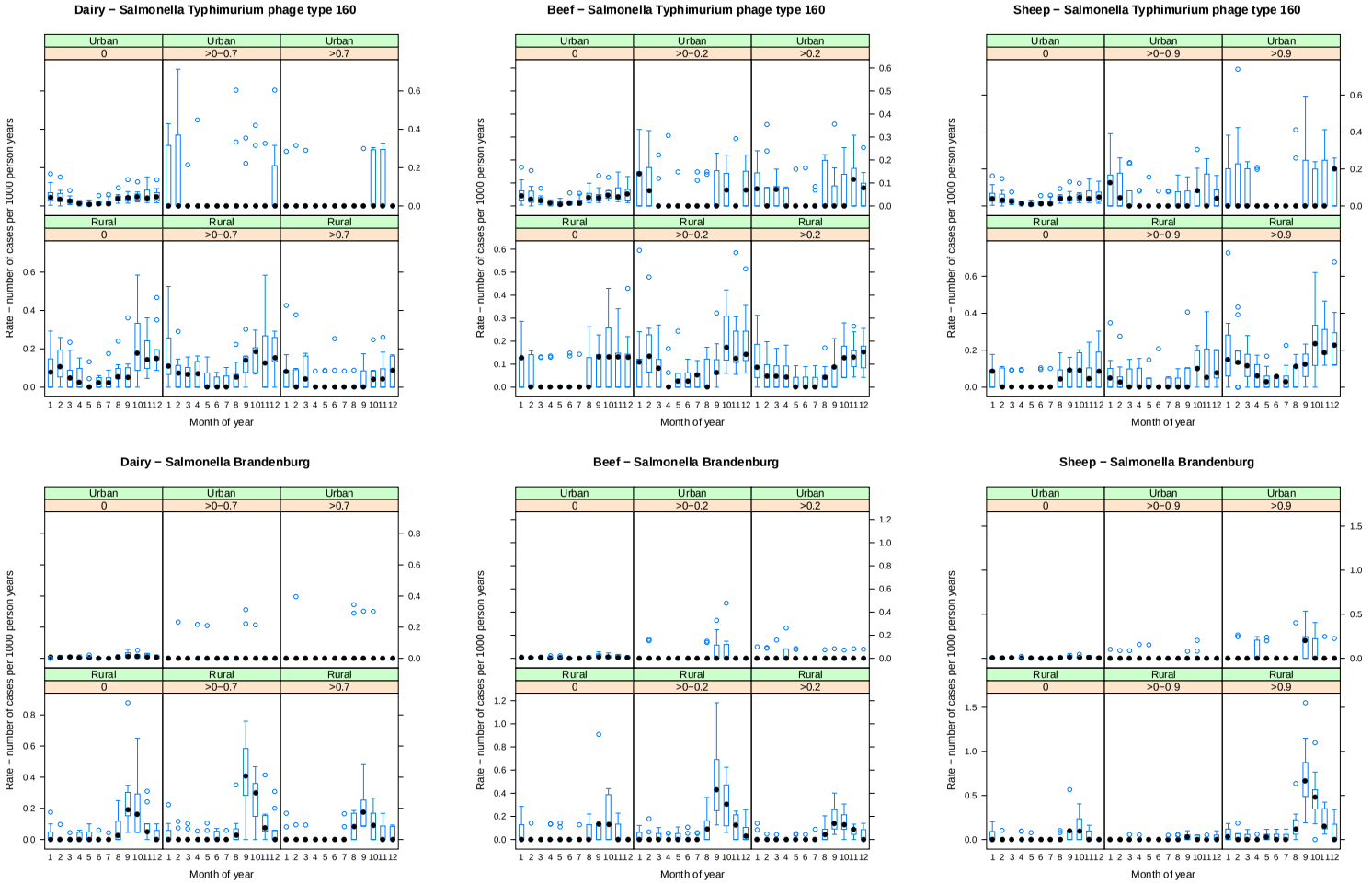

Figure 3.10: Box plots showing the seasonal pattern of rates of human salmonellosis stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep (right) for *Salmonella* Typhimurium DT160 (top row) and *Salmonella* Brandenburg (bottom row). Each box summarises the rates for each month between 2000 and 2009.

Section 4

Modelling Methodology

4.1 Determining Background Spatial and Temporal Trends

4.1.1 Adapting the model developed for campylobacteriosis

The Bayesian hierarchical model developed for campylobacteriosis [20, 27] was based on the collection of models described in Diggle et al. [12]. This model attempts to capture the spatial and temporal variations in infection risk for the data being analysed.

Let $Y_{i,t}$ represent the number of reported cases of *Salmonella* in spatial unit i for week t. The model assumes that $Y_{i,t}$ follows a Poisson distribution with mean $n_i\lambda_{i,t}$, where n_i is the number of people usually resident in spatial unit i, and $\lambda_{i,t}$ is the probability an individual residing in spatial unit i presents with salmonella in week t (i.e. the expected risk at this point in time and space). It is then assumed that the log of this risk may be split into two separate components, so that

$$\log(\lambda_{i,t}) = R_t + U_i, \tag{4.1}$$

where R_t is the purely temporal component and U_i is the purely spatial

component. Structure is provided on these components by specifying prior distributions on R_t and U_i so that they represent the underlying temporal and spatial trends in the data.

For the spatial component a Gaussian Markov Random field prior is assumed (also called a Gaussian intrinsic auto-regression) in which the risk in each spatial unit is assumed to be similar to the mean risk of the neighbouring spatial unit. More formally, the model assumes the following full conditional distribution

$$U_i \sim N\left(\sum_{j \in n(i)} \frac{U_j}{|n(i)|}, \frac{1}{\kappa_U |n(i)|}\right).$$

where $n(i)$ is the set of spatial units neighbouring spatial unit i and the hyper-parameter κ_U is assumed to have a non-informative conjugate Gamma prior.

For the temporal component the model assumes a Gaussian second order random walk prior: that the change in risk from week t to week $t+1$ will be similar to the change in risk from week $t-1$ to week t, i.e. given $R_1, \ldots, R_t$,

$$R_{t+1} - R_t \sim N\left(R_t - R_{t-1}, \frac{1}{\kappa_R}\right).$$

where once again, a conjugate Gamma prior is used for the hyper-parameter κ_R, see [27]. The model also assumes flat priors for R_1 and R_2 so that the temporal component can absorb the baseline level of risk.

An extra term $W_{i,t}$ was added to this model to account for any spatio-temporal interaction and to allow the fitting of epidemic indicators [19], so that

$$\log(\lambda_{i,t}) = R_t + U_i + W_{i,t}, \tag{4.2}$$

Samples from the posterior distribution were obtained using MCMC methods. The spatial component U_i was updated using a mixture of Metropolis Hastings proposals and single site conditional prior proposals (Knorr-Held, 1999). The temporal component R_t was updated using a mixture of Metropolis Hastings proposals and conditional prior proposals in blocks of lengths 4, 5, 9 and 11. The hyper-parameters κ_U and κ_R were updated with Gibb's steps. Multiple chains were run from randomly generated starting values for 40,000 iterations with a thinning of 20 after a burn-in period of 2000

iterations. The spatial units considered for this report were meshblock and territorial authority.

4.1.2 Estimation of Spatial relative Risk

The recently released R package **sparr** [11] offers a state-of-the-art algorithm for calculating kernel-smoothed relative risk estimates. This is achieved via a density-ratio method where by the densities of two point processes (representing case and control data or, in this study, case data from two different serotypes or age groups) are estimated using a newly developed kernel density estimation method. The ratio of the two estimated densities is then used to produce a smoothed log relative risk surface comparing the case density to the control density. This package also provides tools for constructing asymptotically derived p-value/tolerance surfaces, which highlight sub-regions of "extremity" in the risk surface that are statistically significant, along with flexible visualisation tools for displaying the relative risk and tolerance surfaces.

4.2 Identifying Possible Risk Factors

The spatial units considered for the following models were meshblock, water zone, or DHB.

4.2.1 Zero-Inflated Poisson Models

Zero-Inflated Poisson (ZIP) models were used to identify possible risk factors for the 10 most prevalent *Salmonella* serotypes.

Zero-Inflated (ZI) models can be applied in situations where clustered data results in a greater proportion of zero counts than can be expected in a Poisson model. This might for example arise where a proportion of the population of interest have no chance to be infected. This gives rise to a process which starts by determining whether an observation (be that an individual or a spatial unit) is likely to be infected or not and then determines

the number of cases among those who are at risk of infection. Zero-Inflated Poisson (ZIP) models assume that the number of cases amongst those who are at risk of infection follows a Poisson distribution and that the first step in this process can be modelled by a Binomial distribution. As stated by Zeileis et al. [32] ZIP models are two-component mixture models combining a point mass at zero with a truncated Poisson distribution. Thus, there are two sources of zeros: zeros may come from both the point mass and from the count component. For modelling the unobserved state (zero vs. count), a binary model is used: in the simplest case only with an intercept but potentially containing regressors. However it was not computationally feasible to include regressors in the binary model due to the size of the data being analysed (at meshblock level there are 388,390 observations and at the DHB/month level there were 20,160 observations).

More formally, the zero-inflated density is a mixture of the point mass at zero $I_{\{0\}}(\mathbf{y})$, a truncated Poisson distribution $f_{count}(\mathbf{y}; \mathbf{x}, \mathbf{y}, \beta, \gamma)$ and a binomial Generalised Linear Model (GLM) $g(\pi_i) = \mathbf{z}_i^{\mathrm{T}}\gamma$ that may depend on further regressors z_i:

$$f_{zeroinfl}(\mathbf{y} : \mathbf{x}, \mathbf{y}, \beta, \gamma) = \pi \cdot I_{\{0\}}(\mathbf{y}) + (1 - \pi) \cdot f_{count}(\mathbf{y} : \mathbf{x}, \beta) \tag{4.3}$$

where $\mathbf{y}$ is the dependent variable (containing n outcomes), $\mathbf{x}$ is a set of regressors, β and γ are vectors of regression coefficients and π is the unobserved probability of belonging to the point mass component. The corresponding regression equation for the mean is

$$log(\mu_i) = (1 - \pi) \cdot \mathbf{x}_i^{\mathrm{T}}\beta \tag{4.4}$$

The vector of regressors in the zero-inflation model $\mathbf{z}_i$ and the regressors in the count component $\mathbf{x}_i$ need to be distinct; in this case just the intercept, $\mathbf{z}_i = 1$ was considered. The link function $g(\pi)$ in the binomial GLMs was chosen to be the logit link, as this is commonly used in binomial GLMs [32]. Jackson [16] has developed an R package called **pscl** which contains the function `zeroinfl()` for implementing ZIP models as developed by Zeileis et al.[32]. This function has been used in the analysis of the data at the meshblock and DHB level, with population used as an offset.

Unlike the model developed for campylobacteriosis under SCIG-MAS-001 the ZIP models used for this investigation do not account for any spatio-temporal dependencies over and above those contained within the covariate structure. There are at present no computationally efficient methods for implementing ZIP models with the type of complex random effect structures that would be required to account for such spatio-temporal dependencies. In particular MCMC methods for ZIP models are not computationally feasible on datasets of this size. A consequence of this is that we have therefore been conservative in our interpretation of the p-values associated with the ZIP models that were developed.

Initially several regressors were fitted individually (in the count component) at the genus level to assess their relationship with the number of reported cases of *Salmonella*. From these models a multivariate model at the genus level was fitted. The predictors considered in the initial univariate models included; DHB, SDI, SDIFactor, sheepdensC, dairydensC, poultrydensC, sheepdens, dairydens, poultrydens, UrbanRural, URProfile, Island and Zcode. Table A.3 gives a description of these and additional variables considered in the modeling process. From these initial models the regressors that showed a strong association with the number of reported cases of *Salmonella* were used to develop multivariate models for each of the ten most common serotypes. To draw power from the models generated for the ten serotypes investigated, a combined model for all ten of the serotypes was then fitted by nesting the regressors within the variable serotype in the count component. Several of these combined models were fitted testing for an association between different regressors and the number of reported cases of *Salmonella*. Tables A.4 to A.6 list the models fitted in the count component at both the genus and serotype level, along with its corresponding AIC value.

The AIC value (Akaikes Information Criterion [2]) is a popular model selection criteria that can be used to aid in the construction of a suitable regression model. The magnitude of this value indicates how well the model fits the data. The smaller the value is the better the fit. For example, model 3 has a lower AIC value than model 2, indicating that model 3 is capturing more of the variation in the number of reported cases than model 2 (i.e. model 3 is a better fit).

The observations used in models 1 to 24 consisted of all the meshblocks across New Zealand, for model 25 the observations where aggregated to DHB for each month. The first 10 models were fitted for all the reported cases of *Salmonella* over the 10 years of data that was provided (i.e. at genus level). Models 11 to 17 were fitted for each of the ten most prevalent serotypes and models 18 to 25 are the combined models, each of which fitted a single model for all ten serotypes.

The pattern of fitted values from the ZIP model reflects (in a qualitative sense) the relative incidence of disease across the meshblocks/DHBs. However, the absolute magnitudes of the fitted values do not match the observed counts. This is inevitable in this type of ZIP application. Specifically, each meshblock/DHB is modelled as having a very small chance of disease in any given time. When conditioning on disease being present (i.e. a count of one or more) the expected numbers may be moderately large. However, the overall (unconditional) fitted value is calculated as a combination of contributions from the zero-inflated component and the count component, producing an averaged set of meshblock/DHB values that varies far less than the (highly clustered) observed counts themselves.

The results from models 24 and 25 (Table A.6) will be described in the following sections.

4.2.2 Zero-inflated models for water zone

A separate investigation of risk associated with water zones was done in order to ascertain whether the risk of salmonellosis was associated with source (surface, ground or roof) or water treatment, and whether there were any water zones in which risk was abnormally high.

Some 9780 cases of *Salmonella* had water zone information available, with 381 water zones out of a total of 2302 containing cases. Given the large proportion of water zones containing no cases, a zero-inflated analysis was deemed appropriate. It was found that none of the roof supplies had any cases of *Salmonella*, so these were removed from the analysis that followed, leaving a total of 1747 water zones. Of the covariates available at the water-zone level, the main covariates of interest were year, water source (ground,

surface, roof, or a combination thereof), and treatment type (None, Chlorine, UV etc.). Due to the large range over which the counts varied (up to 130 in a single water zone per year), zero-inflated Negative Binomial (ZINB) models were deemed more appropriate than ZIP models, and these were fitted using the **pscl** package in R.

As the data were aggregated per-year to the water zone level, the residuals associated with each water zone may be used to identify those water zones with unusually high counts. Randomized quantile residuals [13] were generated for this purpose, with the residuals for each year for a water zone being averaged to obtain a measure of how far each water zone differs from the average. Water zones with large average residuals are those that have larger than expected counts.

In addition to this analysis, we employed several other modelling strategies, such as including only water regions that have at least one case, and adding water zone as a random effect. These other strategies gave broadly similar results as the above model (data not shown).

Section 5

Results

5.1 Determining Background Spatial and Temporal Trends

5.1.1 Temporal Trends

The temporal analysis of all reported salmonellosis cases at the species level (Figure 5.1) shows an increase in the number of reported cases from January 2000 to early 2002 followed by a large drop in the number of reported cases. From this point the average number of reported cases of salmonellosis remained fairly constant, with only seasonal fluctuations.

The investigation into the temporal variation within the notifications for each of the ten most prevalent serotypes (Table 3.1) revealed a range of different temporal trends for 6 of the 10 serotypes investigated. The remaining 4 serotypes investigated all showed similar patterns over time. The temporal trends for the ten most prevalent serotypes are shown in figures 5.2 to 5.4 and Figures A.15 to A.21.

The trend seen at the species level was also observed at the serotype level for *S.* Typhimurium DT160, as seen in Figure 5.2. This shows an epidemic pattern, with a marked rise, with seasonal fluctuations, to a peak in 2001, followed by a steady decline to 2009. However different trends were observed

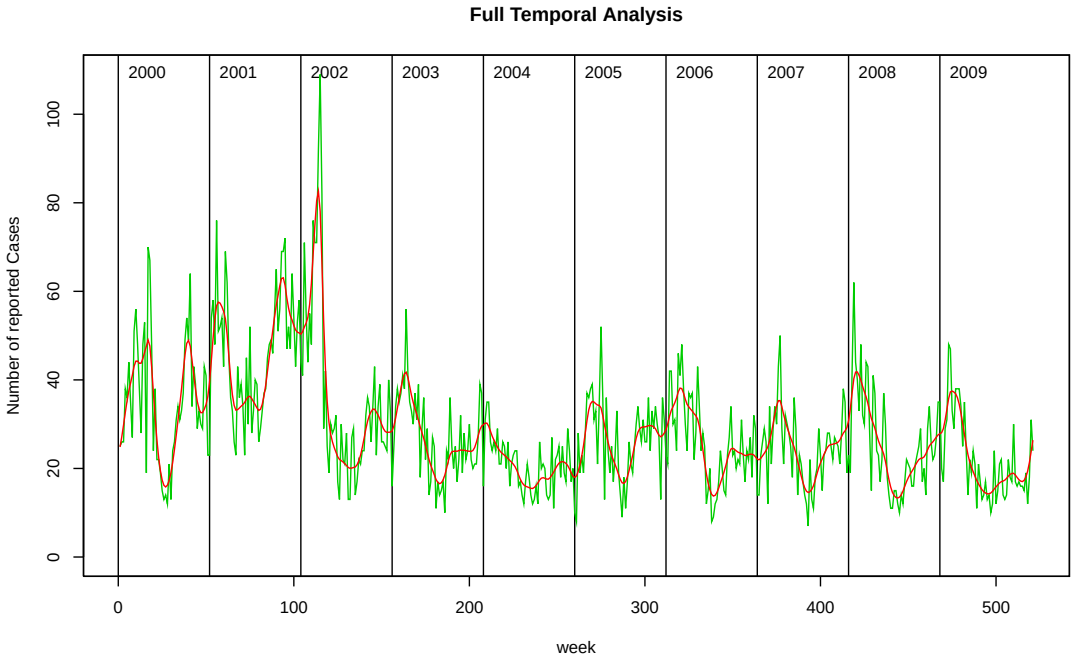

Figure 5.1: Results from the full temporal analysis at the species level, using the model described in section 4.1.1. The green line shows the raw data and the red line is the smoothed model-fit to the raw data.

for the other serotypes investigated. For example, *S.* Typhimurium DT1 showed a fairly constant average number of reported cases across the 10 years of data except at week 110 (early 2002) when there was a very large number of reported cases (see Figure A.15). This peak could have been due to a large outbreak as reported by [1].

The temporal analysis for *S.* Typhimurium DT135 (Figure A.16) showed a fairly rapid decrease in the number of reported cases over the 10 year period, whereas that for *S.* Brandenburg (Figure 5.3) showed a gradual decrease in the number of reported cases over the same period.

Preliminary investigation into differences in reported case numbers between the North and South Islands for the ten most prevalent serotypes showed an interesting temporal trend for *S.* Infantis, Figure A.22. From this figure we see that the majority of reported cases occurred in the North Island whereas

in the South Island there were very few cases in the first 5 years (250 weeks), followed by a noticeable increase in the number of reported cases in the last 5 years of recorded data.

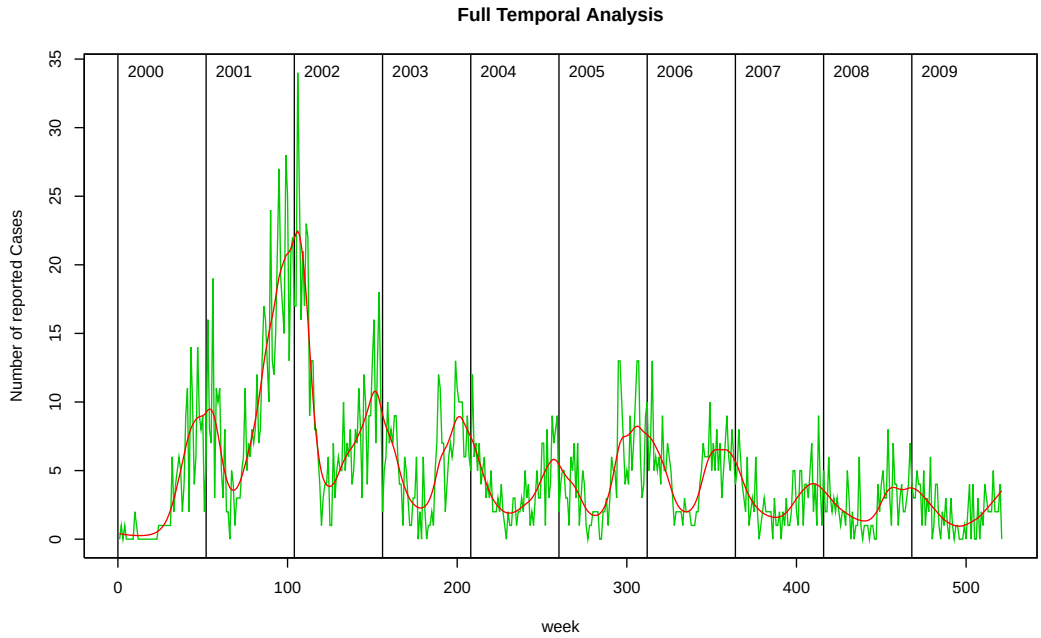

Figure 5.2: Results from the full temporal analysis for *S.* Typhimurium DT160, using the model described in section 4.1.1. The green line shows the raw data and the red line is the smoothed model-fit to the raw data.

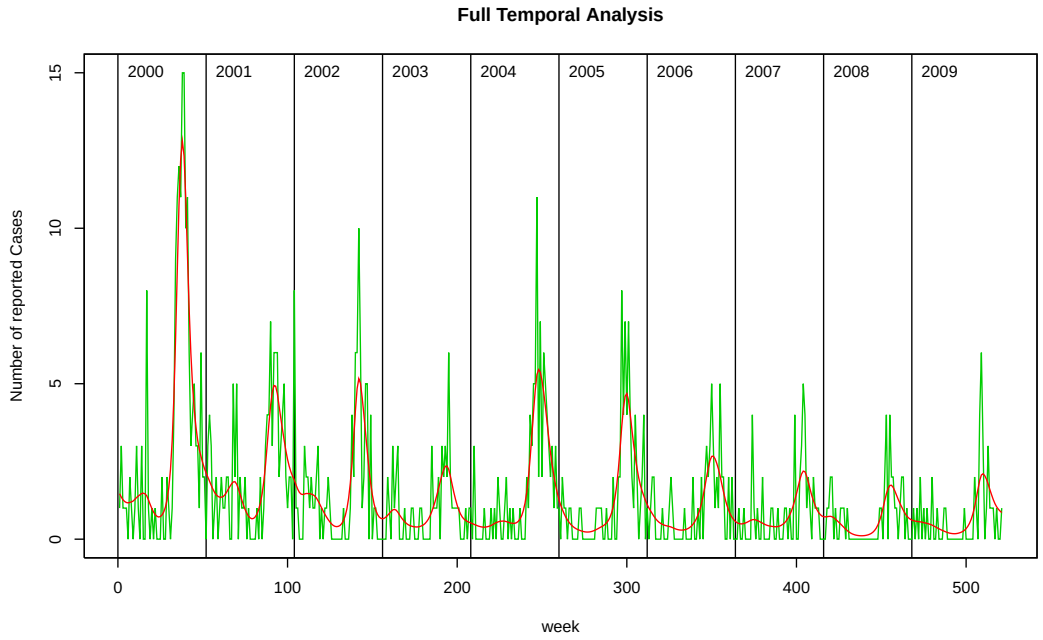

Figure 5.3: Results from the full temporal analysis for *S.* Brandenburg, using the model described in section 4.1.1. The green line shows the raw data and the red line is the smoothed model-fit to the raw data.

5.1.2 Spatial Patterns

The model developed for campylobacteriosis [20, 27] was modified to determine the background spatial and temporal trends of salmonellosis at both the species and serotype level. The model was implemented at the meshblock level when the data were analysed at the species level and at the territorial authority level when the data were analysed at the serotype level.

Figure 5.5 shows the relative risk surface for the analysis at the species level and meshblock. This shows that for the ten year period there was a higher risk of notification of all salmonellosis cases in the South Island compared to the North Island. The relative risk is interpreted as the risk of salmonellosis in each spatial unit (meshblock or TLA as below), compared to the average risk across all units. Therefore a value greater than 1.0 indicates a higher than average risk, whereas a value less than 1.0 indicates a lower than average

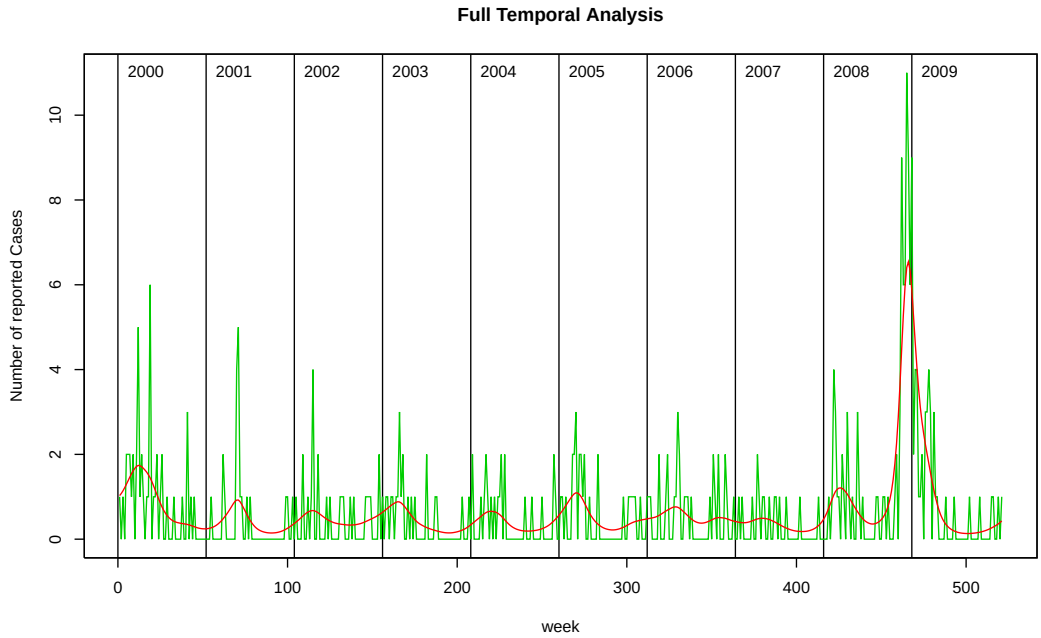

Figure 5.4: Results from the full temporal analysis for *S.* Typhimurium DT42, using the model described in section 4.1.1. The green line shows the raw data and the red line is the smoothed model-fit to the raw data.

risk.

When the model was applied to each of the ten most prevalent serotypes a different spatial pattern in relative risk was observed for each serotype. Most of these patterns were similar to that observed at the species level, in that a higher number of reported cases were seen in the South Island. However for a few serotypes this pattern was reversed. The spatial patterns for the ten most prevalent serotypes are shown in Figures 5.6 and A.23 to A.24.

The spatial trend in relative risk for *S.* Typhimurium DT160 showed a marginally higher risk in Canterbury compared to the rest of the country, as seen in Figure 5.6, however, the risk was relatively uniform across all TLAs. In contrast *S.* Typhimurium DT1 cases were more likely to be notified in upper South Island than anywhere else in New Zealand (Figure 5.6). The spatial pattern for *S.* Brandenburg (Figure 5.6) was similar to that observed at the species level, with a much higher risk of notification in the lower

South Island compared to the rest of New Zealand. The spatial pattern for *S.* Typhimurium DT135 shows areas of elevated risk in the Tararua District, Gisborne and Dunedin.

These spatial patterns were also observed when the fitted values from model 24 were mapped at meshblock level. These maps are presented in Figures A.25 to A.27 in Appendix A.3.2

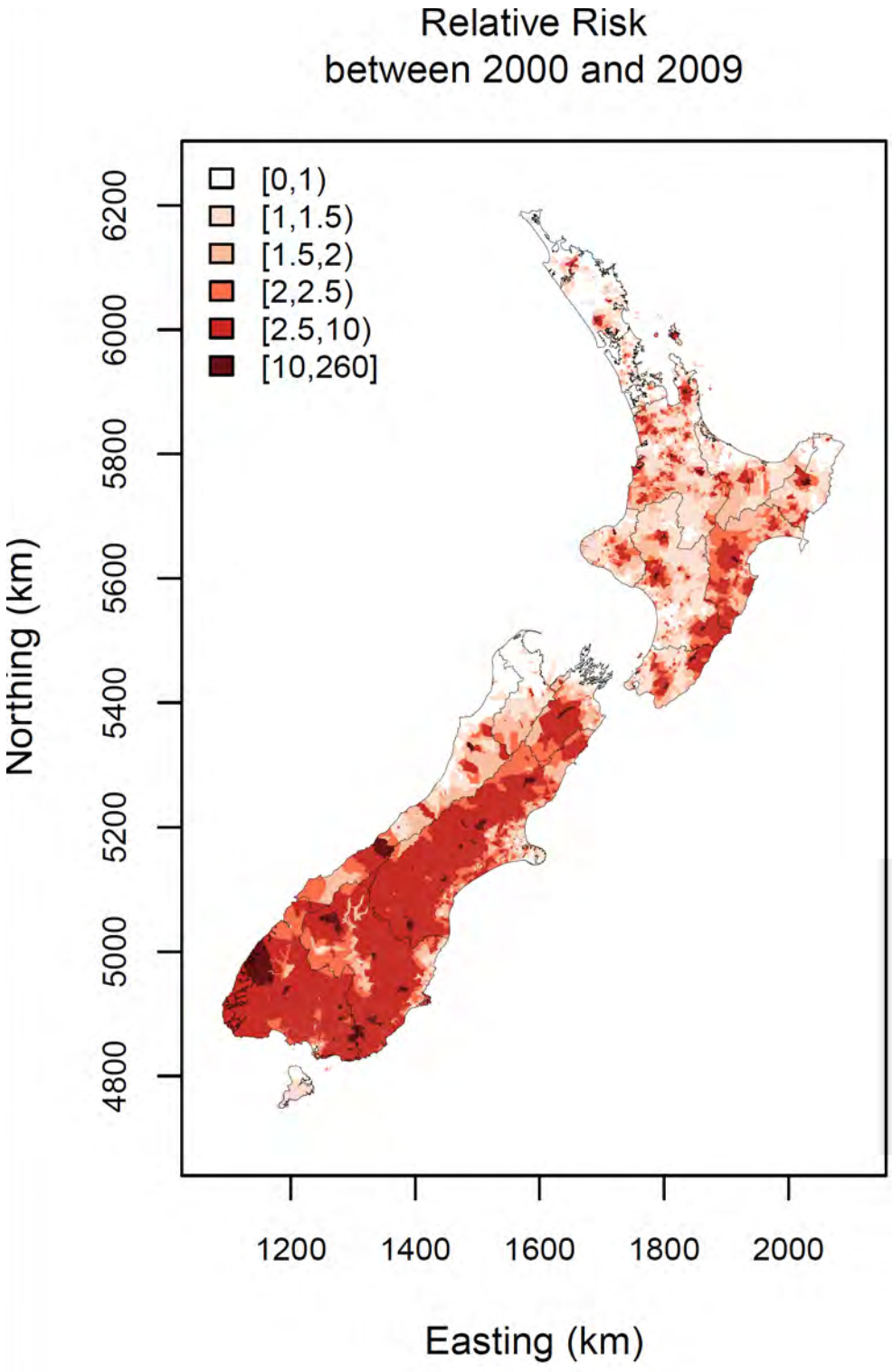

Figure 5.5: Relative risk surface obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at meshblock level for all of the reported cases (that is at species level). In the legend the square bracket denotes inclusive of the adjacent value, whereas the round bracket denotes up to but exclusive of the adjacent value.

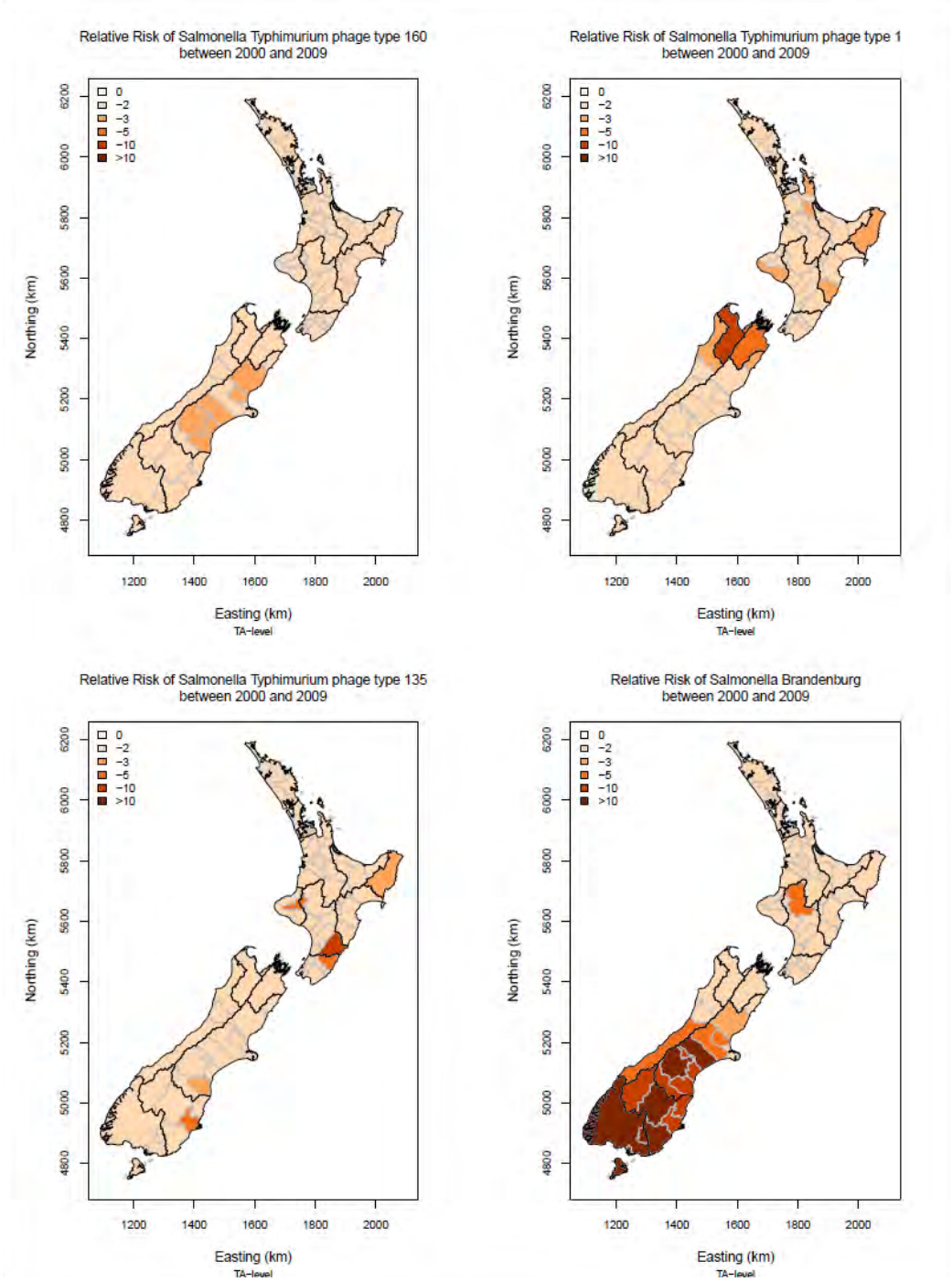

Figure 5.6: Relative risk surface obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *S.* Typhimurium DT160 (top left), *S.* Typhimurium DT1 (top right), *S.* Typhimurium DT135 (bottom left) and *S.* Brandenburg (bottom right).

The R package **sparr** [11] was used to investigate the (log) relative risk of notification for the ten most prevalent serotypes compared to each other on a pairwise basis, and compared to all other reported cases of salmonellosis at the meshblock level. An example of these two sets of analysis can be seen in Figures 5.7 and A.28. Figure 5.7 shows the log relative risk surface comparing *S.* Brandenburg notifications to all other reported cases of salmonellosis. This figure shows a significantly increased risk of *S.* Brandenburg notifications in the lower South Island, compared to all other reported cases of salmonellosis, as noted by the 5% significance line on this figure. Figure A.28 shows the log relative risk surface comparing *S.* Saint Paul notifications to *S.* Brandenburg notifications. This figure shows a significantly increased risk of *S.* Saint Paul notifications compared to *S.* Brandenburg notifications in the upper South Island and the lower North Island, as noted by the 5% significance line on this figure.

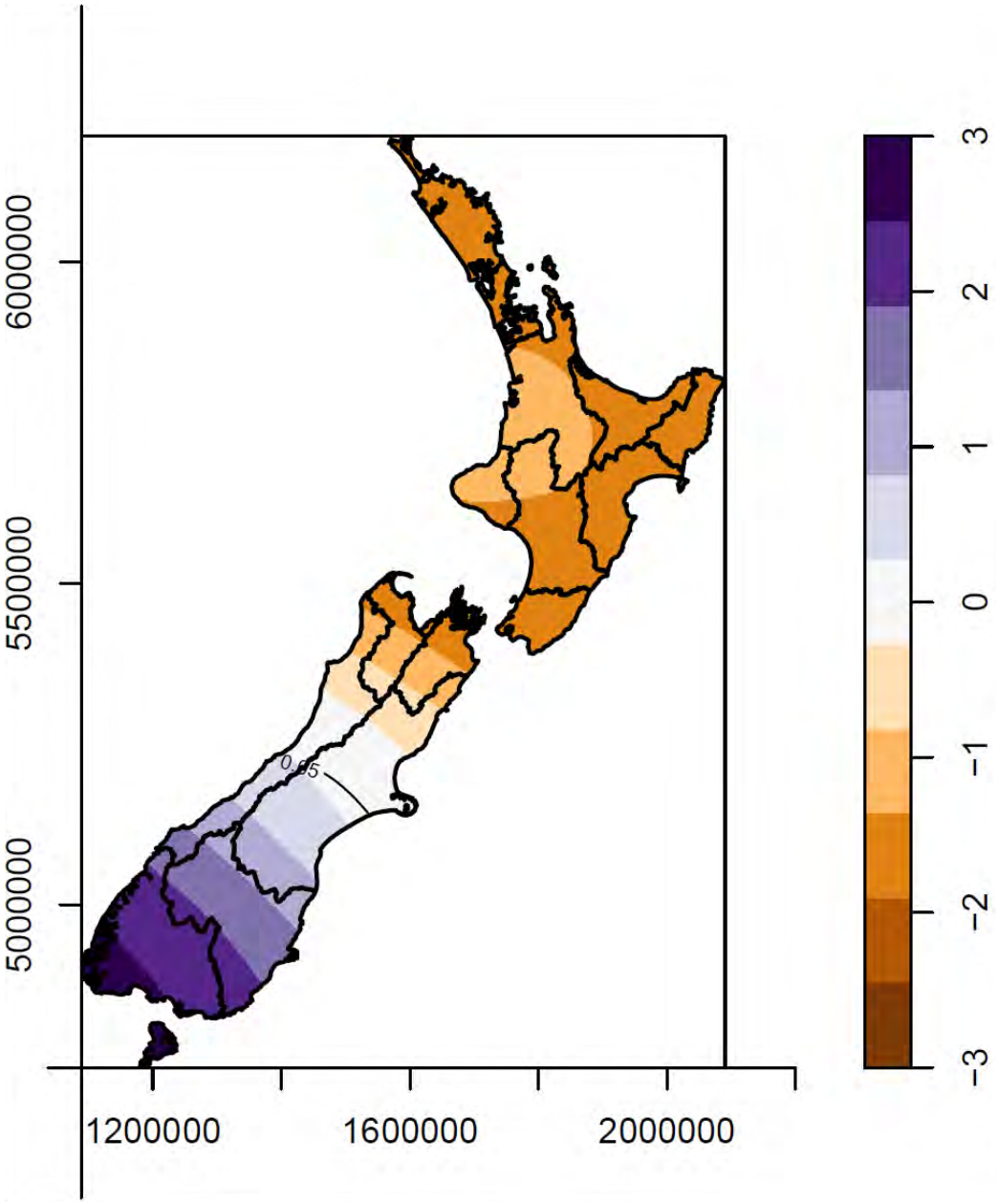

Figure 5.7: Log relative risk of *S.* Brandenburg compared to all other notified cases of *S.* at Meshblock level over the study period.

5.1.3 Anomalous Events / Outbreak analysis

The model developed for campylobacteriosis [20, 27] as shown in Equation 4.2 in the previous section was applied to each of the ten most prevalent *Salmonella* serotypes to detect outbreaks clustered in time and space. The aim of this analysis was to: 1) correctly identify known outbreaks; 2) identity other cases that could have been part of these outbreaks and; 3) identify any cases that were clustered in time and space but not attributed to any known outbreak. Figures 5.8 to 5.10 and Figures A.29 to A.34 show the weekly posterior probability of an outbreak for each of the 10 most common serotypes obtained from applying this model at the Territorial Authority level. The coloured dots indicate the number of reported cases observed at each time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point. These figures indicate that for each of the ten most prevalent serotypes there was at least 1 possible outbreak over the study period.

The Disease Outbreak Manual for Public Health Surveillance in New Zealand [14] states that for the purposes of reporting an outbreak case definition is:

- two or more cases linked to a common source
- a community-wide or person-to-person outbreak (except when the source has become well established as a national epidemic)
- any other situation where outbreak investigation or control measures are undertaken or considered.

Outbreak reporting is not required for single cases due to a specific contaminated source, or secondary cases following person to person spread, with the exception of secondary cases occurring in an institution.

S. Typhimurium DT160

The posterior probabilities of an outbreak for *S.* Typhimurium DT160 (Figure 5.8) identified 10 weeks in which there was a very high probability ($> 80\%$) of localised outbreaks. These all occurred within the first three years of the study period with the majority occurring in the latter half of

2000 and early 2001 in Christchurch City (Table 5.1), where there was a concurrent outbreak of the same serotype in wild birds, particularly sparrows (*Passer domesticus*), causing high mortality [3]. The concentration of DT160 cases in this location is also evident in Figure 3.4. Relatively few cases during this period were associated with notified outbreaks. Later outbreaks were identified in October 2001 and December 2002 and these were located in upper and lower North Island respectively. One major reported epidemic in South Auckland in February 2001 was not detected in our analysis [6]. However, examination of all cases of DT160 in the EpiSurv database during Febraury and March 2001 showed that most of these cases were either not notified, or were not entered onto the database (only 2 cases were notified in Manukau / South Auckland in February 2001, whereas the outbreak was reported to have affected 70 individuals).

Table 5.1: Territorial authorities of the reported cases of *S.* Typhimurium DT160 identified as outbreaks by the model shown in Equation 4.2. The number of reported cases and number associated with notified outbreaks are provided.

Week	Territorial Authority	Cases	Identified as outbreak
32 (Aug-2000)	Christchurch City	3	0
40	Timaru	4	0
(Oct-2000)	Christchurch City	2	0
41 (Oct-2000)	Christchurch City	4	1
43 (Nov-2000)	Christchurch City	2	1
47	Christchurch City	3	1
(Nov-2000)	Palmerston North City	5	0
53	Christchurch City	7	7
(Jan-2001)	Rangitikei District	1	1
59	Ashburton	2	1
(Feb-2001)	Christchurch City	6	6
95	Auckland City	8	1
(Oct-2001)	Manukau City	3	0
145 (Oct-2002)	Gisborne	4	4
154	Lower Hutt City	1	1
(Dec-2002)	Porirua City	2	1
	Wellington City	6	4

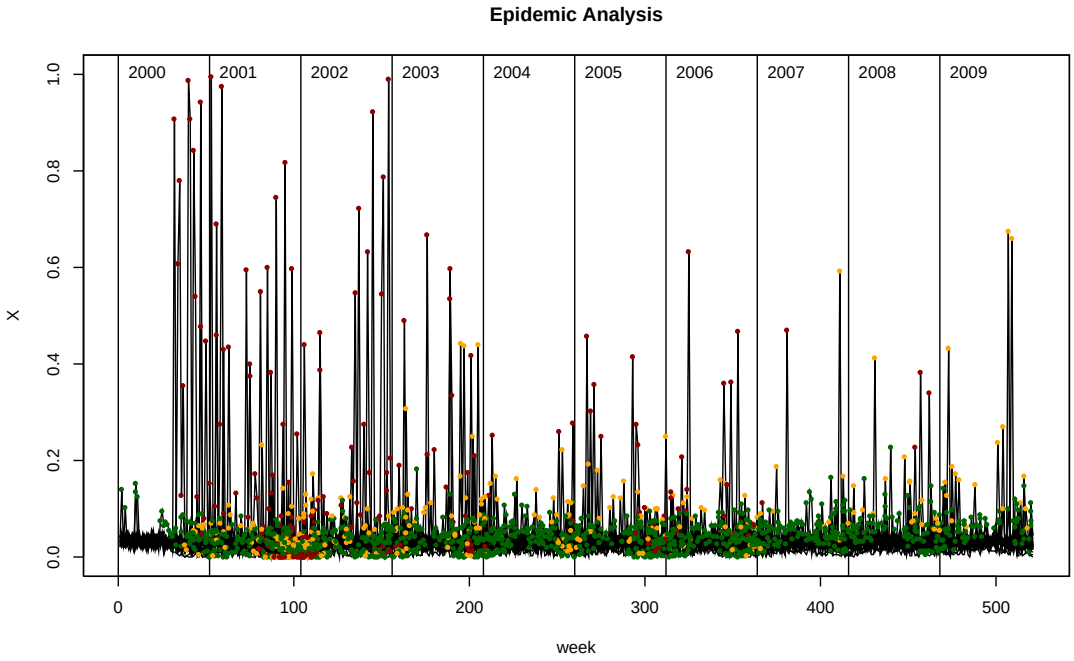

Figure 5.8: The weekly posterior probability of an outbreak (the value 'x' on the y-axis) obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *S.* Typhimurium DT160. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

S. Typhimurium DT1

The posterior probabilities of an outbreak for *S.* Typhimurium DT1 (Figure 5.9) showed 14 weeks in which there was a very high probability ($> 80\%$) of there being an outbreak. As Figure 5.9 shows these were clustered in two time periods; the first centered around weeks 113-117 (Feb-Mar 2002) and the second cluster around weeks 472-476 (Jan-Feb 2009). These two clusters had a total of 181 cases that were identified as being associated with known outbreaks in the following territorial authorities; Buller District, Gisborne District, Marlborough District, Nelson City, Tasman District. The other 4 weeks identified by this model a having a high probability ($> 80\%$) of being

an outbreak were weeks 284-285 (Jun 2005) with 15 cases, week 426 (Feb-Mar 2008) with 7 cases and week 432 (Apr-2008) with 6 cases. Of these only week 426 had cases that were identified as being associated with known outbreaks in Christchurch City, Selwyn District and Waimakariri District. Cases reported in June 2005 (weeks 284-285) and those reported in week 432 (April 2008) were not associated with any known outbreaks, although 6 cases in week 432 were reported within Christchurch city and the Selwyn District. The 15 cases in June 2005 were reported within Marlborough District, Nelson City and the Tasman District. An outbreak in Gisborne associated with contaminated watermelon in January 2009 was reported in the peer-reviewed literature [21].

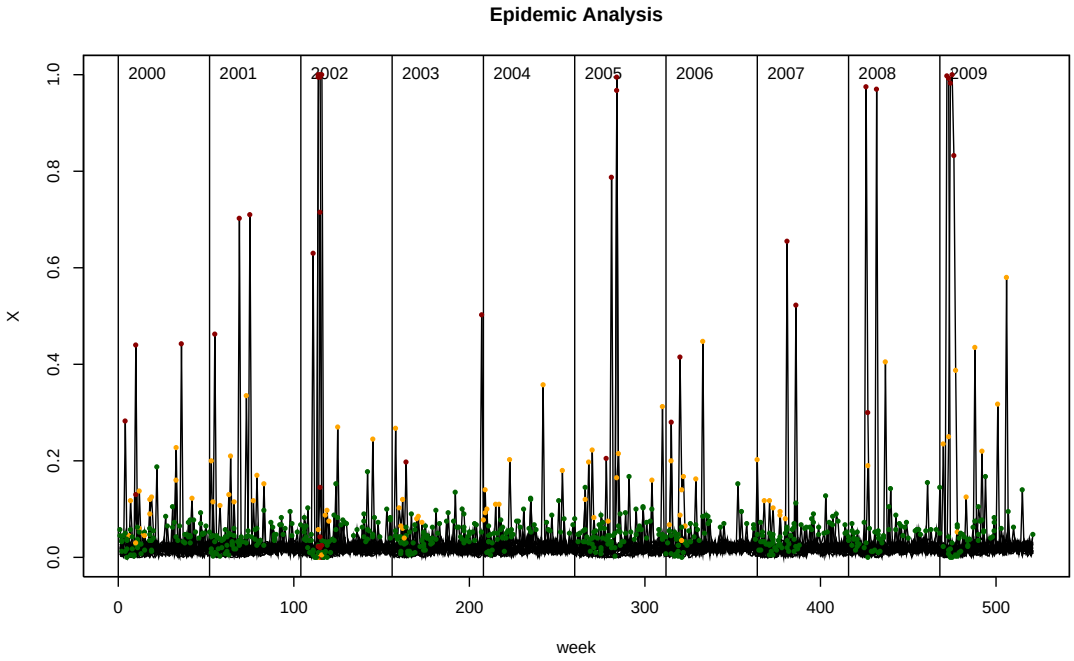

Figure 5.9: The weekly posterior probability of an outbreak (the value 'x' on the y-axis) obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *S.* Typhimurium DT1. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

S. Typhimurium DT135

The posterior probabilities of an outbreak for *S.* Typhimurium DT135 (Figure 5.10) showed 21 weeks having a very high probability ($> 80\%$) of there being an outbreak. The majority of these are clustered within the first two years of the study period (weeks 1-104). In 13 of these weeks, cases were identified as being associated with known outbreaks in both North and South Island in the following territorial authorities; Christchurch City, Lower Hutt City, Selwyn District, Manukau City, Masteron District, New Plymouth District, South Wairarapa District, North Shore City, Palmerston North City, Rotorua District, Stratford District, Tararua District, Tauranga City, Waimakariri District, Wanganui District, Wellington City and Western Bay of Plenty District. Given the large number of outbreaks over 21 weeks, for brevity we have just included those which were identified by the model but not notified as outbreaks in Table 5.2.

Table 5.2: Territorial authorities of the reported cases of *S.* Typhimurium DT135 identified as outbreaks by the model shown in Equation 4.2. In this table we only show the outbreaks identified by the model that occurred in weeks in which there were no formally notified outbreaks (there was a total of 21 weeks in which outbreaks were identified by the model and most of them were notified as outbreaks and are not shown in this table).

Week (Month-Year)	Territorial Authority	Cases
46 (Nov-2000	Dunedin City	2
57 (Feb-2001)	Marlborough District	5
87 (Jun 2001)	Dunedin City	6
95 (Oct-2001)	North Shore City	3
110	Gisborne District	3
(Feb-2002)	North Shore City	2
294	Lower Hutt City	2
(Sep-2003)	Wellington City	2
482 (Mar-2009)	Kawerau District	2
494 (Jun-2009)	Gisborne District	2

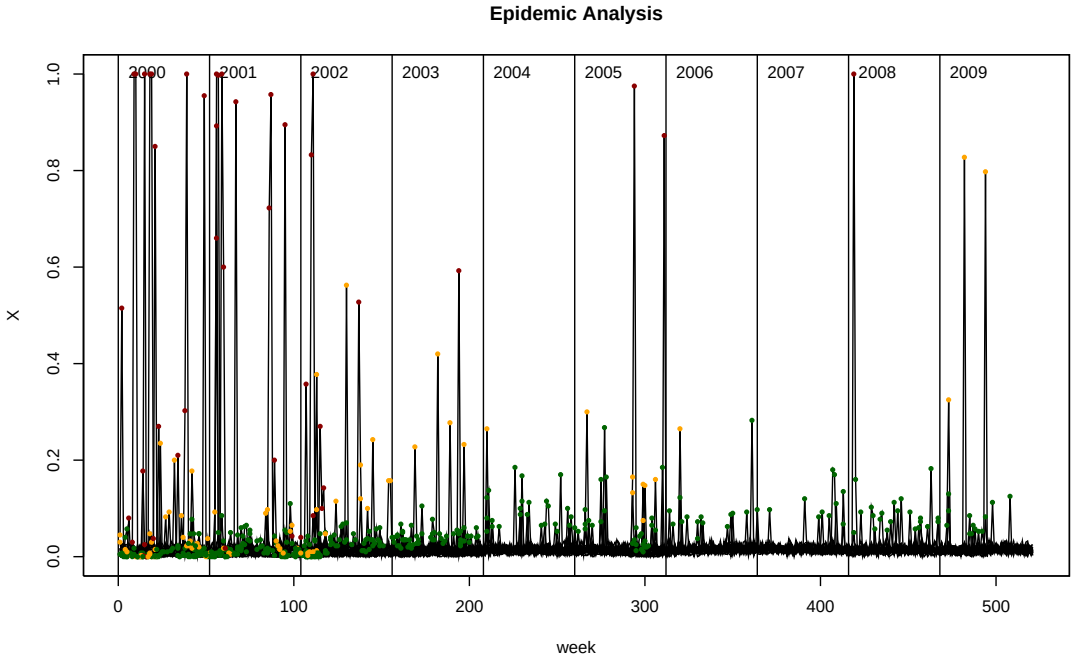

Figure 5.10: The weekly posterior probability of an outbreak (the value 'x' on the y-axis) obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *S.* Typhimurium DT135. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

S. Brandenburg

In contrast to other common serotypes, spatio-temporal clusters of cases caused by *S.* Brandenburg were relatively rare. The posterior probabilities of an outbreak for *S.* Brandenburg (Figure 5.11) shows two weeks (week 17 (Apr-May 2000) and week 68 (Apr 2001)) as having a very high probability ($> 80\%$) of there being an outbreak in those weeks. The number of notified cases for these two weeks were 8 and 5 respectively and none of these cases were attributed to a known outbreak. In week 17 (Apr-May 2000) 4 of the 8 notified cases were reported in the Clutha District, the other 4 cases were reported in Dunedin City (2 cases), Queenstown-Lakes District (1 case), and within Manukau City (1 case). In week 68 (Apr 2001) all 5 of the reported cases occurred in the Ruapehu District.

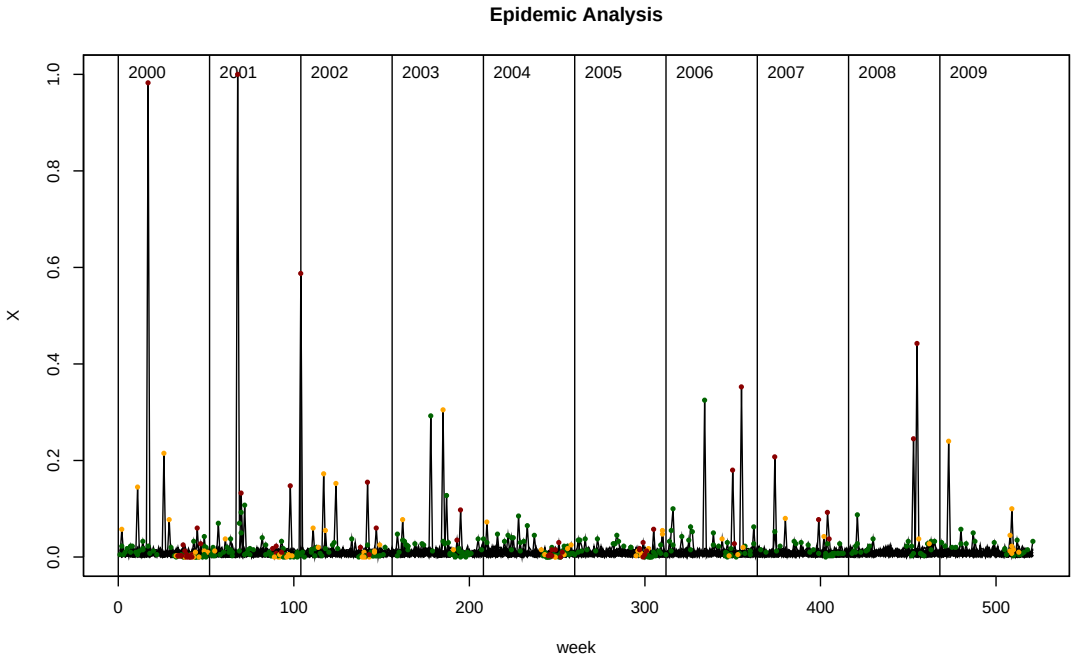

Figure 5.11: The weekly posterior probability of an outbreak (the value 'x' on the y-axis) obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *S.* Brandenburg. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

S. Typhimurium DT156

This model correctly identified a known outbreak of *S.* Typhimurium DT156 that occurred in March-April 2007 (weeks 376-378), by associating a posterior probability of being part of an outbreak of approx 99% to week 377 (Figure A.29). However of the 18 cases reported over this 3 week period only 12 were identified as being associated with a known, notified outbreak. As shown in Table 5.3 all but 1 of these 12 cases were reported in the lower North Island.

Table 5.3: Territorial authorities of the reported cases of *S.* Typhimurium DT156 in week 376 to 378 (March-April 2007) identified as outbreaks by the model shown in Equation 4.2. The number of reported cases and number associated with notified outbreaks are provided.

Week	Territorial Authority	Cases	Identified as outbreak
376 (Mar-2007)	Lower Hutt City	1	1
	Porirua City	2	1
	Rangitikei District	1	0
	Wellington City	2	1
377 (Apr-2007)	Lower Hutt City	5	3
	Porirua City	2	1
	Wellington City	5	4
378 (Apr-2007)	Central Otago District	1	1

S. Infantis

This model correctly identified the outbreak of *S.* Infantis that occurred within Dunedin City in March-April 2005 (week 274), by associating a posterior probability of being part of an outbreak of approx 99% to week 274 (Figure A.30). However of the 9 notified cases in this week only 5 of them were identified as being part of the notified outbreak. There was 1 reported case within each of the following TA's that was not identified as part of this outbreak, these were Dunedin City, Christchurch City, Invercargill City and Napier City.

From Figure A.30 we can also see that this model associated a high probability ($> 80\%$) of there being an outbreak in weeks 56-57 (Jan-Feb 2001), 170-171 (Apr 2003) and in week 520 (Dec 2009). None of the notified cases within these weeks were associated with any known, notified outbreak. Of the 16 cases reported in weeks 56 and 57, six were reported within Auckland City, 5 within North Shore City, 2 within Manukau City and 1 within each of

the Franklin, Rotorua and South Waikato Districts. For the cases reported in weeks 170-171 (17 in total) the majority (7) were reported in Hamilton city. The 4 cases reported in week 520 (Dec 2009) came from the Ashburton and Timaru Districts with 3 and 1 cases respectively.

S. Typhimurium DT101

The model detected several spatio-temporal clusters of *S.* Typhimurium DT101 with high probabilities, but none were notified as outbreaks. Figure A.31 shows three weeks (weeks 32, 467 and 468) where there was a very high probability ($> 80\%$) of an outbreak having occurred. The number of notified cases for these three weeks were 10, 4, and 5 respectively. In week 32 (Aug 2000), the outbreaks were identified in South Island: 6 of the 10 notified cases were reported in Dunedin City, the other 4 cases were reported in Invercargill City (1 case), Christchurch City (1 case), and within the Selwyn District (2 cases). In weeks 467 and 468 (Dec 2008) the majority of cases were reported in upper North Island, in Manukau City (7 out of the 9 notified cases) and Auckland City.

S. Enteritidis phage type 9a

High posterior probabilities (approx 99%) of outbreaks of *S.* Enteritidis phage type 9a (Figure A.32) were identified in two weeks in April 2005 (weeks 275 and 276). For these two weeks 11 and 10 cases were notified respectively. Most of the cases for these were notified in the north of the greater Auckland region. More specifically the TA's Rodney and Waitakere. 17 (9 and 8 respectively) out of these 21 cases had been identified and notified as being part of an outbreak, with all 8 cases for week 276 observed in Waitakere City.

S. Typhimurium DT42

This model correctly identified the outbreak of *S.* Typhimurium DT42 associated with contaminated flour that occurred in Nov-Dec 2008 (weeks 462-468), by associating a posterior probability of being part of an outbreak of approx 40% to week 465 (Figure A.33). However of the 56 reported cases over this period, three were not attributed to this outbreak. These three cases were reported in weeks 462, 464 and 465 in three different TA's across the North Island (Papakura District, Ruapehu District and North Shore City respectively). The model also correctly identified (with a posterior probability of

approx 90%) an earlier outbreak of this serotype within Wellington City in May 2001 (week 71).

This analysis also identified 6 cases in May 2000 (week 19) as having a high posterior probability of being part of an outbreak (approx 99%). Of these 6 cases 3 occurred in Palmerston North City, 2 in the Rangitikei District and the final case was reported in Dunedin City. None of these 6 reported cases were identified as being part of an outbreak of this serotype.

S. Saint Paul

The posterior probability of an outbreak for *S.* Saint Paul (Figure A.34) shows three weeks (weeks 290 to 293, Jul-Aug 2005) as having a high probability of there being an outbreak. For these three weeks 5, 9 and 5 cases were observed respectively. Most of the cases for the latter two weeks were seen in the north of the greater Auckland region. More specifically the TA's Auckland City, North Shore, Rodney and Waitakere, where 12 (2, 7 and 3 respectively) out of these 19 cases had been identified as being part of an outbreak.

5.2 Identifying Possible Risk Factors

This section will highlight the results from some of the ZIP models described in Section 4. Specifically results from models 24 and 25 in Table A.6 will be discussed.

5.2.1 Age

Initial investigation showed a strong association between age and notification rates. Specifically cases aged under 5 years old were over-represented compared to cases which were 5 years or older. Table 5.4 shows the notification rate (per 100,000 person years at risk) in each age group for each of the ten most prevalent serotypes. From this table we can see that *S.* Typhimurium DT160 was the most prevalent serotype within each age group and that *S.* Typhimurium DT1 was the second most prevalent serotype for all but three of the age groups (these were the 20-29, 40-49 and 60+ year old groups). It is also interesting to note that for the 40-49, 50-59 and 60+ year old groups the second most prevalent serotype was *S.* Infantis.

Temporal trends in the distribution of reported cases with respect to age show that, for the majority of the ten most prevalent serotypes, the proportion of reported cases occurring in the 0-4 year old age range was much higher than that for any other age group over the ten year period (as illustrated in Figures 5.12 and 5.13 and in Figures A.35 and A.36). A noteworthy exception is *S.* Infantis, where in 2002 a higher proportion of 20-29 and 50-59 year olds were infected compared to 0-4 year olds and in 2009 a higher proportion of 15-19 year olds were infected compared to 0-4 year olds (as shown in Figure 5.12).

The distribution of reported cases of *S.* Typhimurium DT135 show a high level of fluctuation over time for each age group, as illustrated in Figure A.35. Although this fluctuation is not uncommon in all of the prevalent serotypes it is however most noticeable for this particular serotype especially within the age groups of 0-4 and 20-29 years of age.

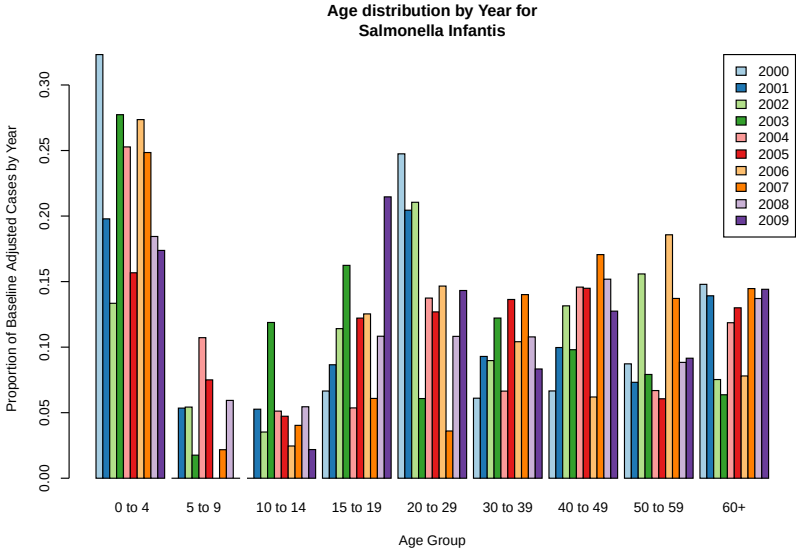

Figure 5.12: Age distribution by Year for *S.* Infantis

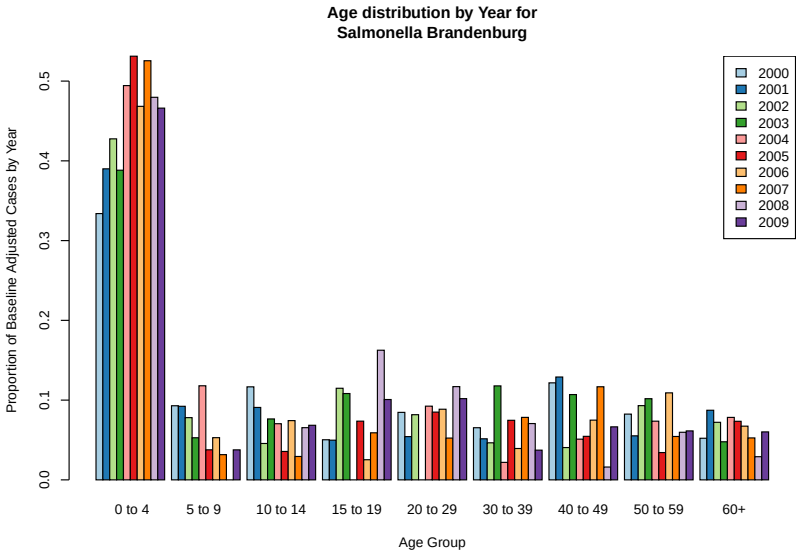

Figure 5.13: Age distribution by Year for *S.* Brandenburg

Investigation into spatial patterns with respect to age involved the production of log relative risk surfaces comparing under 5 year olds to those 5 years and older for each of the ten most common serotype. The choice to stratify age into under 5 years and 5 years and older was based on the observation that cases under 5 years old were over represented compared to those cases that were 5 years and older. Figure 5.14 show the log relative risk of *S.* Typhimurium DT160 for under 5 year olds to those 5 years and older for a

selection of the ten years worth of data. Only this serotype is presented here due to it being the most prevalent over the study period. In 2004 there was a significantly higher relative risk of *S.* Typhimurium DT160 for under 5 year olds in the north-west of the South Island, as shown by the 5% significance tolerance line in the figure for this year. In 2007 a significantly higher risk of *S.* Typhimurium DT160 for under 5 year olds was observed in three different areas of New Zealand: the south of the North Island, the south-east and north-east of the South Island, as noted by the 5% significance tolerance lines on the figure for 2007 in Figure 5.14. The following year (2008) showed a significantly higher risk of *S.* Typhimurium DT160 for under 5 year olds in the upper half of the North Island. Although not displayed here, log relative risk surfaces were also computed comparing the risk for under 15 year olds to those 15 years and older by year for each of the ten most prevalent serotypes.

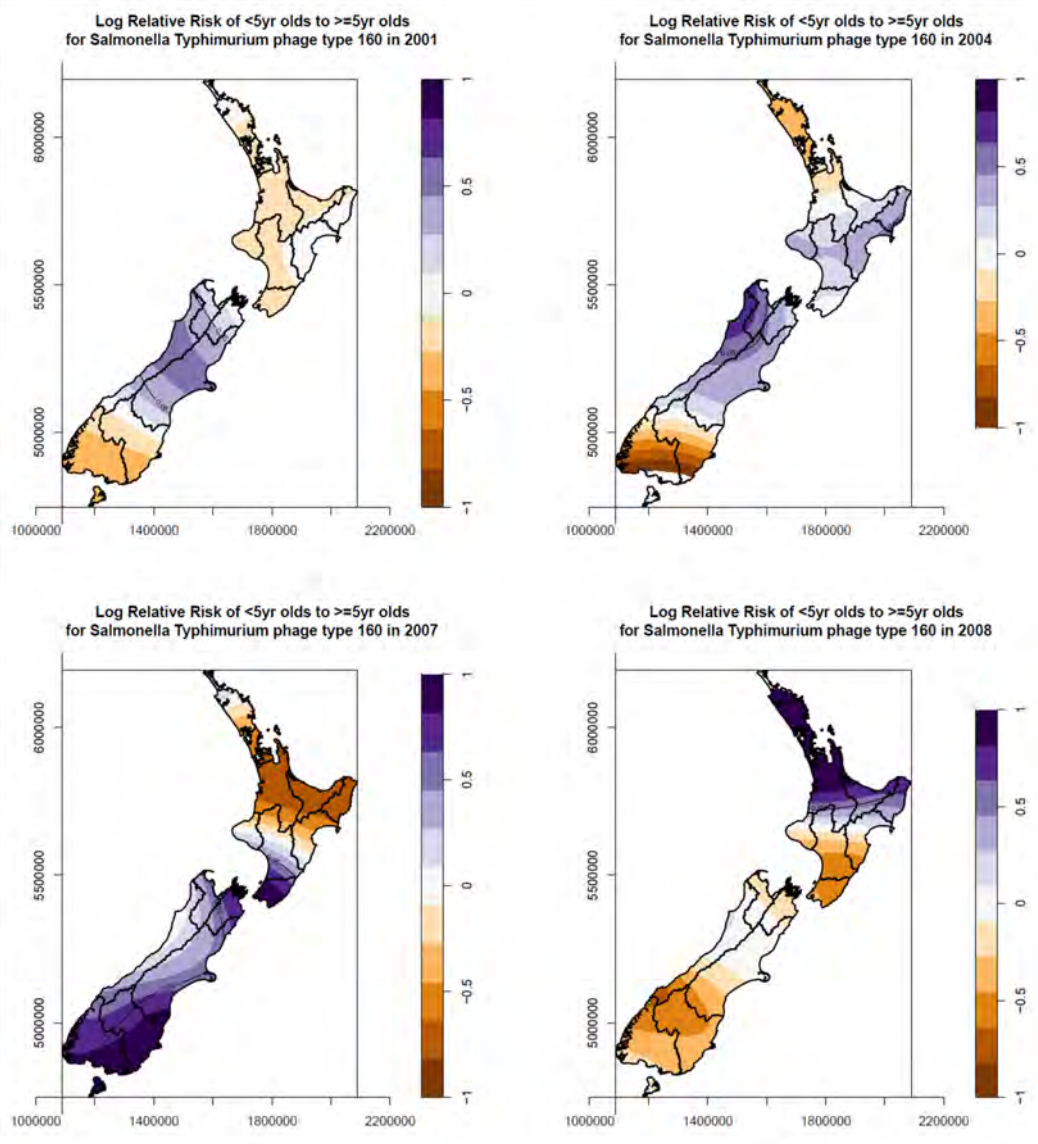

Figure 5.14: Log relative risk in 2001, 2004, 2007 and 2008 of *S.* Typhimurium DT160 for Under 5 year old to those 5 years and older at Meshblock level

5.2.2 Gender

Males account for approximately 51% of all the reported cases of *S.* over the years 2000 to 2009. Females accounted for approximately 48% of all of the reported cases of salmonellosis over the same 10 year period, the remaining 1% were not identified as male or female in the provided data. Table 5.5 gives the rate (per 100,000 person years at risk) of reported cases of *S.* over the 10 year period cross-classified by gender and serotype. This table also gives the relative rate of males to females for each serotype and at species level, with their corresponding 95% confidence interval. From Table 5.5 we can see that males had a significantly higher rate (per 100,000 person years at risk) than females at the species level and for the following serotypes; *S.* Typhimurium DT1, Brandenburg, Typhimurium DT156, Infantis and Enteritidis phage type 9a. From this table we can also see that the relative rate of males to females was the highest for *S.* Brandenburg, with a relative rate of 1.44 with a 95% confidence interval of 1.24 to 1.67.

Figure 5.15 shows the log relative risk of males to females for all reported cases of salmonellosis over the years 2000 to 2009 at meshblock level. From this figure we can see that there is a significant increase in the log relative risk for males in the lower South Island.

Figures 5.16 and 5.17 show the log relative risk of males to females for all reported cases of *S.* Brandenburg and *S.* Typhimurium DT160 over the years 2000 to 2009 at meshblock level respectively. From these figures we can see that there is no statistically significant spatial structure to the increase in the log relative risk for males relative to females for either serotypes.

Table 5.4: Notification rates per 100,000 person years at risk for the most prevalent serotypes for each age group

Aged	0 to 4	5 to 9	10 to 14	15 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60+
S. Typhimurium DT160	29.8	9.5	5.2	3.1	3.9	4.7	4.3	4.1	4.5
S. Typhimurium DT1	10.8	3.7	1.8	2.5	2.5	2.1	1.2	1.5	1.2
S. Typhimurium DT135	6.6	2.5	1.5	2.0	2.9	2.1	1.6	1.0	0.9
S. Brandenburg	8.4	1.4	1.4	1.2	1.5	1.1	1.6	1.3	1.2
S. Typhimurium DT156	9.9	3.1	1.7	1.3	1.3	0.8	0.8	0.7	0.8
S. Infantis	3.1	0.6	0.7	1.7	1.9	1.5	1.8	1.5	1.6
S. Typhimurium DT101	7.0	2.3	0.9	0.9	1.5	0.8	0.6	0.8	0.8
S. Enteritidis phage type 9a	6.4	1.8	0.7	0.6	0.8	0.9	0.9	1.2	1.2
S. Typhimurium DT42	4.2	1.5	0.6	0.6	0.6	0.7	0.4	0.3	0.4
S. Saint Paul	3.4	1.5	1.0	0.5	0.6	0.6	0.4	0.2	0.4
2006 Population Count	275079	286488	306009	300198	513417	578115	607119	486300	675222

Table 5.5: Rate of reported cases by Gender per 100,000 person years at risk and relative risk of males to females with associated 95% Confidence interval at species level and for the ten most prevalent serotypes.

	Male	Female	relative rate (95% CI)
species	37.97	33.66	1.13 (1.09, 1.16)
S. Typhimurium DT160	6.42	5.98	1.07 (0.99, 1.16)
S. Typhimurium DT1	2.70	2.16	1.25 (1.10, 1.41)
S. Typhimurium DT135	2.04	1.97	1.03 (0.90, 1.19)
S. Brandenburg	2.05	1.42	1.44 (1.24, 1.67)
S. Typhimurium DT156	1.88	1.48	1.27 (1.10, 1.48)
S. Infantis	1.39	1.23	1.72 (1.05, 1.44)
S. Typhimurium DT101	1.46	1.26	1.16 (0.98, 1.36)
S. Enteritidis phage type 9a	1.51	1.09	1.39 (1.17, 1.65)
S. Typhimurium DT42	0.79	0.79	1.00 (0.81, 1.24)
S. Saint Paul	0.78	0.71	1.11 (0.88, 1.38)

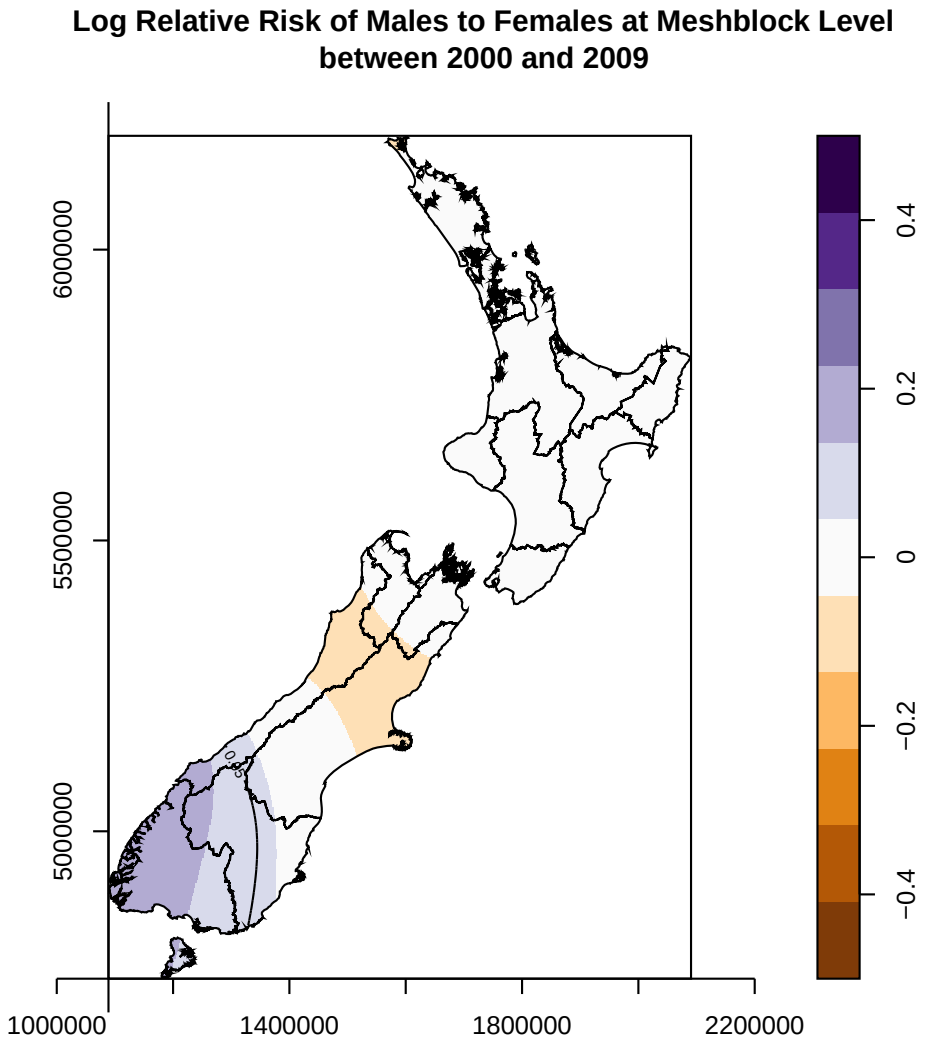

Figure 5.15: Log relative risk of males to females for all serotypes from 2000 to 2009

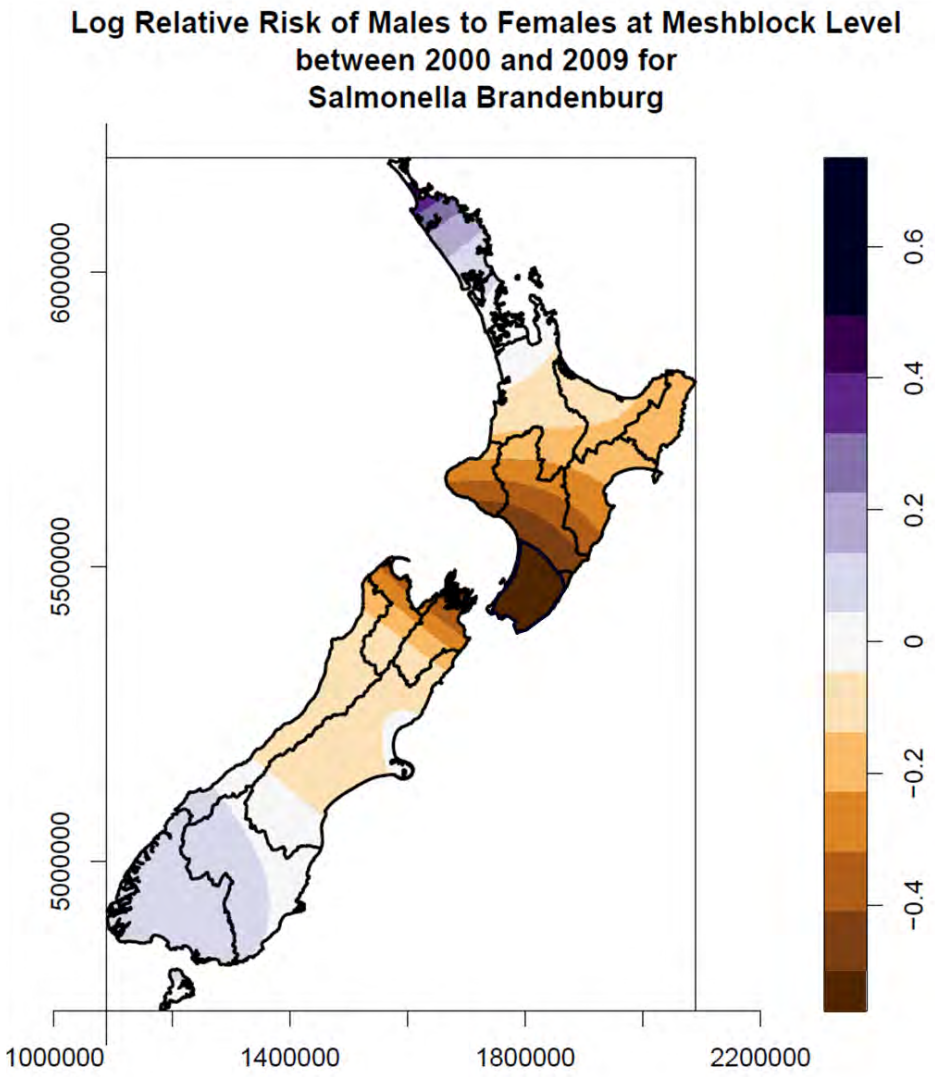

Figure 5.16: Log relative risk of males to females for *S.* Brandenburg from 2000 to 2009

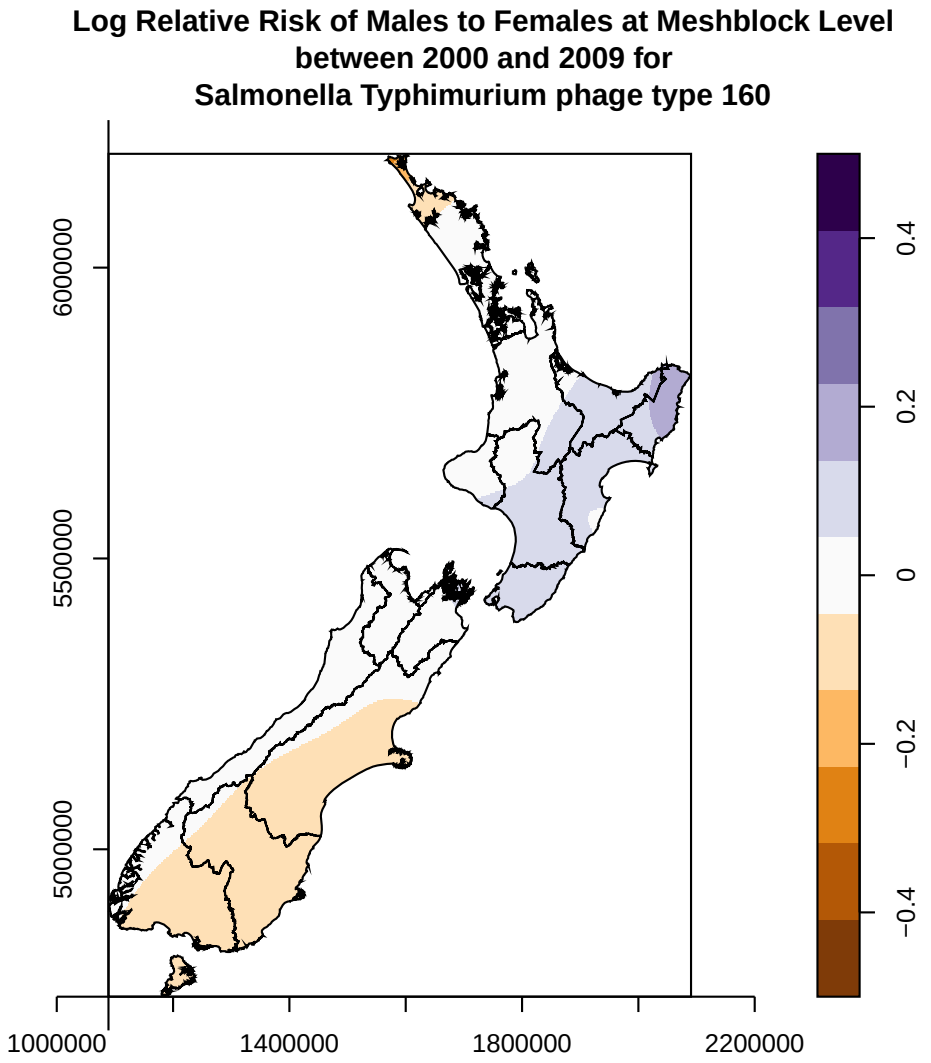

Figure 5.17: Log relative risk of males to females for *S.* Typhimurium DT160 from 2000 to 2009

5.2.3 Social Deprivation Index (SDI)

As SDI increased the expected number of reported salmonellosis cases decreased. The univariate ZIP models developed for each serotype showed that when SDI was nested within urban/rural, SDI was significantly negatively associated with the number of reported cases for all but the serotype *S.* Typhimurium DT101. This relationship was seen to be fairly constant over the serotypes, so SDI was added as a separate predictor outside the nested structure of model 24 (see Section 4 for this model).

5.2.4 Urban/Rural Classification

The rate of notification in rural areas was seen to be nearly double that in urban areas (201 and 120.4 per 100,000 person years at risk (pyar) respectively). This relationship was most noticeable for *S.* Typhimurium DT160 and *S.* Brandenburg. In highly rural remote areas *S.* Brandenburg was seen to have the highest rate of notification per 100,000 pyar (14.6) and *S.* Typhimurium DT160 had the second highest rate of notification per 100,000 pyar (8.8). This pattern is reversed for rural areas with a high urban influence where *S.* Typhimurium DT160 was seen to have a rate of notification of 13.5 per 100,000 pyar whereas *S.* Brandenburg had one of the lowest notification rates per 100,000 pyar (1.6). These figures and those for the other eight most prevalent serotypes and for all notified cases of *S.* can be seen in Table 5.6. Model 18 (Table A.6) was used to test the hypothesis that certain serotypes were associated with either urban or rural areas, having adjusted for the confounding effect of SDI. Five serotypes were significantly associated with rural areas at the 0.1% level, they were: *S.* Brandenburg, *S.* Enteritidis phage type 9a, *S.* Saint Paul, *S.* Typhimurium DT1 and *S.* Typhimurium DT160 (Table 5.7).

Table 5.6: Notification rates per 100,000 person years at risk for the most prevalent serotypes within each category in the 2006 census Urban/Rural Profile

Salmonella	Independent urban area	Main urban area	Satellite urban area	Rural area with high urban influence	Rural area with low urban influence	Rural area with moderate urban influence	Highly rural remote area
Typhimurium DT160	7.9	6.1	8.1	13.5	7.6	9.5	8.8
Typhimurium DT1	3.8	2	3.8	5.4	4.1	4.3	3.9
Typhimurium DT135	3	2	1.9	2.6	2.1	3.2	2
Brandenburg	2.4	0.9	1.3	1.6	8.1	3.2	14.6
Typhimurium DT156	1.7	1.8	1.6	2.1	2.1	2.7	1.9
Infantis	2	1.8	1.8	0.8	1.2	1.3	1.2
Typhimurium DT101	2	1.2	1.7	2	2.6	2	2.6
Enteritidis DT9a	1.8	1.2	2	2	2	1.7	1.6
Typhimurium DT42	1.2	0.7	1.2	1.1	1.2	0.7	1.7
Saint Paul	1.8	0.5	0.4	1.5	1.6	1.2	2.9
All Notified Cases	41.2	36.8	42.4	50.6	48.4	46.3	55.6
2006 Population Count	438120	2540355	112845	98754	216999	139416	69024

Table 5.7: Estimated coefficients for serotypes significantly associated with urban / rural status from ZIP model 18. Negative values indicate the serotypes are negatively associated with urban areas, and therefore positively associated with rural areas.

Serotype	Coefficient	Std Err.	P value
S. Brandenburg	-2.73	0.18	<0.00001
S. Enteritidis phage type 9a	-0.74	0.21	<0.001
S. Saint Paul	-1.15	0.24	<0.00001
S. Typhimurium DT1	-0.83	0.15	<0.00001
S. Typhimurium DT160	-0.53	0.10	<0.00001

5.2.5 Sheep, Dairy and Beef Densities

The association between ruminant densities and the number of reported cases was interpreted within the four possible combinations of North and South Islands with Urban and Rural areas, using Model 24. Only three serotypes were significantly associated (at the 0.1% level, equivalent to a threshold P value of 0.001) with at least one of the ruminant densities within at least one of the four possible combinations of North/South Island and Urban/Rural areas. These were:

- *S.* Typhimurium DT160 in rural North Island areas with sheep density Figure 5.18
- *S.* Brandenburg in rural South Island areas with dairy density Figure 5.19
- *S.* Brandenburg in rural and urban South Island areas with sheep density Figures 5.19 and 5.20
- *S.* Saint Paul in rural South Island areas with sheep density Figure 5.21

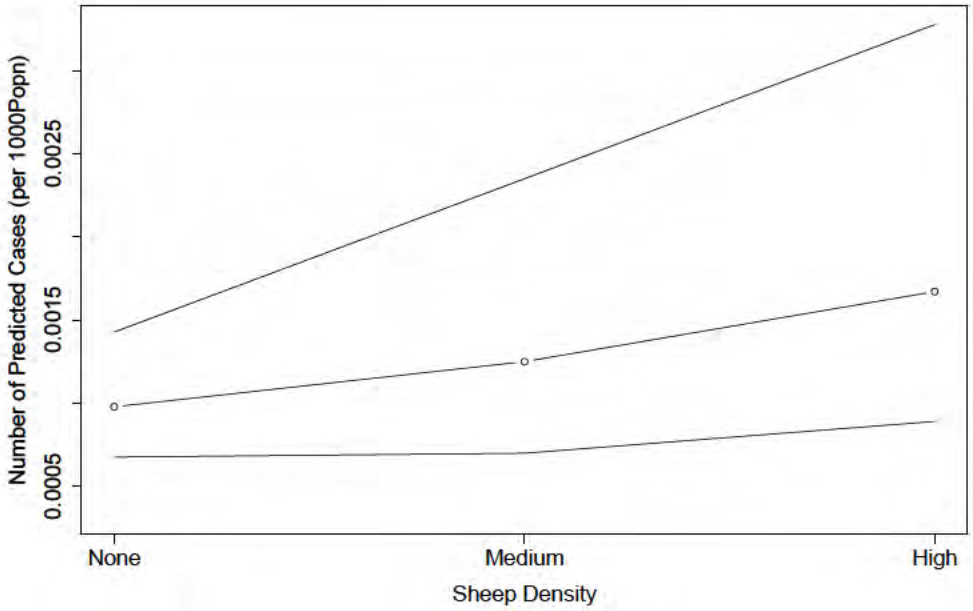

Figure 5.18: Sheep density per hectare against the predicted number of reported cases per 1000 population for *S.* Typhimurium DT160 in Rural North Island areas

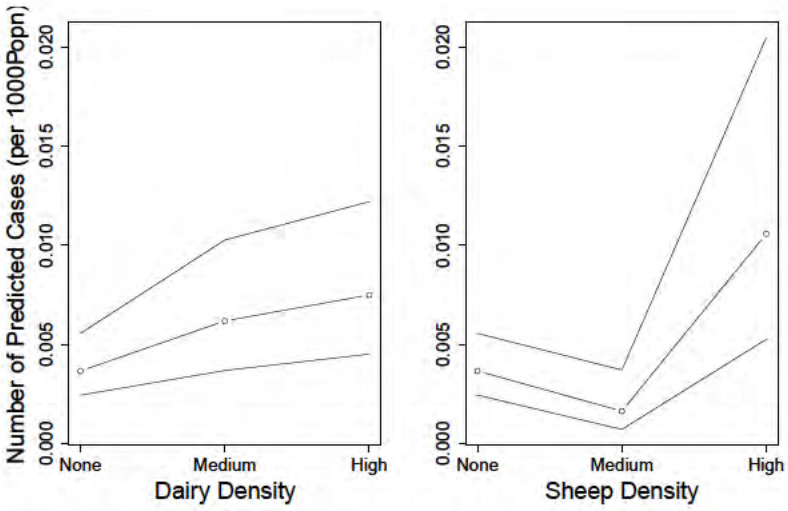

Figure 5.19: Dairy (left) and Sheep (right) Density per hectare against the predicted number of reported cases per 1000 population for *S.* Brandenburg in Rural South Island areas.

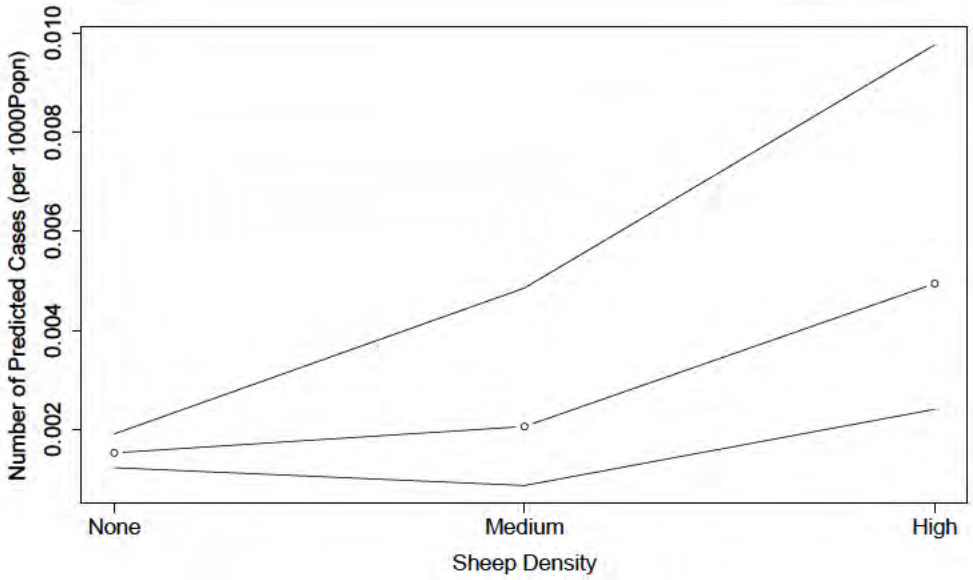

Figure 5.20: Sheep Density per hectare against the predicted number of reported cases per 1000 population for *S.* Brandenburg in Urban South Island areas.

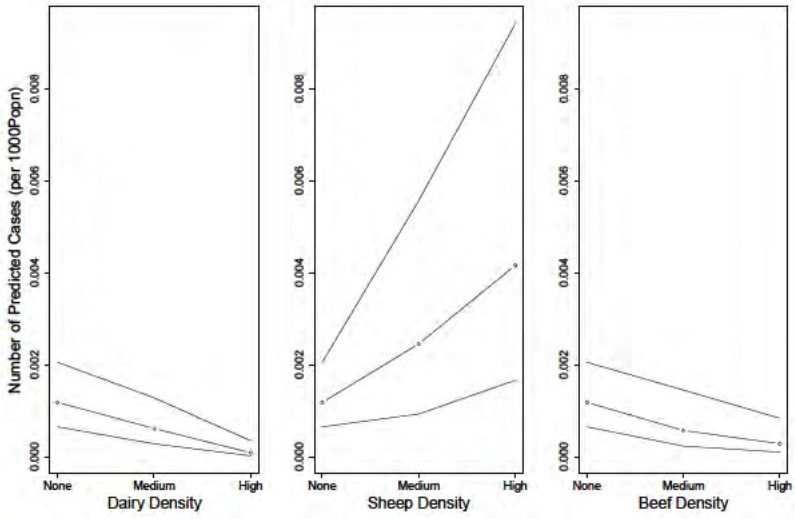

Figure 5.21: Dairy (left), Sheep (middle) and Beef (right) Density per hectare against the predicted number of reported cases per 1000 population for *S.* Saint Paul in Rural South Island areas.

5.2.6 Water Zone (Z-Code)

Whether the water zone was supplied by a surface supply was the only significant contributor to the probability that a water zone was had at least one case of salmonellosis of any serotype (the zero-inflation part of the ZINB model), where as both year and the treatment type were significantly related to the number of cases (Table 5.8). The proportion of people living in water zones supplied by roof water was small compared to ground and surface supplies as they are rarely the source of water for a community, and as a result no cases of salmonellosis were observed in these areas (N.B. this analysis did not include individual domestic house supplies). The negative value for the zero-inflation part of the model indicates that water zones supplied by surface water were more likely to have at least one case of salmonellosis than those supplied by ground water, but this is likely to be confounded by the larger size of populations supplied by surface compared to ground water. As expected the estimated coefficients for each year reflect the trend from 2000 (intercept) to 2009 showing an increase in 2001 and lower case rates from 2002 to 2009. The positive association between all treatment combinations, with the exception of UV, may reflect the need to treat based on the quality of intake water (the intercept was 'no treatment').

The residual analysis highlighted 6 regions as having particularly high counts, which are presented in Table 5.9. All with the exception of Woodend Town are supplied by surface water.

Table 5.8: Model parameters for the water zone analysis of all salmonellosis cases (all serotypes).

	Estimate	Std. Error	P-value
Count model			
Intercept	-7.87	0.064	<0.0001
2001	0.2	0.071	0.005
2002	-0.034	0.074	0.65
2003	-0.34	0.077	<0.0001
2004	-0.57	0.08	<0.0001
2005	-0.37	0.078	<0.0001
2006	-0.37	0.078	<0.0001
2007	-0.56	0.08	<0.0001
2008	-0.35	0.078	<0.0001
2009	-0.58	0.08	<0.0001
Chlorine	0.15	0.044	0.001
Chlorine and Ozone	0.4	0.14	0.004
Chlorine and UV	0.16	0.073	0.03
Ozone	12.2	297	0.97
UV	-0.49	0.13	0.0002
Zero-inflation model			
Estimate			
Intercept	-0.8607	0.1105	<0.0001
Surface water	-0.91	0.1533	<0.0001

Table 5.9: Water zones with unusually high counts of salmonellosis cases after accounting for year, treatment and source type, ordered from most to least unusual.

Code	Name	Regional Council	Cases	Popn.
OXF102OR	Oxford Rural No2	Canterbury	21	450
PAL001FW	Fitzherbert West	Manawatu-Wanganui	18	450
WEL002ON	Onslow	Wellington	111	11883
ROC001RO	Rocklands/Shannon/Pukerangi	Otago	8	40
WOO002WO	Woodend Town	Canterbury	17	1200
OAM001RR	Reservoir Road, Oamaru	Otago	18	900

5.2.7 Meteorological Variables

Model 25 was fitted to investigate any association between notification rates at the serotype level and the meteorological variables of average rainfall, temperature and absolute humidity. Measurements for these variables were aggregated to the DHB level for each month over the 10 year period. This aggregation meant that general seasonal patterns could be incorporated into

the model so that any association over and above these trends could be observed.

A significant association between the notifications rates of 7 of the 10 most prevalent serotypes with at least one of the three meteorological variables investigated was observed. These serotypes were *S.* Typhimurium DT160, Typhimurium DT1, Typhimurium DT135, Brandenburg, Typhimurium DT156, Typhimurium DT101 and Saint Paul. Table 5.10 shows the model coefficients from the count component for those meteorological variables that were significantly associated with notification rates for the 7 serotypes stated above. From this table we see that increases in notification rates for *S.* Typhimurium DT160 were associated with increases in average rainfall and average temperature. Whereas, increases in notification rates for *S.* Brandenburg were only significantly associated with increases in average rainfall. The only other serotype to be significantly positively associated with average temperature was *S.* Saint Paul. Increases in average absolute humidity were seen to be significantly associated with increases in notification rates for *S.* Typhimurium DT135, DT156 and DT101.

Table 5.10: Model coefficients for the meteorological variables that were significant at the 0.1% level from the count component of model 25, with 95% confidence intervals in brackets. The units for the variables are the average daily value in each month in each meshblock (mm, degrees Centigrade and kg/m^3 for rainfall, temperature and absolute humidity respectively) as determined by NIWA for the virtual climate network.

Serotype	Variable	Estimate (95% Confidence Interval)
Salmonella Typhimurium DT160	Average Rainfall (mm)	0.04 (0.02, 0.06)
	Average Temperature (°C)	0.19 (0.13, 0.25)
	Average Absolute Humidity %	-238.20 (-345.97, -130.44)
Salmonella Typhimurium DT1	Average Rainfall (mm)	-0.10 (-0.14, -0.06)
Salmonella Typhimurium DT135	Average Rainfall (mm)	-0.07 (-0.12, -0.03)
	Average Absolute Humidity %	404.40 (196.50, 612.23)
Salmonella Brandenburg	Average Rainfall (mm)	0.12 (0.08, 0.16)
Salmonella Typhimurium DT156	Average Absolute Humidity %	408.30 (172.25, 644.43)
Salmonella Typhimurium DT101	Average Absolute Humidity %	505.20 (275.45, 734.91)
Salmonella Saint Paul	Average Temperature (°C)	0.40 (0.25, 0.56)

Section 6

Discussion

Salmonella are ubiquitous organisms that are carried in the gastrointestinal tract of many animal species [15]. Salmonellosis can be a very severe disease especially in young children and those compromised. Of the approximately 1100 cases occurring each year (2004 to 2009) there have been 200 hospitalisations annually. Consuming food from retail premises and contact with farm animals are the two most common risk factors for notified disease (ESR Annual surveillance report 2009). In its 2010-2013 Salmonella risk management strategy, the New Zealand Food Safety Authority has set itself a goal of achieving a 30% reduction in the reported annual incidence of food-borne salmonellosis by 2013. In order to reduce the incidence of salmonellosis, we must have a thorough understanding of the epidemiology of this complex disease. There is an expanding body of valuable work recently reported on salmonellosis attribution in New Zealand including the following recent reports:

1. In a 2009 systematic review of the aetiology of salmonellosis in New Zealand Wilson and Baker conclude that contaminated food is very likely to be the cause of the majority cases, person-to-person spread a moderate cause, and direct animal contact and contaminated water relatively minor causes [30].

2. Adlam et al [1] analysed New Zealand human salmonellosis surveillance data from EpiSurv using a case-case analysis, and an analysis of serotype and outbreak data with the aim of attributing non-typhoidal

salmonellosis. While acknowledging their results do not allow accurate quantification of foodborne salmonellosis, these authors conclude that food is an important route of transmission. With regard to other transmission routes they recognize that their findings have the potential to be confounded by cases having contact with multiple risk factors, such as contact with farm animals, consumption of untreated drinking water and contact with recreational water.

3. King and lake [18] conducted a detailed analysis of 251 salmonellosis outbreaks between September 1997 and December 2006 with the aim of identifying modes of tranmission and the proportion of outbreak cases associated with food. They concluded that food borne outbreaks were the most common (approximately 40% of outbreaks), involving a wide variety of food sources, and identified infected food handlers as an important source. Although outbreak-associated cases are only a small fraction of the total number of notified cases, this analysis provided useful information to be considered alongside other studies of sporadic cases.

4. Mullner et al. (2009) [22] used multiple data sources with a Bayesian approach to estimate food source attribution to salmonellosis. They conclude that pork and poultry were the major food sources, however the model did not include routes of transmission other than food borne.

Our report uses multiple techniques to analyse salmonellosis notification data (2000 to 2009) to extend the understanding further. Specifically we examined all salmonellosis cases and the ten most prevalent serotypes and:

1. Visualised notified cases using choropleth and relative risk mapping to provide hypotheses for further investigation.

2. Conducted a detailed analysis of temporal trends, examining differences in seasonality by geographical region and livestock density.

3. Quantified potential risk factors for disease, and adjusted for the confounding effects of environmental, social and individual determinants of *Salmonella* infection.

4. Identified potential outbreaks after adjusting for underlying spatial and temporal trends.

We found much diversity between serotypes with regard to spatial and temporal trends and risk factors. Nevertheless, there is an over-arching theme across this diversity: salmonellosis notification risk is generally associated with children, living in rural, often livestock dense areas. The caregivers of these children are a target group for interventions around personal hygiene, and safe animal handling and safe food preparation and handling.

6.1 Understanding gained through examining seasonality of notifications:

Month-plots of cases stratified by rurality and livestock density have revealed interesting patterns associations with springtime (August to November). We hypothesise this is due to increased exposure associated with lambing and calving seasons and changing populations of wild birds. As the New Zealand livestock industry is pasture based, ruminant-breeding cycles are synchronised to capture the flush of feed associated with spring. Thus the increased exposure at this time may be acting through the following pathways:

1. Pregnant females in the transition period (between pre- partum and peak lactation) are under significant metabolic stress and so more likely to be shedding bacteria in their faeces.
2. Young ruminants are generally more likely to shed bacteria.
3. *Salmonella* Brandenburg causes abortion in ewes late in the third trimester.
4. The spring-time of year is associated with increased human contact (both occupational and recreational) with livestock.
5. The spring-time is a high rain-fall time enhancing faecal run-off from pasture into waterways.

The seasonal pattern in rural areas with high ruminant densities was most marked for *S.* Brandenburg, indicating that the primary factors driving infection with this serotype are environmental and the result of direct contact with infected sheep and cattle. The association with spring and lambing and calving season was observed early in the epidemic of *Salmonella* Brandenburg in Southland [9]. The seasonal pattern for the most common serotype, Typhimurium DT160, was similar to the pattern for Brandenburg, but as observed by Adlam et. al. [1], the spring peak was approximately one-month later (October compared to September), particularly in rural areas. Their observation that this may be due to increased exposure to the rise in young bird populations in late spring is supported by evidence that the early outbreaks in Canterbury coincided with a major epidemic of the same serotype in wild birds in the same region [3].

6.2 Understanding gained through risk-factor analysis:

6.2.1 Ruminant density

Three of the ten most common serotypes were seen to be significantly associated (at the 0.1% level) with at least one of the ruminant densities within one of the four possible combinations of island and rurality area. These associations with livestock density remained significant after adjustment for rurality, SDI and island within the model . Notifications of the most numerous of the serotypes, Typhimurium DT160, showed a dose-response association with sheep density in rural North Island (Figure 5.18). Brandenburg notifications showed a dose response relationship with sheep densities in rural and urban South Island meshblocks, and with dairy densities in the rural South Island. Potential exposures to animal faeces may either be direct (children handling animals) or indirect (drinking unpasteurised milk, practicing home killing).

6.2.2 Rurality

Rural living may be acting through the following exposure pathways:

- Home killed meat either home butchered or commercially home killed.
- Contaminated drinking water supply.
- Young children and adult occupational mixing with ruminants (as above livestock density).
- Contact with contaminated recreational water source.
- Access to unpasteurised milk.
- More temperature abuse of food due to long traveling distances from the supermarket.
- Use of animal effluent for home garden fertiliser with subsequent produce contamination.

Despite the high relative risk associated with rural living, it is worth noting that, of all notified cases over the study period 2581 (17%) were rural living, 11622 (77% urban) with the remaining 6% unknown. Adlam et al. [1] noted differences in the distribution of *S.* serotypes affecting urban and rural populations, but we believe they erroneously concluded that serotype DT160 was associated with urban areas. Although a greater proportion of urban cases were DT160 (18.5%) compared to the proportion of rural cases that were DT160 (16.5%) their analysis was biased by the relative frequency of other *Salmolella* serotypes and did not take into consideration the population at risk in urban and rural areas. As can be seen in Table 5.6 the rates of DT160 in rural areas were generally higher than urban areas, and this is supported by the evidence provided from modelling (Table 5.7)

6.2.3 Age

For all notifications children under 5 years of age had significantly higher notification rates than other age groups, with the majority of these notifications coming from rural populations. Young children have an immature

immune system, they are inquisitive, have poor hygiene, practice frequent hand to mouth behaviours, and they interact closely with animals, particularly young animals. In dairying areas there is a common practice of family share-milking where twice daily in the busy spring time, one parent is milking cows while the other is raising calves, often accompanied by young children.

6.2.4 SDI

As with our work on campylobacteriosis a negative association between salmonellosis notification and deprivation was found. Baker et al. (2007) [5] suggests that this relationship is likely to be due to inferior access to and use of health services and therefore lower rates of notification among lower socioeconomic groups. This inferior access and use may be due to the barrier of having to pay a fee, lack of transport or communication difficulties.

6.2.5 Water zones and the water supply

The analysis of water zones did not reveal any major insight, other than the low number of cases associated with roof supplies, and the identification of high risk zones that may warrant further investigation. *Salmonella* typhimurium was isolated from 1 out of 115 roof water supplies [25] in an earlier study.

6.3 Implications for attribution:

This work was not designed to provide definitive information on *Salmonella* source attribution. However we believe our results with regard to livestock density across multiple serotypes suggest the role of animal contact in salmonellosis may play a larger role than previously considered ([1], [30]). A more detailed examination of routes of exposure based on strain-typing of samples from human cases and potential sources (ruminant faeces, recreational water, drinking water and food) is an important area for future work. We also recommend continued enhancement of the strong links between hu-

man and animal health in New Zealand.

6.3.1 Food or environmental exposure? Summary of likely exposure pathways for the most prevalent serotypes

In this section we compare the epidemiology of the four most prevalent serotypes, using information provided by this study and other recent peer and non-peer reviewed reports. We consider *Salmonella* Brandenburg to be a good 'benchmark' serotype for comparison with other serotypes - it has a number of distinct epidemiological features that identify it as predominantly environmentally acquired, rather than food borne. These features - spatial and temporal patterns and risk factors - are used to identify the likely role of environmental and food pathways for other serotypes. We therefore consider *S.* Brandenburg first and then the other three most prevalent serotypes in order of prevalence.

***Salmonella* Brandenburg**

S. Brandenburg was first isolated in New Zealand in 1966, but it was not until 1996 that it became an important epidemic strain in sheep, cattle and humans [7, 4]. The emergence of variant strains may have contributed to the recent epidemic [8]. This serotype displays a strong spatial pattern, with a high case rates in lower South Island and it has a characteristic seasonal pattern that is most marked in rural areas with high ruminant, particularly sheep, densities. Further, this serotype has the highest rates in highly rural remote areas than any other serotype examined. A relatively high proportion of cases are observed in pre-school children, and the relative rate for males compared to females was highest for this serotype, compared to the other nine most prevalent serotypes. Although it has been most strongly associated with sheep, it has also been associated with outbreaks of disease in cattle, and disease in dogs, horses, goats and deer [7, 4]. Together, this information, combined with the results of other epidemiological studies [4], provide strong evidence that this serotype is predominantly associated with environmental

exposure to ruminant faeces (household and occupational exposures).

This serotype has been isolated from food products, and may be present in high counts [31] but the predominance of rural cases, particularly during the calving and lambing season, and the relatively low number of outbreaks associated with this serotype, suggest that food borne exposure is relatively unimportant for transmission of *S.* Brandenburg to humans.

Salmonella Typhimurium DT160

S. Typhimurium DT160 first appeared in New Zealand in 1998, causing a major epidemic that peaked in 2001 [28] and has persisted as the most common serotype into 2009/10. The epidemic in humans coincided with a similar epidemic in wild birds, causing major mortality in species such as the house sparrow (*Passer domesticus*) [3, 10]. Although there are fewer reports of large scale mortality in wild birds, *S.* Typhimurium DT160 continues to be isolated from sparrows and other passerines, and recent work has shown that genotypes identified in the year 2000 were still circulating in sparrows and other birds in 2009 [23] (these were indistinguishable by two-enzyme pulsed-field gel electrophoresis and virulotyping). This serotype has also been recovered from a wide range of domestic animal species, such as dogs, cats and horses, but is not the dominant serotype recovered from cattle and sheep. The spatial pattern of *S.* Typhimurium DT160 is more diffuse than the pattern for *S.* Brandenburg, affecting both North and South Island. Like *Salmonella* Brandenburg, *S.* Typhimurium DT160 is associated with rural areas, but not as strongly associated with highly rural remote areas (note this finding is contrary to the conclusions drawn by Adlam et al. [1]) or with livestock densities (the only significant association was with high sheep densities in North Island). The age distribution shows a similar high rate of notification in young children, and there is evidence that this association with young children has varied in space and time. Unlike *S.* Brandenburg, *S.* Typhimurium DT160 is not significantly associated with males.

The time series of *S.* Typhimurium DT160 was characterised by a number of localised outbreaks, some of them very large, [6], most notably in the early years of the epidemic. The seasonal pattern is consistent with a relatively

high exposure to infection in the spring months, but in rural areas the peak was approximatly one-month after the peak observed for *S.* Brandenburg; the serotype most strongly associated with lambing and calving. This is consistent with the observation that DT160 was associated with widespread infection in wild birds, and increased exposure could be the result of increases in avian populations in the spring and early summer associated with nesting behaviour [1, 28].

From the evidence presented above we can tentatively conclude that environmental exposure to livestock is important for the tranmission of *S.* Typhimurium DT160, but not as important as it is for *S.* Brandenburg. The frequency of outbreaks, in both humans and animals, and the epidemiology of sporadic cases, is consistent with a higher level of food borne transmission than *S.* Brandenburg. Infection in wild birds, resulting in contamination of both food and the environment (and transmission to other animal species), may be the most important factor driving the epidemiology of this serotype.

Salmonella Typhimurium DT1

This serotype also had a distinctive spatial pattern, with a high concentration of cases in the north of South Island, where one of the largest outbreaks of salmonellosis occurred in February and March 2002 in Nelson City and Tasman District TAs. It is also implicated a number of other localised outbreaks identified by our spatio-temporal modelling, some of which were identified as food borne [21], and others that were not notified as outbreaks in EpiSurv- including a large cluster of cases in the Nelson and Marlborough DHB in June 2005.

S. Typhimurium DT1 was associated with young children and males, and rates were higher in rural compared to urban areas. The relationship with rurality, and males was stronger than that observed for *S* Typhimurium DT160, but the seasonal pattern did not show any association with the calving or lambing season. *S.* Typhimurium DT1 has been isolated from a range of animals, food [31] and environmental sources, and was the most commonly recovered serotype from cattle [1] (see Appendix 1 of that report).

Given the information provided by this study and other recent reports, it is

difficult to determine what is the predominant source of either sporadic or outbreak-related cases. The association with rural areas, and the frequent isolation from cattle suggests that environmental pathways resulting from exposure to cattle faeces may be important, but this is not supported by the seasonal pattern or any association with either dairy or beef cattle densities.

Salmonella Typhimurium DT135

S. Typhimurium DT135, like *S.* Typhimurium DT160 and *S.* Brandenburg, emerged in New Zealand in the form of a major epidemic that has subsequently declined. The peak of the *S.* Typhimurium DT135 epidemic was in 1999, some two-years prior to the peak in *S.* Typhimurium DT160 [1, 24] cases, and in the first year of our analysis this serotype was the most prevalent. Our analysis therefore focusses on the post-epidemic phase, which was characterised by a rapid decline in the number of cases, in which there was a large number of localised outbreaks, especially in the early years. Many of these were notified as food and food-handler related outbreaks, in regions such as the Wairarapa where there was evidence of a marked elevation in risk. Indeed, one of the largest outbreaks of salmonellosis over the time period was a *S.* Typhimurium DT135 outbreak in the Wairarapa in 2000, and this was linked to an infected food handler [1]. However, a large number of spatio-temporal clusters identifed by our model were not notified as outbreaks, and these were located in widely different geographical regions mainly in lower North Island, Gisborne and Dunedin.

S. Typhimurium DT135 was also associated with young children, but not with males, populations resident in rural areas or livestock densities. The seasonal pattern did not show any association with calving or lambing (although Adlam et al [1] did observe a rise in the number of cases in September and October, this was not associated with rural areas). Other studies have shown this serotype to be prevalent in a wide range of animal and food sources [1].

Together, the information from our detailed epidemiological analysis of *S.* Typhimurium DT135 cases and other recent reports, is consistent with a serotype in the post-epidemic phase, that is predominantly associated with

food borne exposures. The number of sporadic cases and outbreak associated cases (both notified and non-notified), have declined markedly since 1999. The reason for this decline is not known, but is has been suggested that it may be attributable to the introduction and expansion of *S.* Typhimurium DT160 and bacteriophage ST160. Unlike many of the major causes of salmonellosis worldwide, *S.* Typhimurium DT135 is sensitive to this phage [24].

6.4 Limitations of our study:

In our analysis we did not attempt to augment the EpiSurv data provided by the Ministry of Health and ESR and did not distinguish between sporadic cases and those identified to be part of an outbreak. Given the large number of latent outbreaks, characterised by clusters of cases in space and time, we considered it was important to consider all confirmed cases in our analysis for comparative purposes and for completeness. The number of cases in our study, particularly those associated with outbreaks, therefore differs to the analysis carried out by Adlam et. al [1]. In their analysis they identified, and examined separately, a total of 1426 outbreak-associated cases, of which 1082 were confirmed, whereas in our analysis we only exmained the 928 confirmed cases in our EpiSurv datset. A large proportion of the differences between these figures may be attributed to a small number of large outbreaks, such as that reported by Callaghan et al. [6], which did not appear in our EpiSurv dataset. In most cases the putative index case will be entered into the EpiSurv database but not necessarily all related cases that appear in the separate database of outbreaks.

Our analyses were first run for all salmonella and subsequently on the ten most prevalent serotypes. The ten represent 8300 of the 15033 cases in total while other or unknown serotypes (n=376) account for the remainder of the data. This has meant that some important outbreaks associated with rarer serotypes will not have not been identified. These include the summer-time 2007-2008 outbreak of *Salmonella* Chester (n=89 cases) and the 2008 outbreak of *Salmonella* Mbandaka (n=21). In the later outbreak pulse-field gel electrophoresis linked a poultry source to human cases while excluding alfalfa sprouts.

As the ZIP models used in this investigation did not account for any spatio-temporal dependencies over and above those contained within the covariate structure, we needed to be conservative in our interpretation of the p-values associated with these ZIP models. This meant that only those regression coefficients that were highly significant (p-value less that 0.001) were interpreted. However this may not have been conservative enough for a small number of variables as a simple Bonferroni correction for the largest model (model 24, with 290 regression coefficients) would suggest using a cut off of 0.00017 (0.05/290).

6.5 Recommendations for control

Control of environmental exposures is arguably more difficult than control of food pathways. Reducing infection in animal reservoirs may be helped by measures such as vaccination, which is currently practiced on some farms for the control of *Salmonella* Brandenburg. Localised spread between sheep farms has been associated with wild birds, such as gulls [7]. Understanding the epidemiology of infection in animal populations can help in the development of control measures, such as the control of *Salmonella* Brandenburg in sheep [17].

Improvements in the identification and investigation of large outbreaks involving the most prevalent serotypes may help to reduce the impact of outbreaks and help to determine common high risk sources and behaviours for future prevention strategies [29]. The implementation of new model-based techniques may help with the rapid identification of outbreaks leading to targetted subtyping and follow-up case control studies, possibly with the use of standardised control sets.

Exposure to faecal material from wild birds appears to be an important contributor to the large epidemic of *S.* Typhimurium DT160. Wild birds, particulary house sparrows, are frequently observed in food production and retail premises with open access to the outdoors. Additional control measures, particularly during the spring months, may help to reduce the likelihood of direct contact, and the contamination of food. Similar control of wild bird populations in animal feed producing premises would also help to reduce

transmission to food-producing animals.

6.6 Acknowledgements

We acknowledge the contribution made by Ruth Pirie (ESR) in the preparation and delivery of EpiSurv data and to Drs Isabel Castro and Maurice Alley for information on the ecology of wild birds and infection in house sparrows.

Appendix A

Appendix

A.1 Tables and Figures for Data Overview Section

Table A.1: Total number of cases for the 30 most prevalent serotypes along with the total number that were attributed to a known outbreak. The last three rows summarise the remaining notifications.

Serotype	Total	Number Part of an Outbreak
S. Typhimurium phage type 160	2592	141
S. Typhimurium phage type 1	1010	123
S. Typhimurium phage type 135	844	123
S. Brandenburg	734	7
S. Typhimurium phage type 156	705	25
S. Infantis	657	19
S. Typhimurium phage type 101	570	6
S. Enteritidis phage type 9a	544	31
S. Typhimurium phage type 42	334	72
S. Saint Paul	310	23
S. Typhimurium phage type 12a	269	7
S. Typhimurium phage type 9	217	9
S. Typhimurium phage type 74	171	0
S. Virchow	167	2
S. Typhimurium phage type RDNC-May 06	161	1
S. Heidelberg	158	2
S. Typhimurium phage type 23	148	1
S. Thompson	138	16
S. Typhimurium phage type RDNC	125	0
S. Enteritidis phage type 4	118	0
S. Mbandaka	118	22
S. Mississippi	113	2
S. Typhimurium phage type RDNC-Aug 01	110	7
S. Agona	109	1
S. Montevideo	107	20
S. Typhimurium phage type untypable	105	0
S. species 4,5,12 : d : -	102	0
S. Weltevreden	100	3
S. Chester	98	87
S. Newport	78	0
Other Typhimurium	694	52
Other Non-Typhimurium	2131	26
Unknown	1219	99

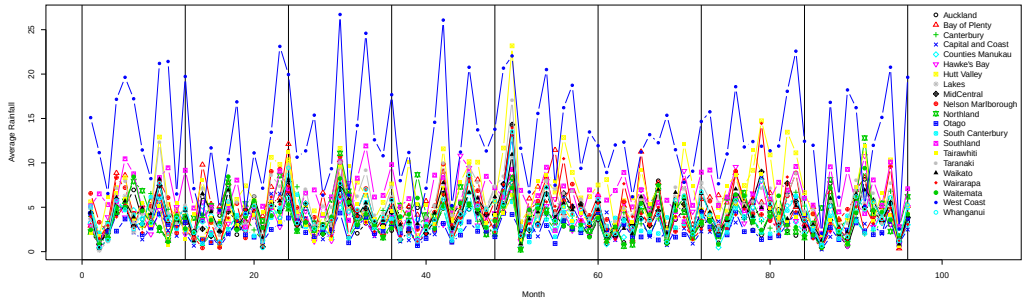

Figure A.1: Average rainfall (mm) by Month for each DHB over the 8 year period of 2000 to 2007.

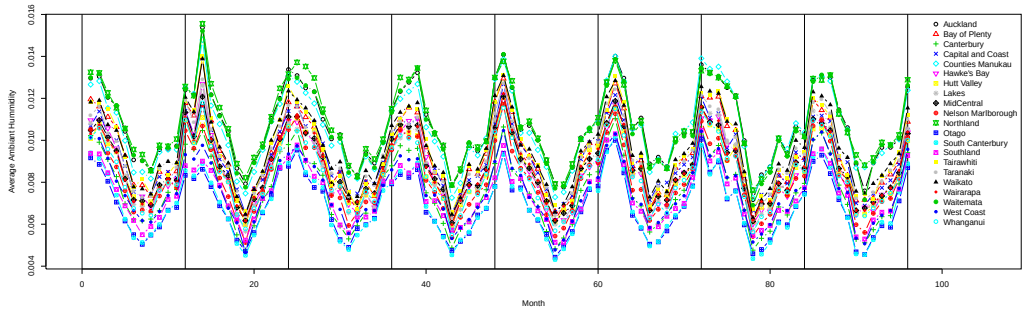

Figure A.2: Average absolute humidity kg/m^3 by Month for each DHB over the 8 year period of 2000 to 2007.

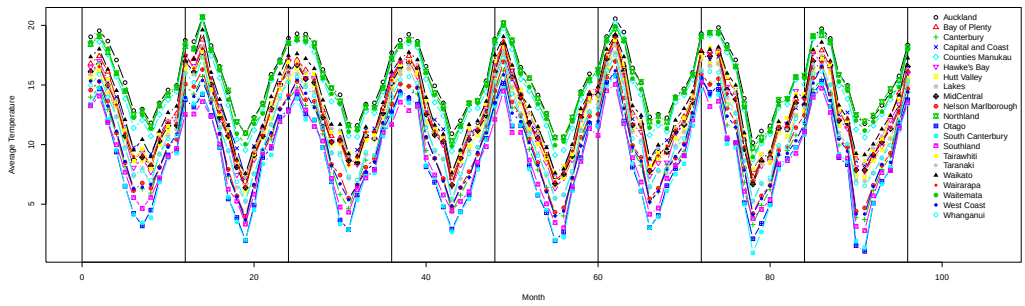

Figure A.3: Average temperature (°C) by Month for each DHB over the 8 year period of 2000 to 2007.

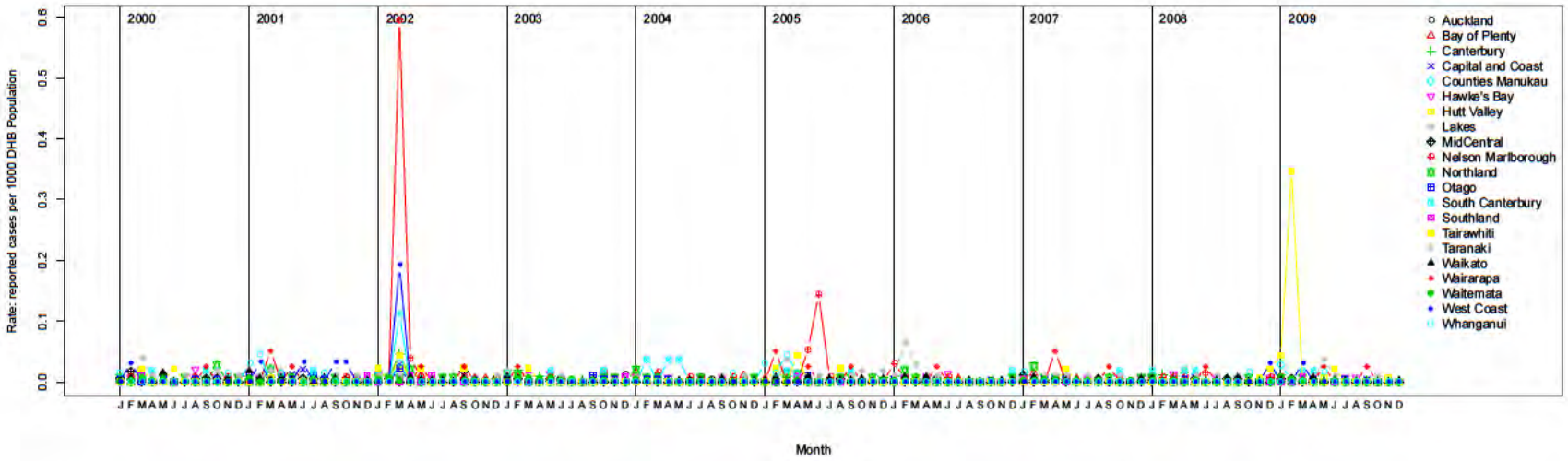

Figure A.4: Number of reported cases of *Salmonella* Typhimurium DT1 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

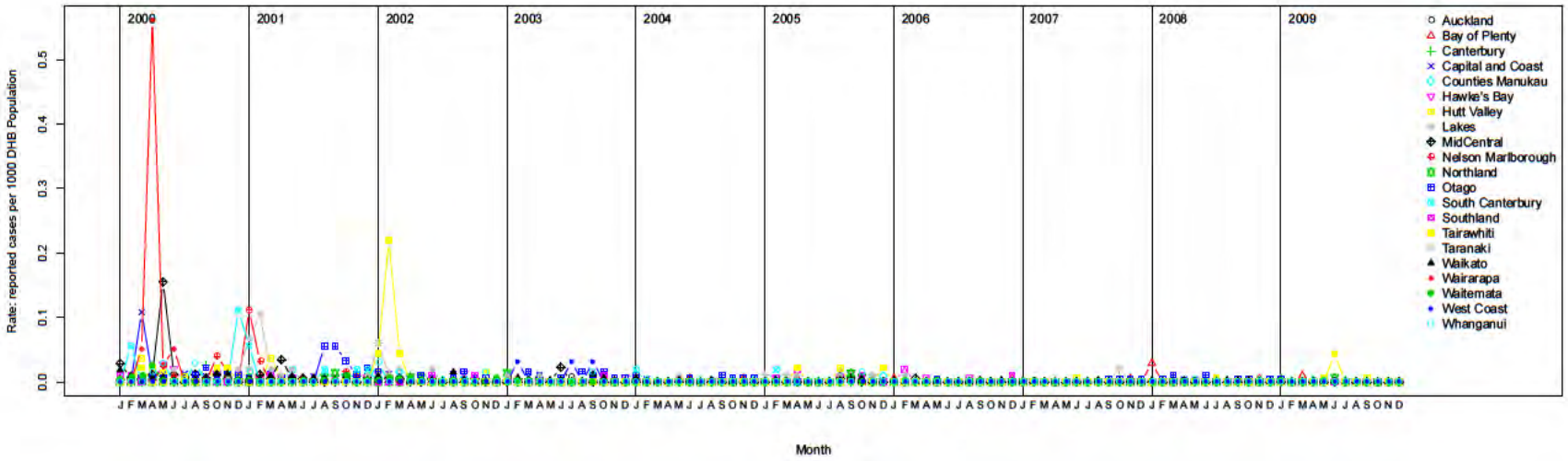

Figure A.5: Number of reported cases of *Salmonella* Typhimurium DT135 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

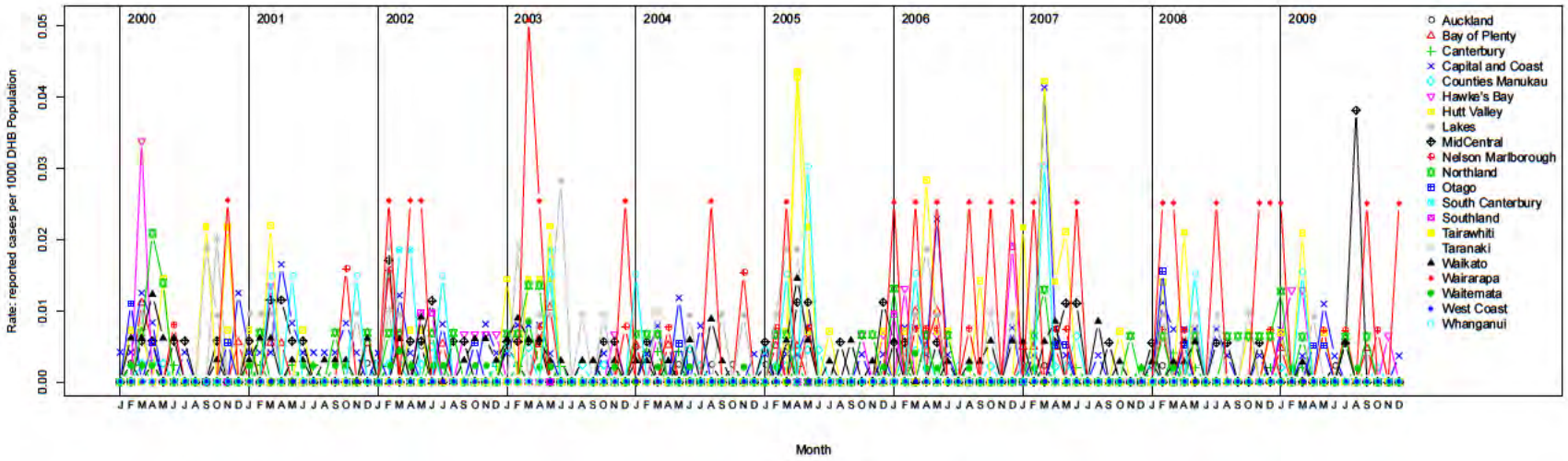

Figure A.6: Number of reported cases of *Salmonella* Typhimurium DT156 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

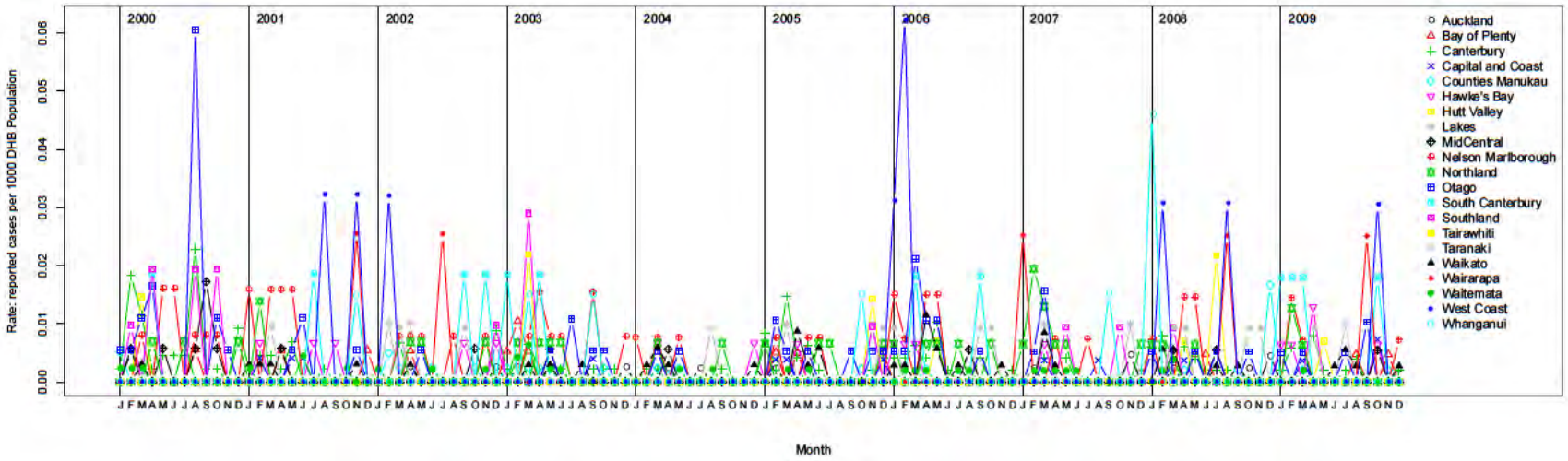

Figure A.7: Number of reported cases of *Salmonella* Typhimurium DT101 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

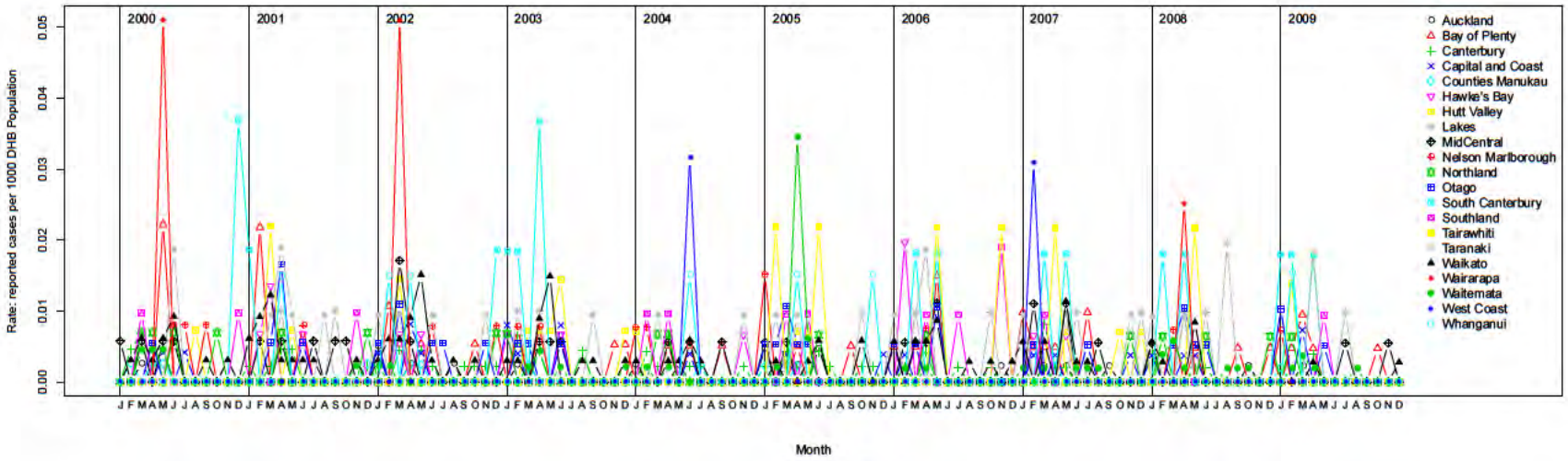

Figure A.8: Number of reported cases of *Salmonella* Enteritidis DT9a per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

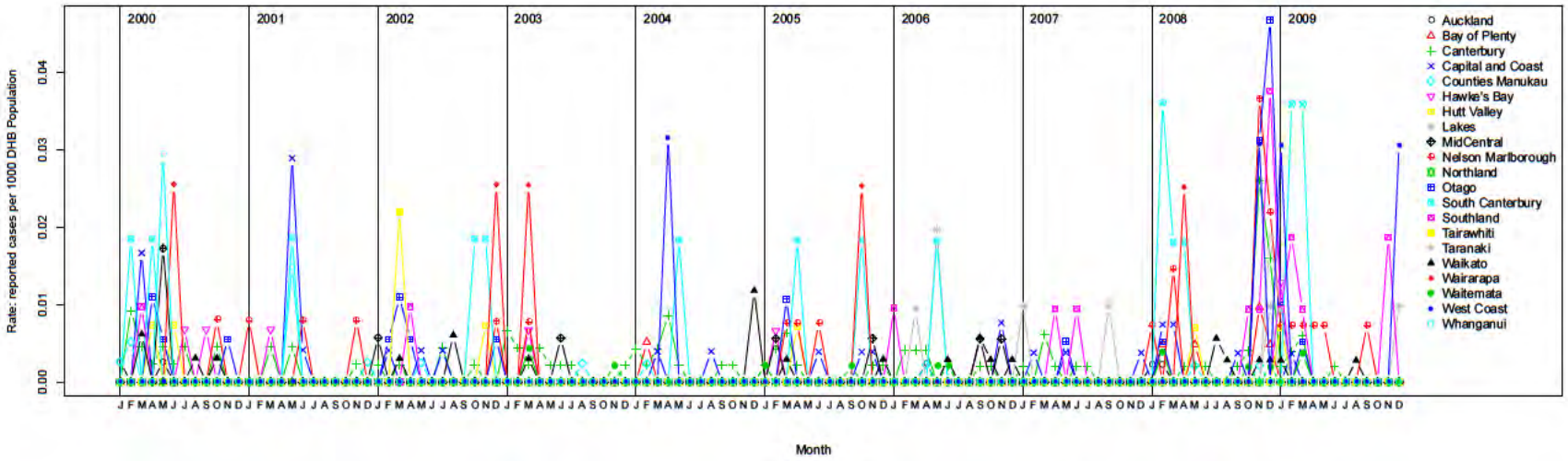

Figure A.9: Number of reported cases of *Salmonella* Typhimurium DT42 per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

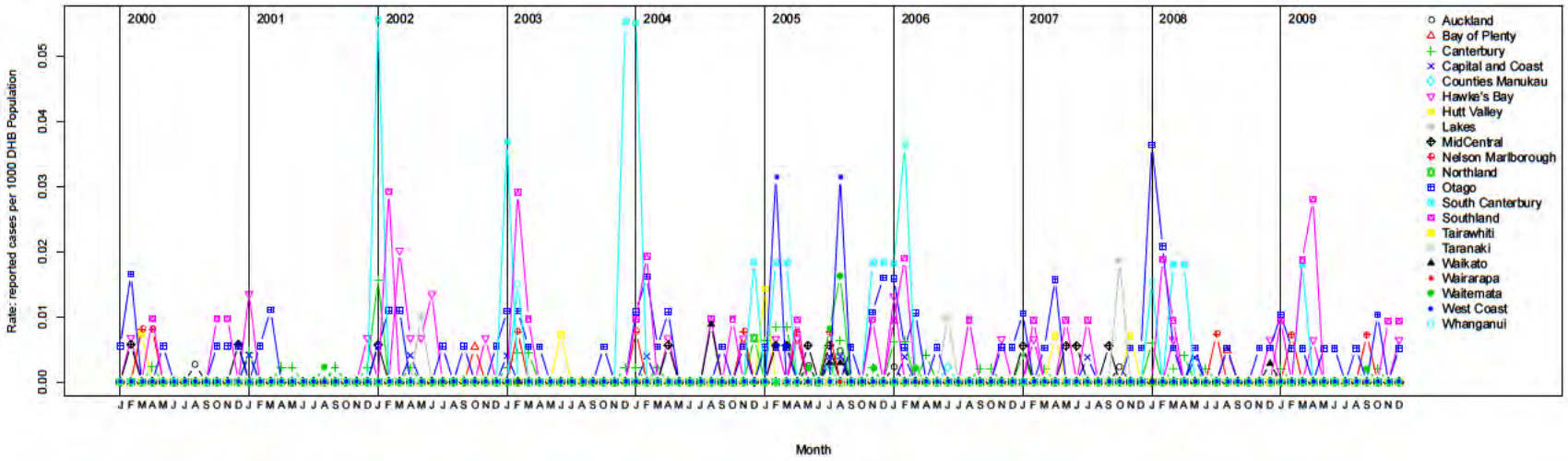

Figure A.10: Number of reported cases of *Salmonella* Saint Paul per 1000 DHB Population by Month for each DHB over the 8 year period of 2000 to 2007.

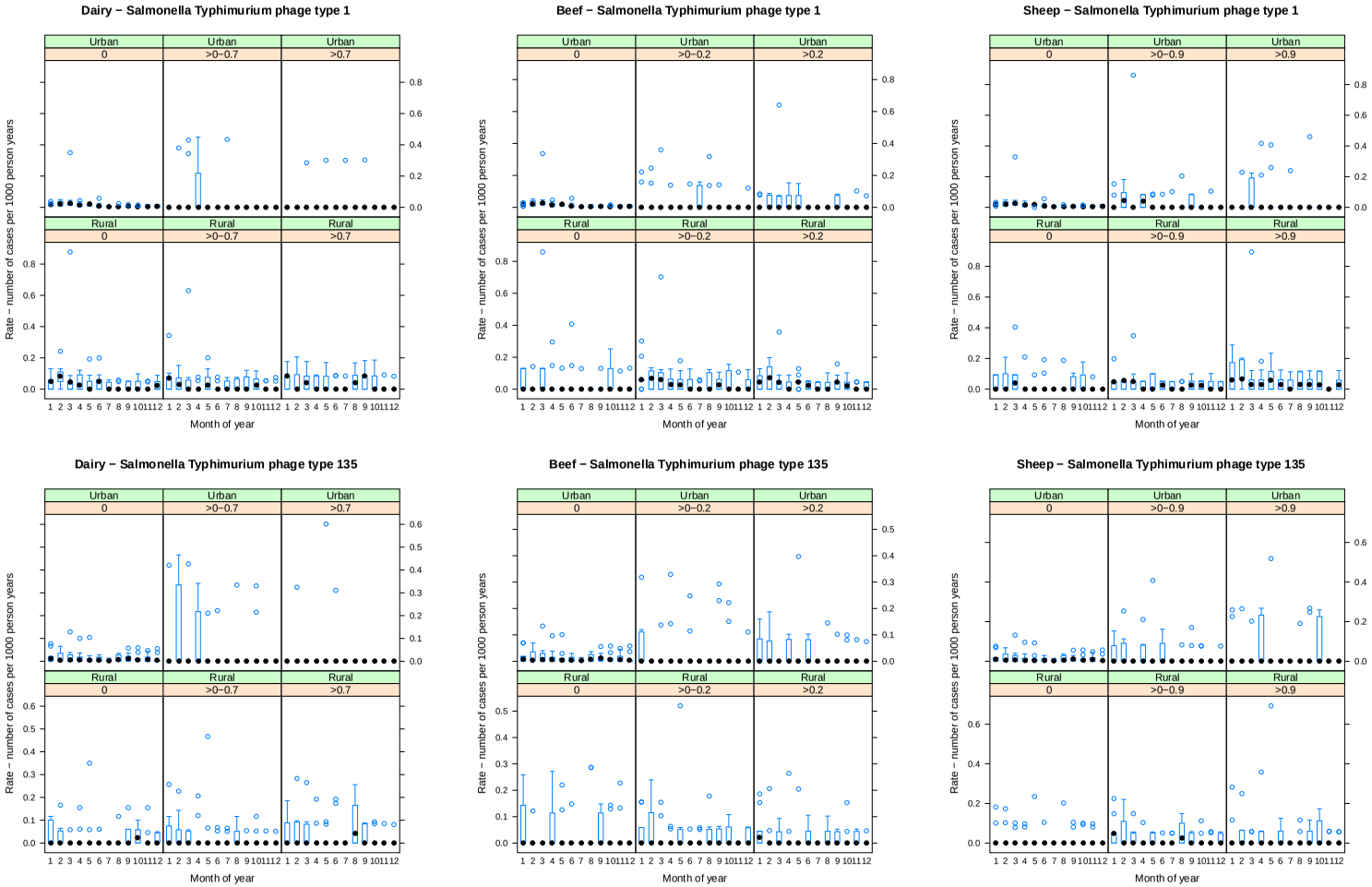

Figure A.11: Box plots showing the seasonal pattern of rates of human salmonellosis stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep herds (right) for *Salmonella* Typhimurium DT1 (top row) and *Salmonella* Typhimurium DT135 (bottom row). Each box summarises the rates for each month between 2000 and 2009.

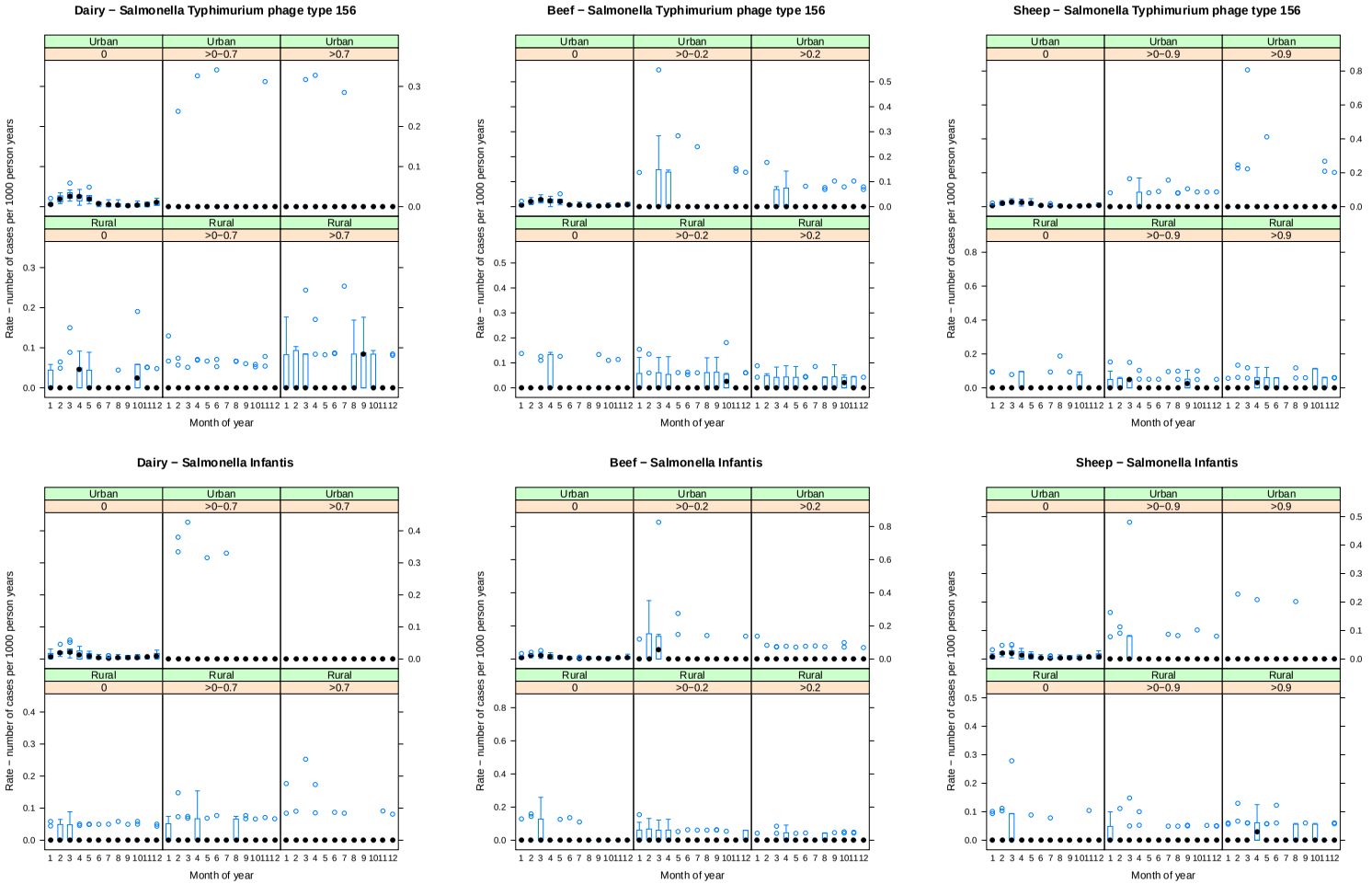

Figure A.12: Box plots showing the seasonal pattern of rates of human salmonellosis stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep herds (right) for *Salmonella* Typhimurium DT156 (top row) and *Salmonella* Infantis (bottom row). Each box summarises the rates for each month between 2000 and 2009.

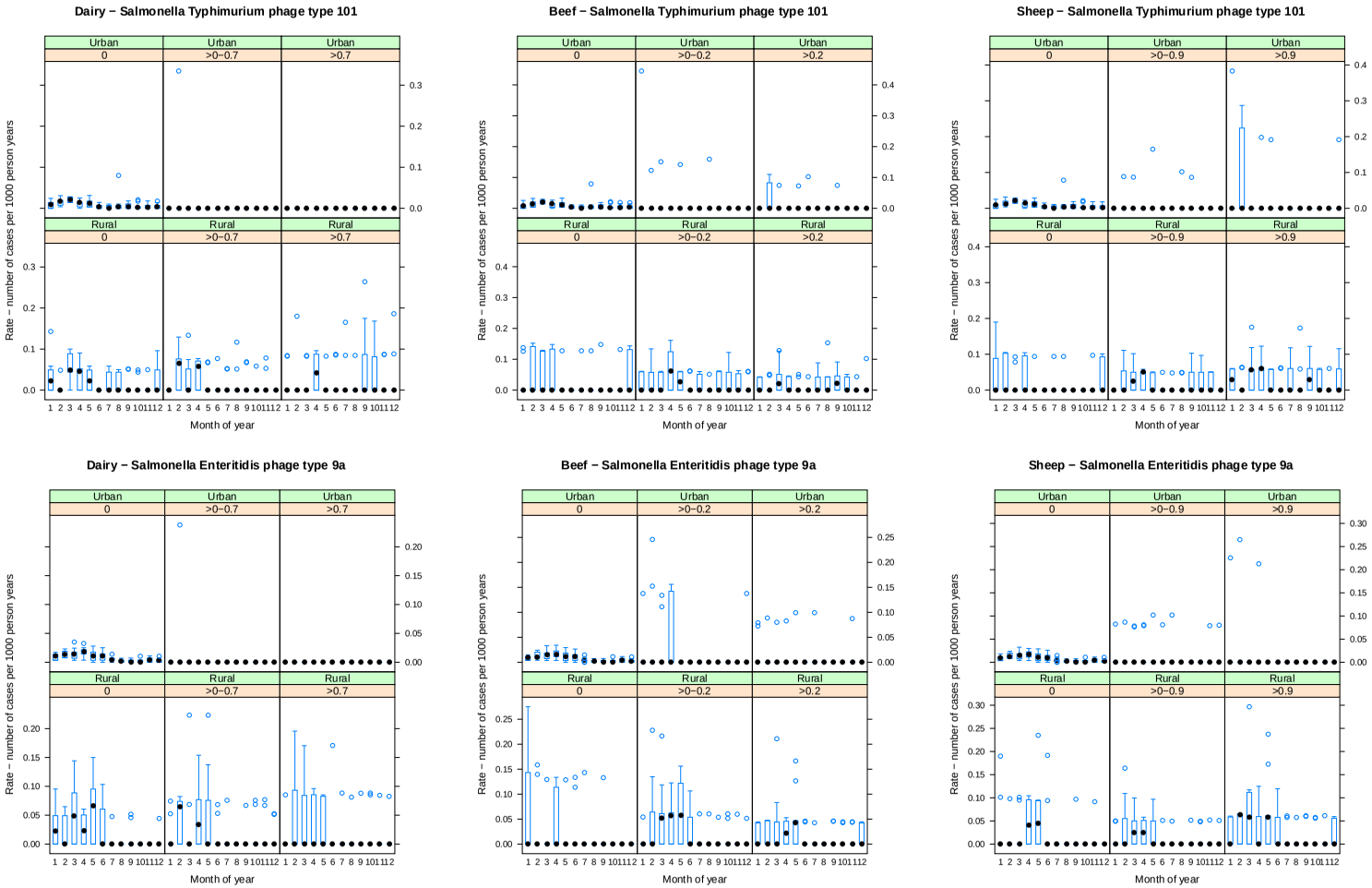

Figure A.13: Box plots showing the seasonal pattern of rates of human salmonellosis stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep herds (right) for *Salmonella* Typhimurium DT101 (top row) and *Salmonella* Enteritidis DT9a (bottom row). Each box summarises the rates for each month between 2000 and 2009.

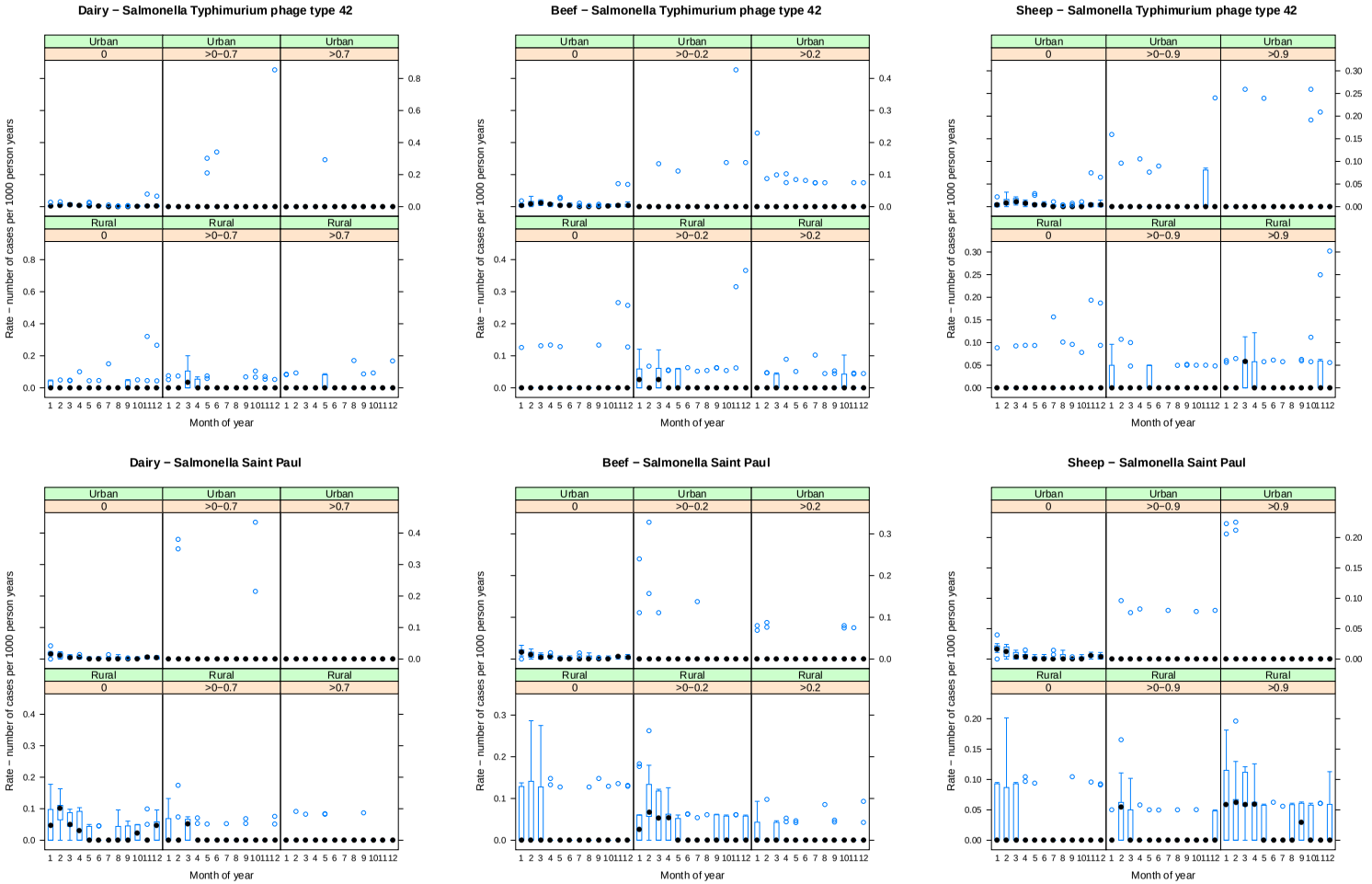

Figure A.14: Box plots showing the seasonal pattern of rates of human salmonellosis stratified by urban and rural status and the density (number per hectare) of dairy cattle (left), beef cattle (middle) and sheep herds (right) for *Salmonella* Typhimurium DT42 (top row) and *Salmonella* Saint Paul (bottom row). Each box summarises the rates for each month between 2000 and 2009.

Table A.2: Annual number of cases for the most prevalent serotypes for the years 2000 to 2009

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
S. Typhimurium phage type 160	151	570	494	286	204	249	251	151	135	101
S. Typhimurium phage type 1	105	102	212	96	62	109	73	87	72	92
S. Typhimurium phage type 135	297	189	141	64	27	53	17	10	27	19
S. Brandenburg	168	113	80	51	86	70	51	43	36	36
S. Typhimurium phage type 156	74	66	79	84	52	77	82	73	64	54
S. Infantis	24	56	85	78	57	65	59	79	84	70
S. Typhimurium phage type 101	96	47	40	58	27	62	70	42	71	57
S. Enteritidis phage type 9a	56	56	70	59	39	70	54	58	46	36
S. Typhimurium phage type 42	45	20	25	21	22	28	27	15	93	38
S. Saint Paul	19	14	34	26	31	65	35	25	35	26
Inferred Population Total	3713489	3737286	3795418	3853550	3911683	3969815	4027947	4086079	4144211	4202344

A.2 Tables for Methodology Section

Variable Label	Description
NumCases	The number of reported cases for each spatial unit (Response variable)
PopnCount	The usually resident population count for each spatial unit (Offset variable)
serotype	A Factor variable giving the serotype for each case
Month	A numeric variable (1:12) giving the Month for each year
dairydensC	An Ordinal variable categorising Dairy Density (per hectare) into None, Medium and High for each spatial unit
sheepdensC	An Ordinal variable categorising Sheep Density (per hectare) into None, Medium and High for each spatial unit
beefdensC	An Ordinal variable categorising Beef Density (per hectare) into None, Medium and High for each spatial unit
poultrydensC	An Ordinal variable categorising Poultry Density (per hectare) into None, Medium and High for each spatial unit
dairydens	A numeric variable giving the Dairy Density (per hectare)
sheepdens	A numeric variable giving the Sheep Density (per hectare)
URProfile	A categorical variable denoting the Urban/Rural Profile for each spatial unit. Based on the 2006 Census
UrbanRural	A categorical variable denoting whether each spatial unit was in a Rural or Urban area
Zcode	A categorical variable giving the Water Zone for each spatial unit
DHB	A categorical giving the District Health Board for each spatial unit
SDI	A numeric variable (1:10) giving the Social Deprivation Index for each spatial unit
SDIFactor	An ordinal variable giving the Social Deprivation Index for each spatial unit
Island	A categorical variable denoting whether each spatial unit was in the North or South Island
AveRain	A numeric variable giving the Average Rain Fall for each spatial unit over each month
AveTemp	A numeric variable giving the Average Temperature for each spatial unit over each month
AveAbHum	A numeric variable giving the Average Absolute Humidity for each spatial unit over each month

Table A.3: Variables considered in models

Model ID#	Model Formula: NumCases offset(PopnCount)+	AIC
1	DHB	63195.47
2	SDI	63603.35
3	SDIFactor	63543.09
4	URProfile	63401.6
5	Island	63532.16
6	Zcode	63575.53
7	sheepdensC	62887.82
8	dairydensC	63069.82
9	poultrydensC	62954.96
10	SDIFactor+DHB+Zcode+URProfile+Island	65096.25

Table A.4: Models fitted in the count component of the ZIP models developed at the Genus level with corresponding AIC values. Only the intercept and the offset were fitted in the zero-inflated component of these ZIP models

Model ID#	Model Formula: NumCases offset(PopnCount)+	AIC									
		ST1	ST2	ST3	ST4	ST5	ST6	ST7	ST8	ST9	ST10
11	UrbanRural/SDIFactor	3444.86	3772.91	5604.11	5824.84	6505.12	6923.51	6647.97	7998.06	9182.06	18735.69
12	DHB	3212.25	3575.04	5498.86	5608.92	6503.54	6594.02	5735.69	7817.15	8871.50	18676.49
13	UrbanRural/log(0.5+sheepdens)	3514.39	3766.29	5592.42	5807.76	6510.29	6905.79	6481.20	8005.04	9177.79	18729.23
14	UrbanRural/log(0.5+dairydens)	3473.34	3766.93	5588.45	5805.00	6513.57	6892.10	6777.24	8005.85	9178.86	18746.91
15	UrbanRural/sheepdensC	3481.46	3750.53	5558.04	5803.09	6490.34	6895.52	6517.30	7965.77	9150.86	18709.50
16	UrbanRural/dairydensC	3484.03	3767.31	5578.08	5811.34	6509.87	6895.69	6738.38	7980.50	9159.65	18724.00
17	UrbanRural/poultrydensC	3508.60	3750.90	5564.52	5800.94	6486.33	6910.12	6669.37	7984.68	9170.58	18718.95

Table A.5: Models fitted in the count component of the ZIP models developed at the Serotype level with corresponding AIC values. Only the intercept and the offset were fitted in the zero-inflated component of these ZIP models

Model ID#	Model Formula: NumCases offset(PopnCount)+	AIC
18	serotype/(UrbanRural/SDI)	74557.31
19	serotype/(UrbanRural/dairydensC)	74695.32
20	serotype/(UrbanRural/sheepdensC)	74460.43
21	serotype/(UrbanRural/(dairydensC+sheepdensC+SDIFactor))	74340.49
22	serotype/(SDIFactor + UrbanRural/(dairydensC+sheepdensC))	74364.65
23	serotype/(UrbanRural/(dairydensC+sheepdensC))+SDIFactor	74346.48
24	serotype/(Island/(UrbanRural/(dairydensC+sheepdensC+beefdensC)))+SDIFactor	73038.43
25	serotype/(DHB+AveRain+AveTemp+AveAbHum+sin(Month)+cos(Month))	26107.14

Table A.6: Models fitted in the count component of the ZIP models developed by nesting within Serotype level with corresponding AIC values. Only the intercept and the offset were fitted in the zero-inflated component of these ZIP models

A.3 Tables and Figures for Results Section

A.3.1 Temporal Analysis

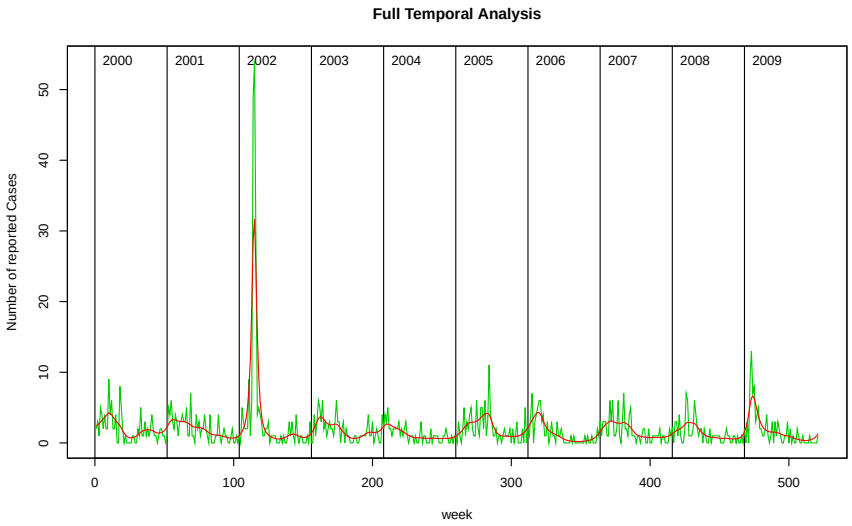

Figure A.15: Results from the full temporal analysis for *Salmonella* Typhimurium DT1

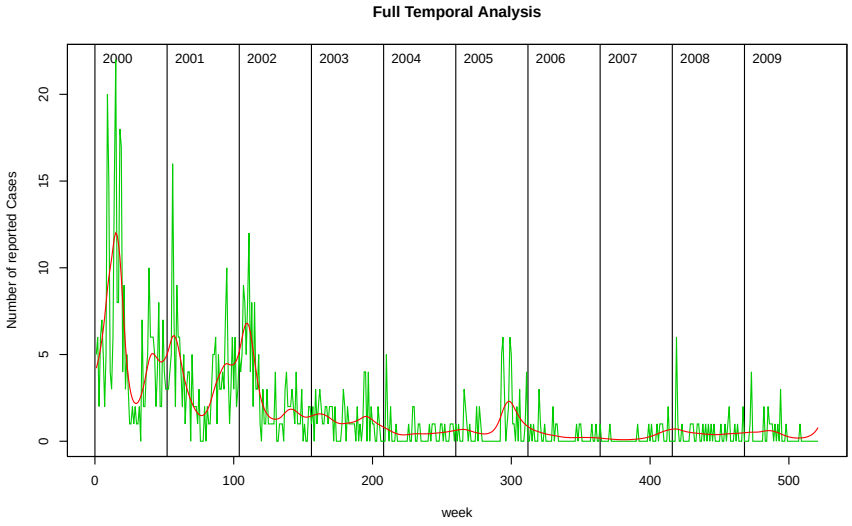

Figure A.16: Results from the full temporal analysis for *Salmonella* Typhimurium DT135

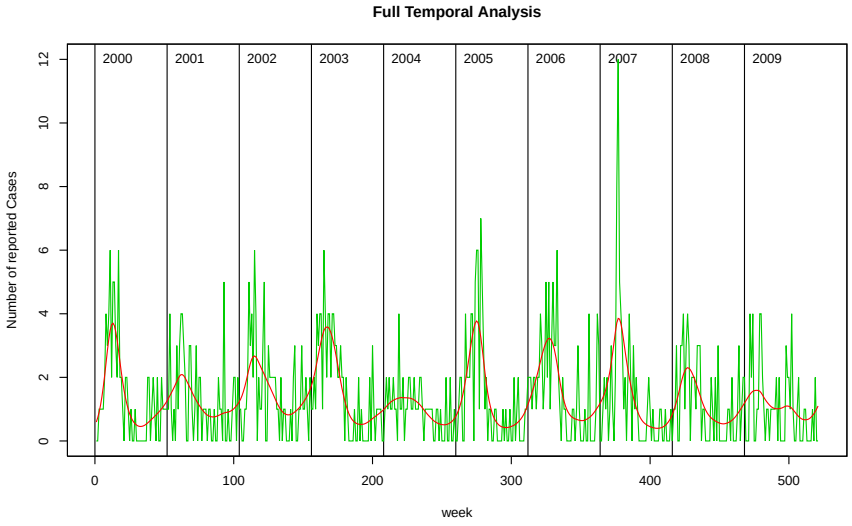

Figure A.17: Results from the full temporal analysis for *Salmonella* Typhimurium DT156

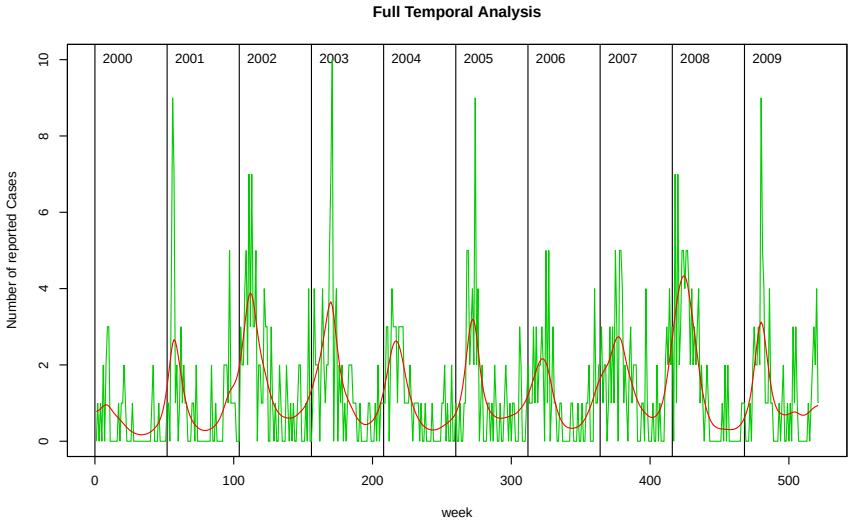

Figure A.18: Results from the full temporal analysis for *Salmonella* Infantis

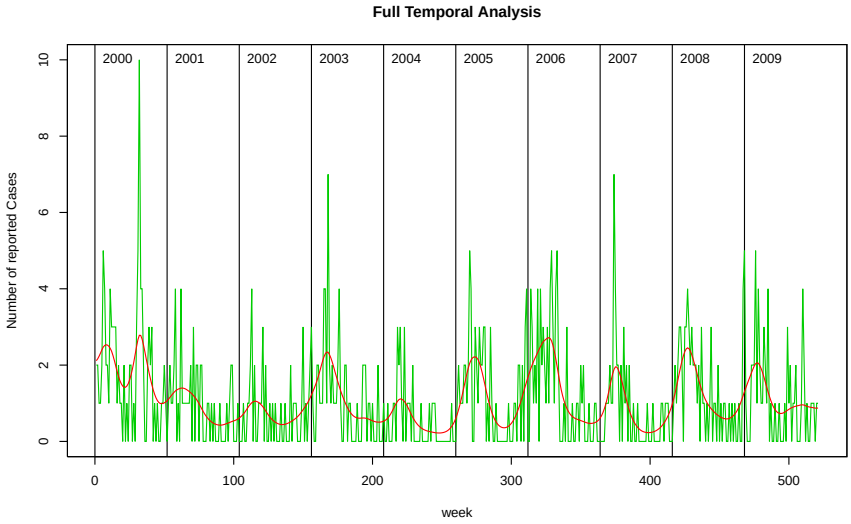

Figure A.19: Results from the full temporal analysis for *Salmonella* Typhimurium DT101

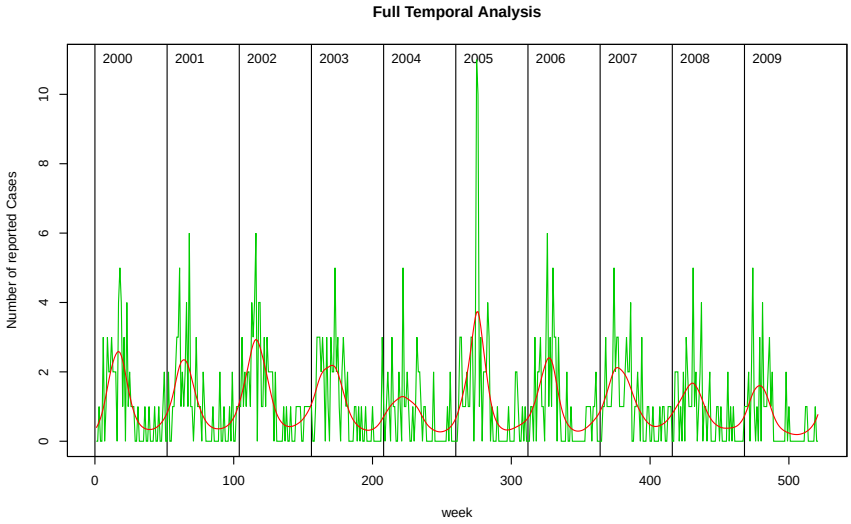

Figure A.20: Results from the full temporal analysis for *Salmonella* Enteritidis DT9a

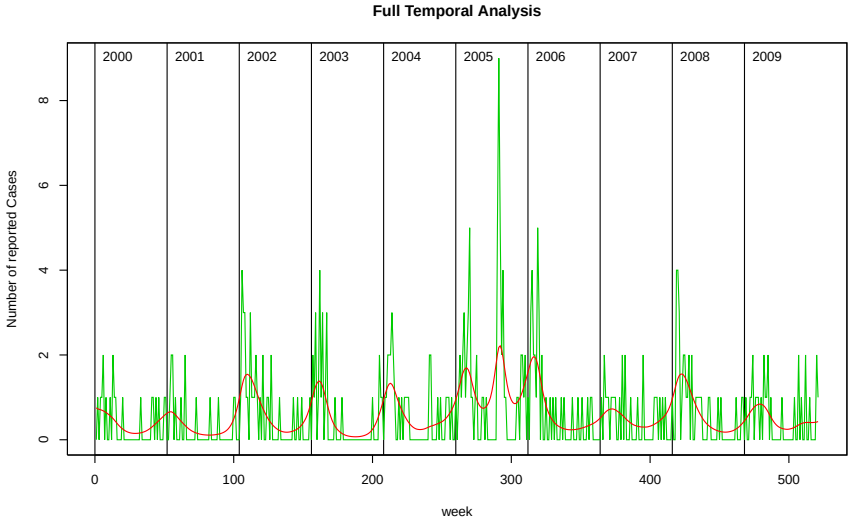

Figure A.21: Results from the full temporal analysis for *Salmonella* Saint Paul

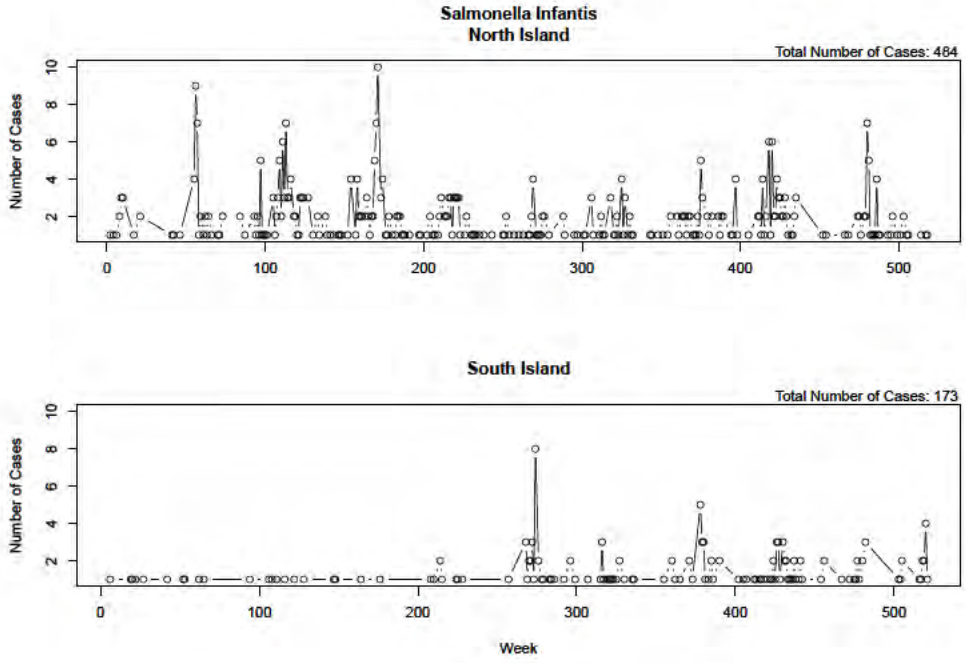

Figure A.22: Number of reported cases, over the ten year period, of *Salmonella* Infantis by Week for the North and South Islands

A.3.2 Spatial Trends

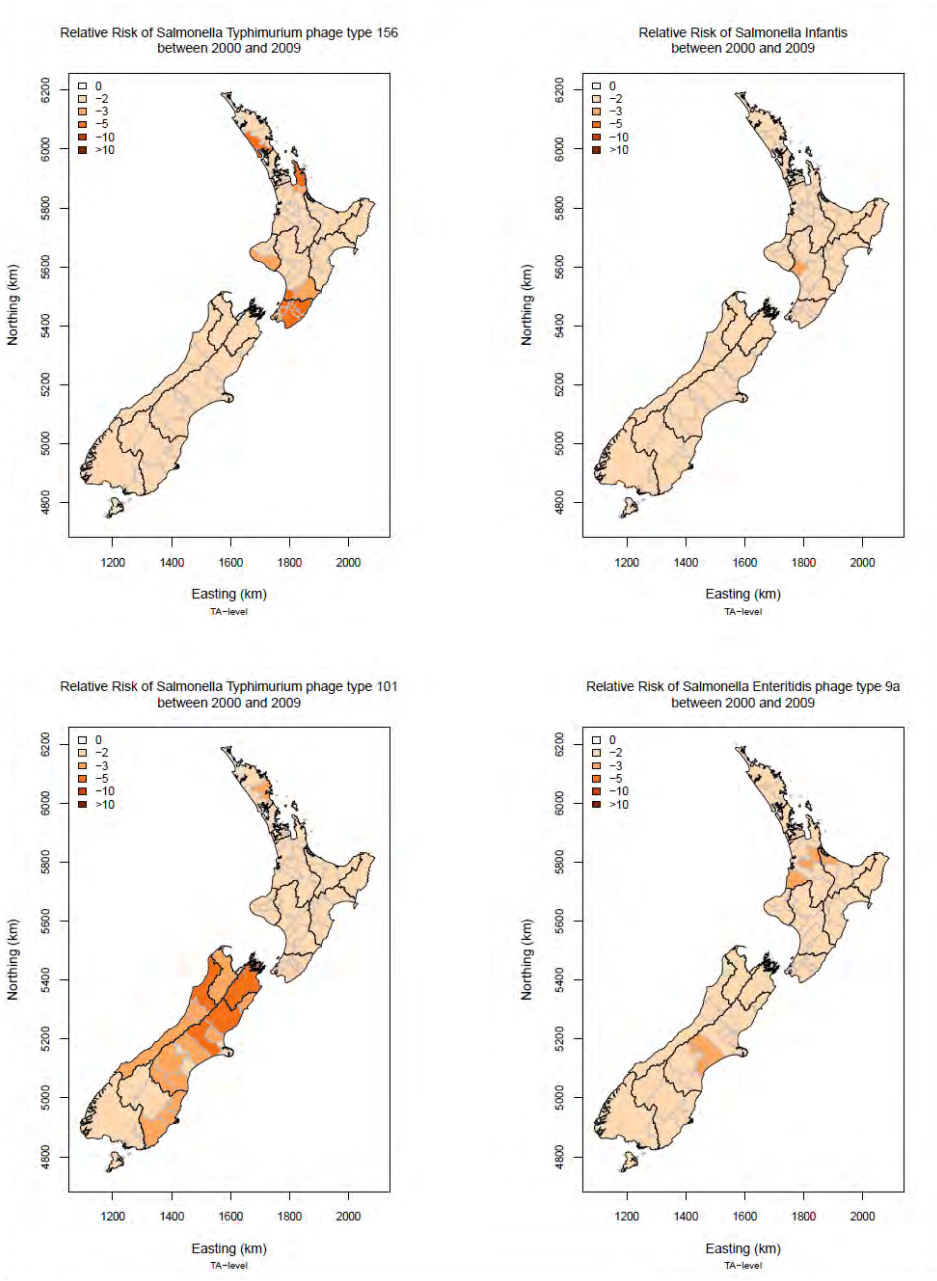

Figure A.23: Relative risk surface obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Typhimurium DT156 (top left), *Salmonella* Infantis (top right), *Salmonella* Typhimurium DT101 (bottom left) and *Salmonella* Enteritidis DT9a (bottom right).

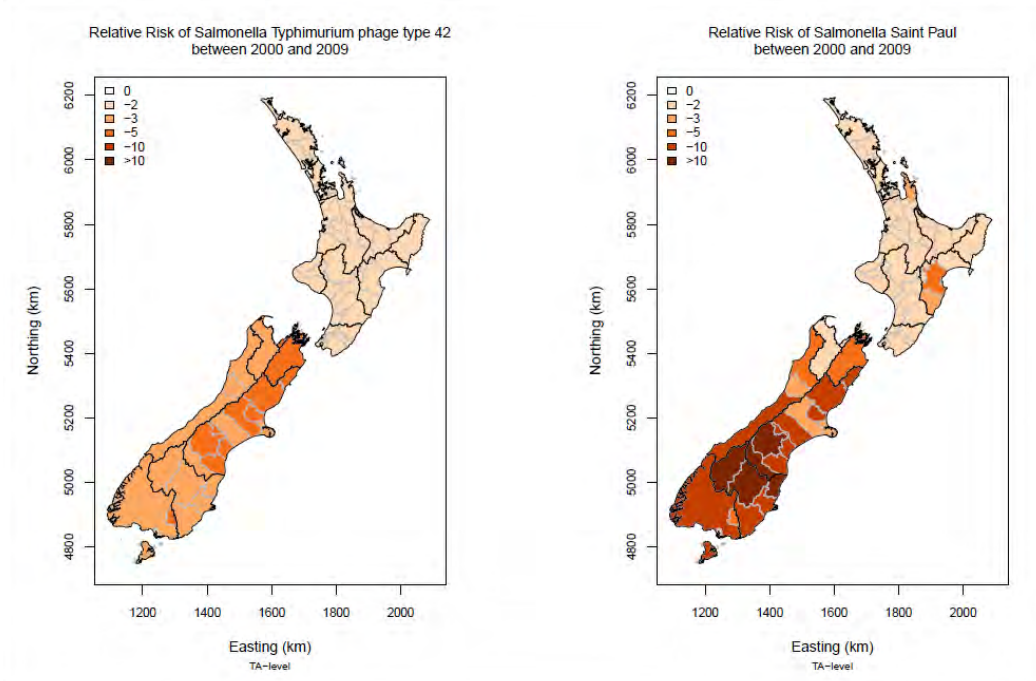

Figure A.24: Relative risk surface obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Typhimurium DT42 (left) and *Salmonella* Saint Paul (right)

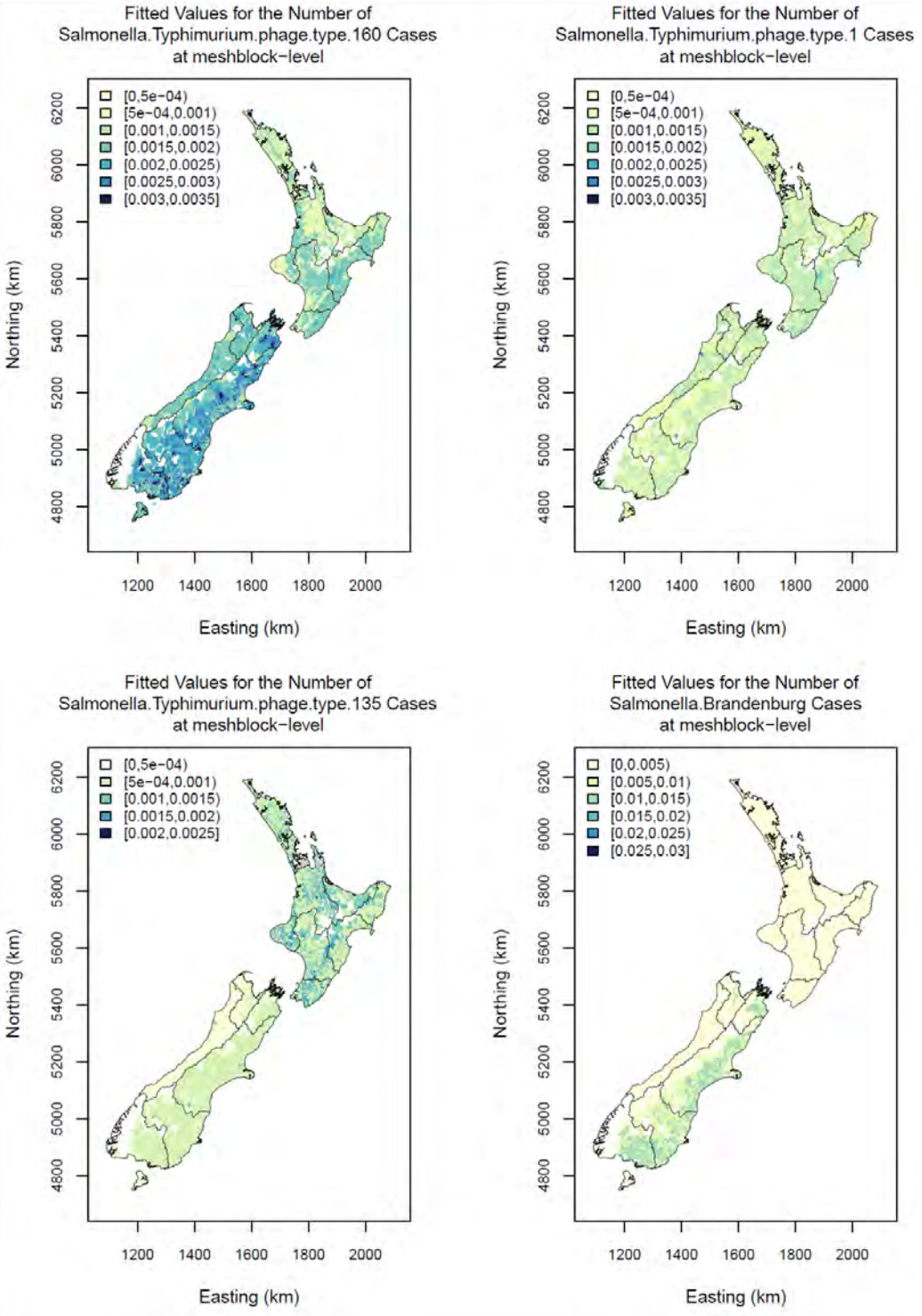

Figure A.25: Fitted values for the number of *Salmonella* Typhimurium DT160 (top left), *Salmonella* Typhimurium DT1 (top right), *Salmonella* Typhimurium DT135 (bottom left) and *Salmonella* Brandenburg (bottom right) cases at meshblock level from Model 24 described in Section 4

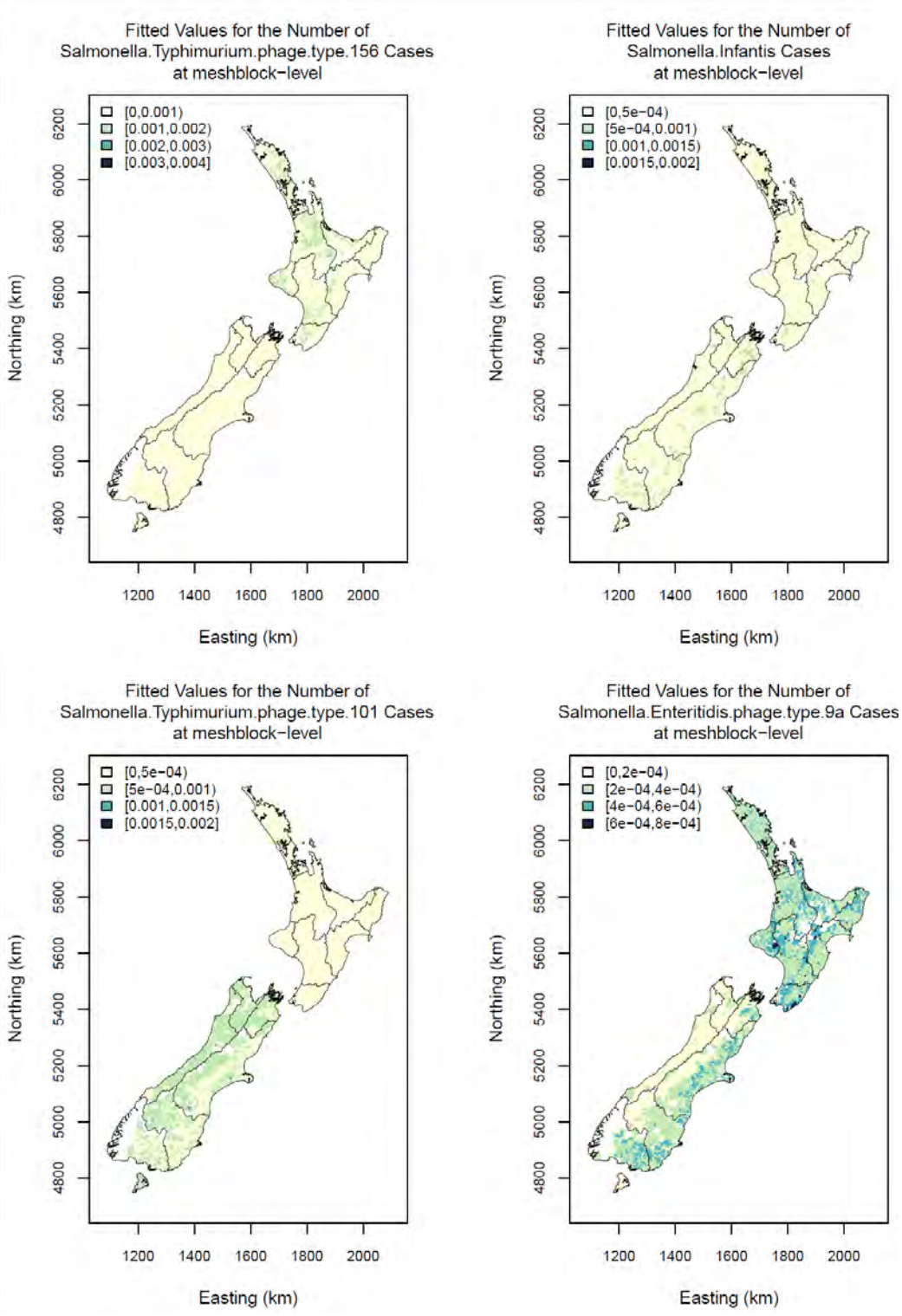

Figure A.26: Fitted values for the number of *Salmonella* Typhimurium DT156 (top left), *Salmonella* Infantis (top right), *Salmonella* Typhimurium DT101 (bottom left) and *Salmonella* Enteritidis DT9a (bottom right) cases at meshblock level from Model 24 described in Section 4

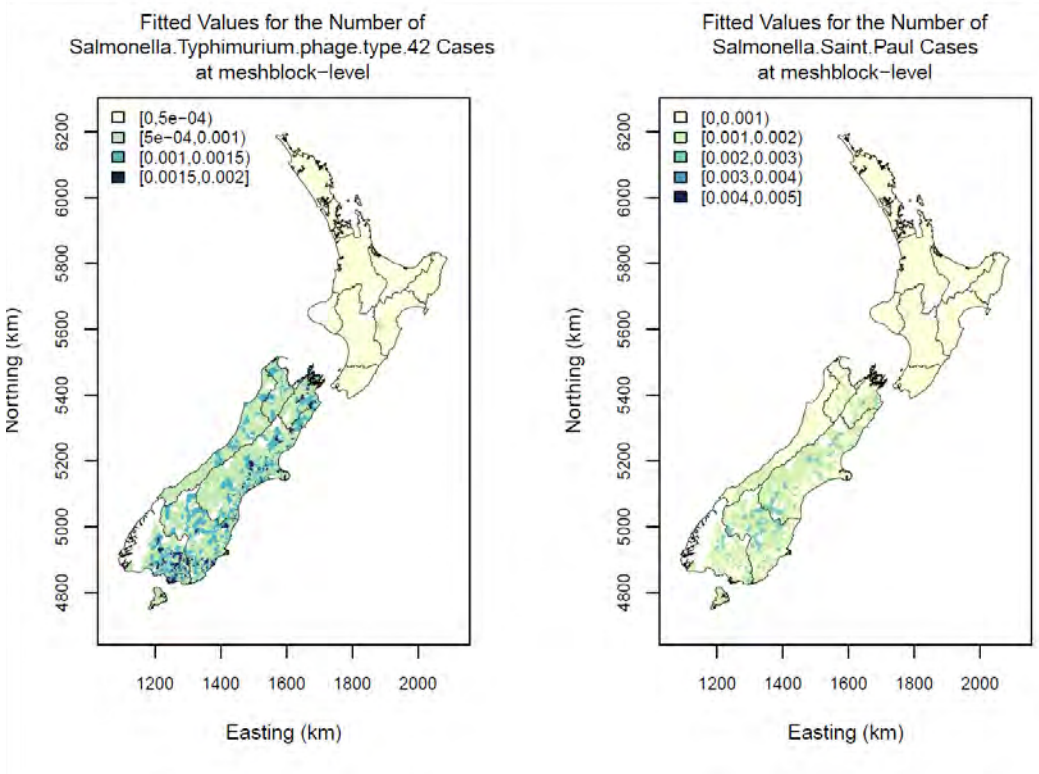

Figure A.27: Fitted values for the number of *Salmonella* Typhimurium DT42 (left) and *Salmonella* Saint Paul (right) cases at meshblock level from Model 24 described in Section 4

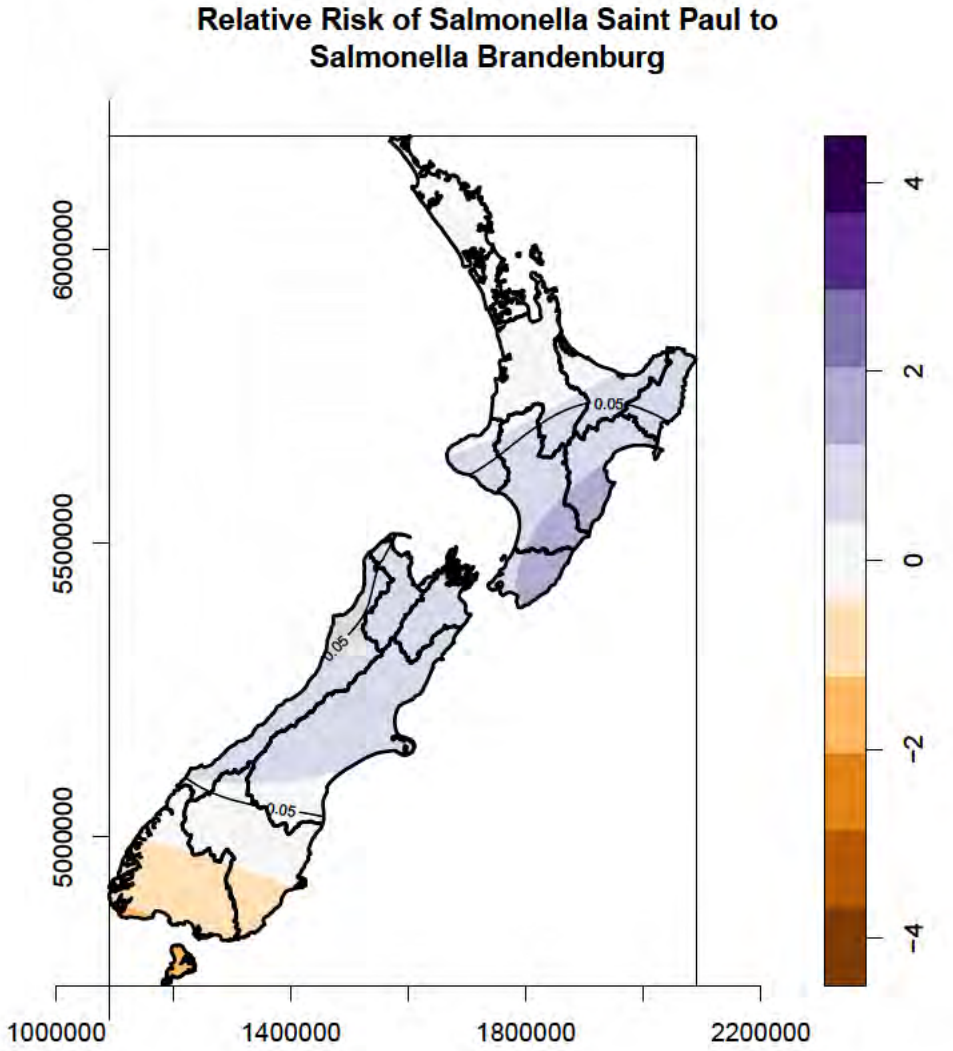

Figure A.28: Log relative risk of *Salmonella* Saint Paul to *Salmonella* Brandenburg at Meshblock level over the study period.

A.3.3 Epidemic Analysis

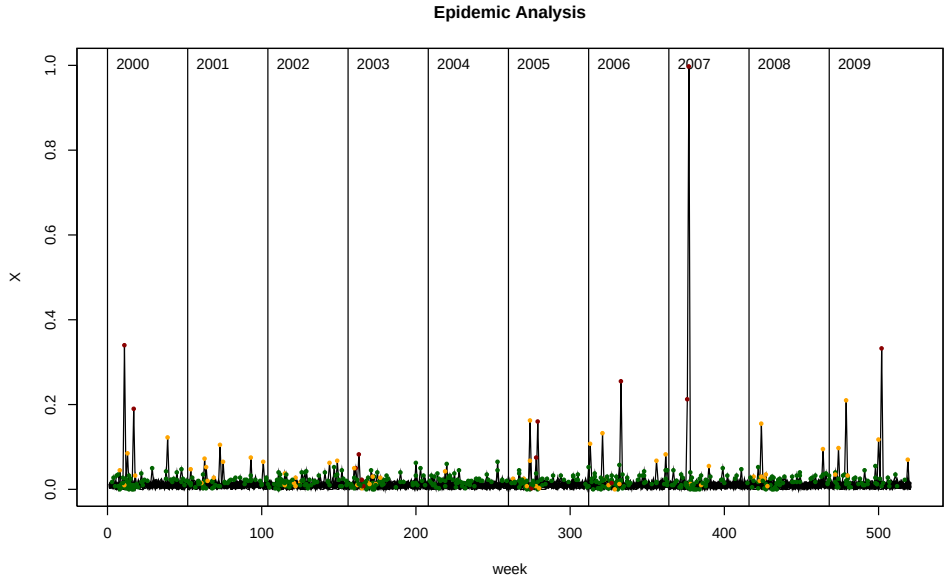

Figure A.29: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Typhimurium DT156. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

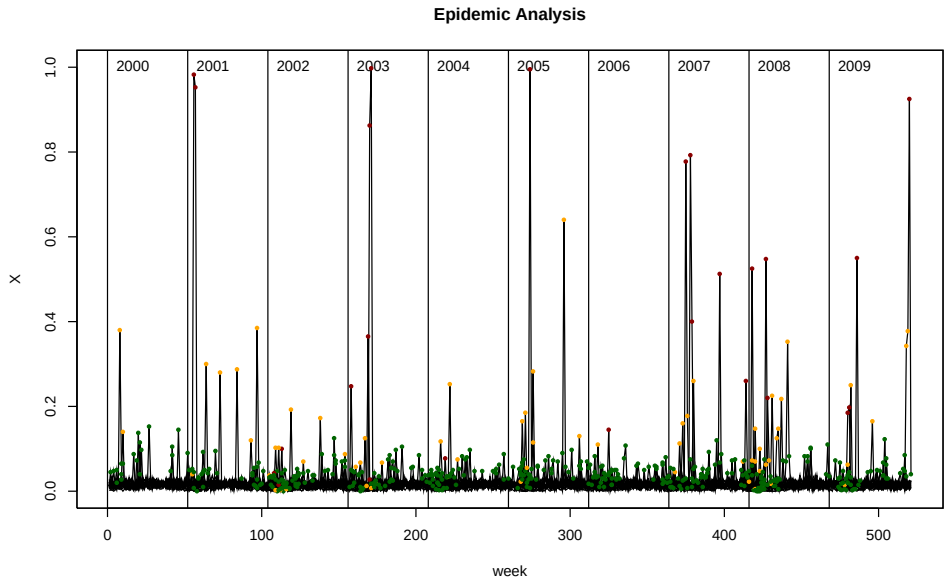

Figure A.30: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Infantis. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

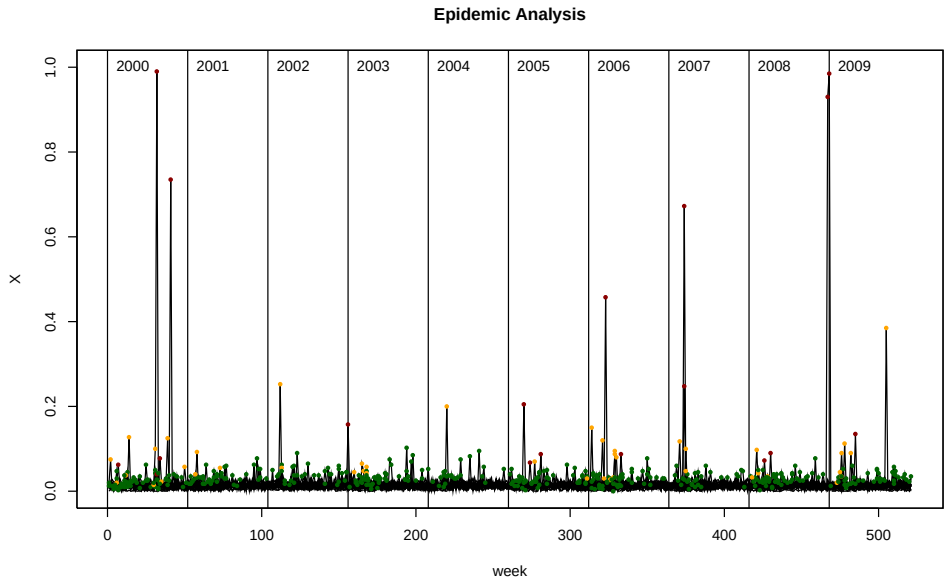

Figure A.31: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Typhimurium DT101. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

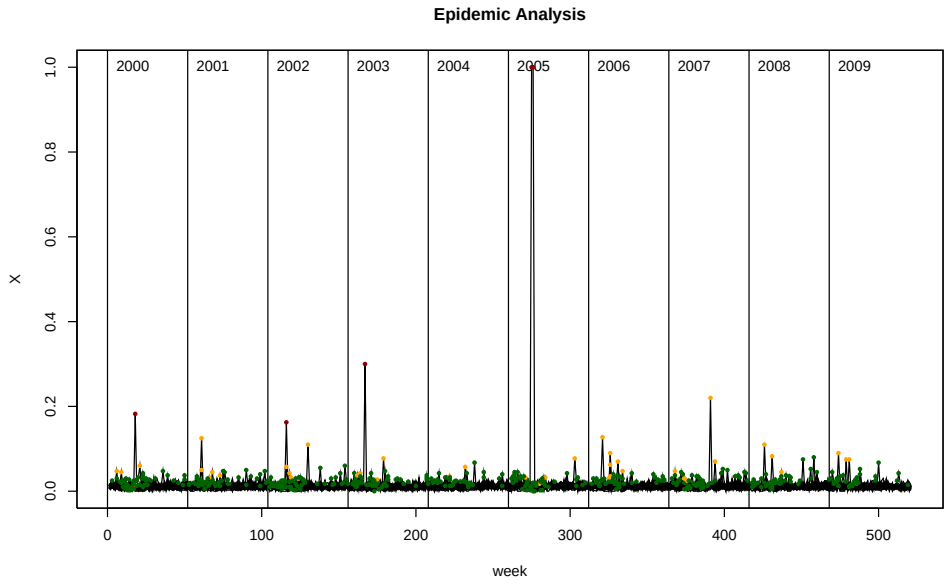

Figure A.32: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Enteritidis DT9a. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

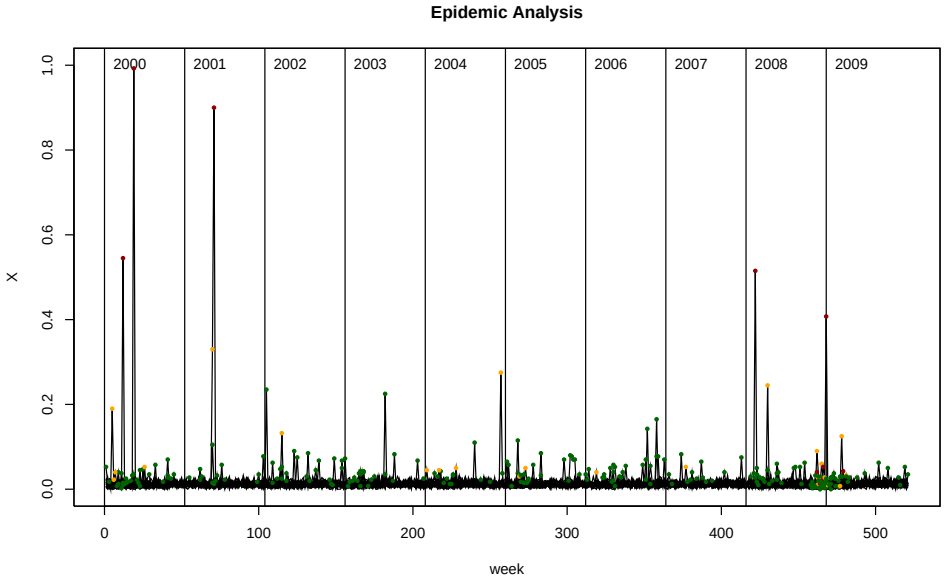

Figure A.33: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Typhimurium DT42. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

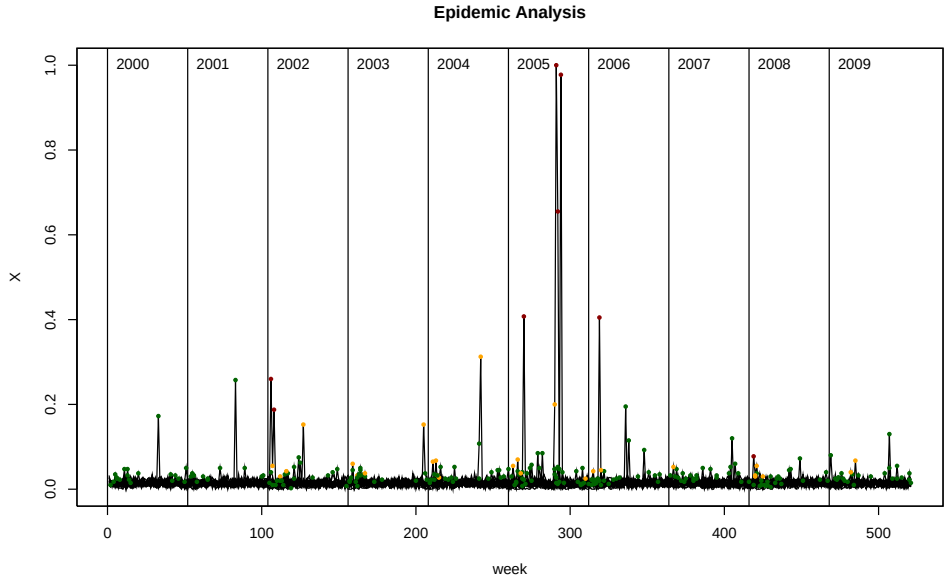

Figure A.34: The weekly posterior probability of an outbreak obtained from applying the model developed for salmonellosis as shown in Equation 4.2 at the Territorial Authority level for *Salmonella* Saint Paul. The coloured dots indicate the number of reported cases observed at that time point. Green dots represent 1 reported case, yellow 2 reported cases and a red dot means 3 or more reported cases were observed at that time point.

A.3.4 Age

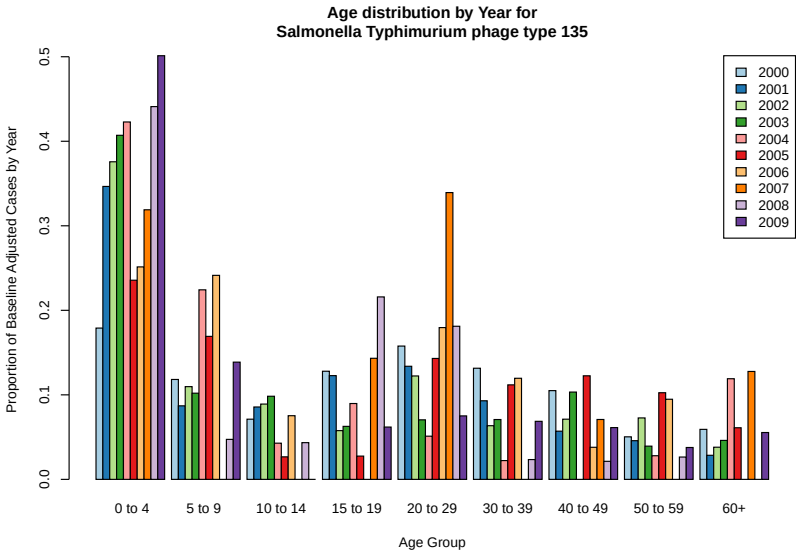

Figure A.35: Age distribution by Year for *Salmonella* Typhimurium DT 135

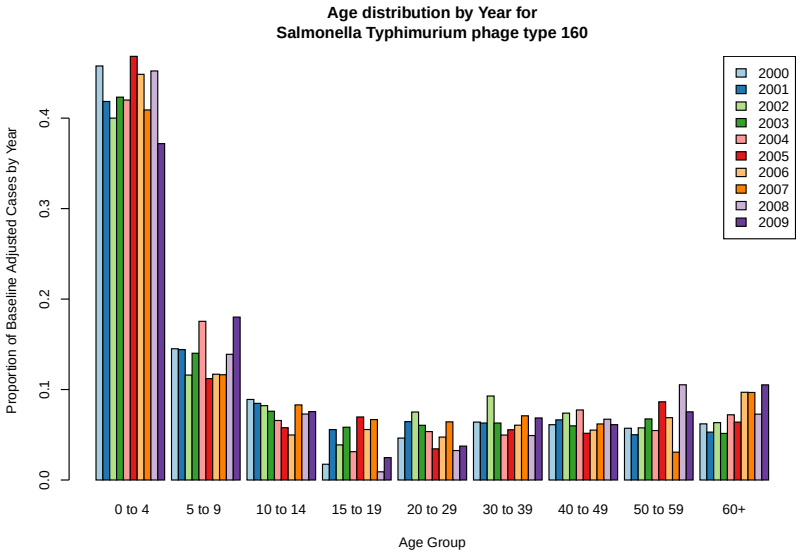

Figure A.36: Age distribution by Year for *Salmonella* Typhimurium DT 160

A.3.5 Water Supply Regions (Zcode)

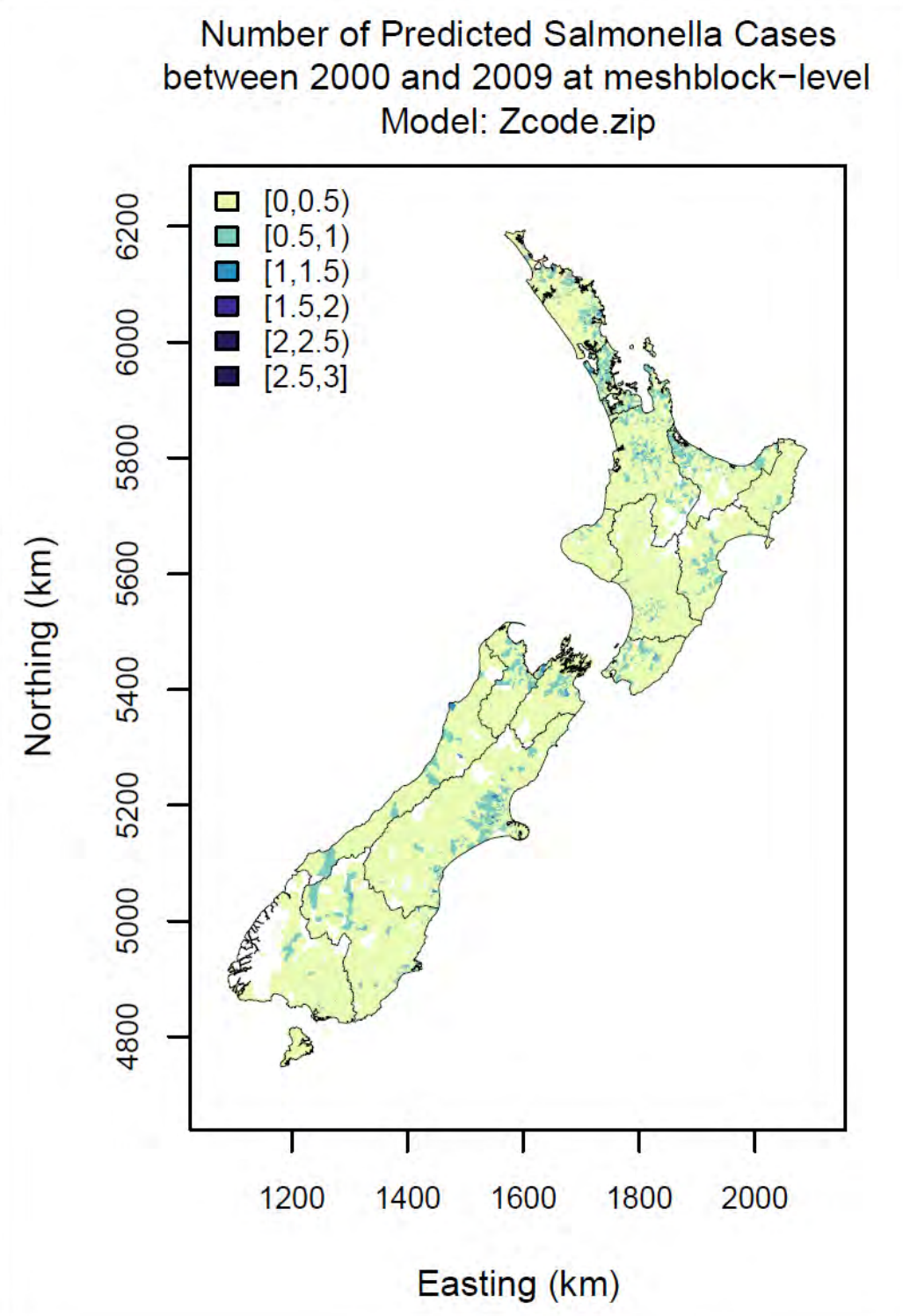

Figure A.37: Predicted number of reported *Salmonella* cases between 2000 and 2009 using Z-code as the only predictor

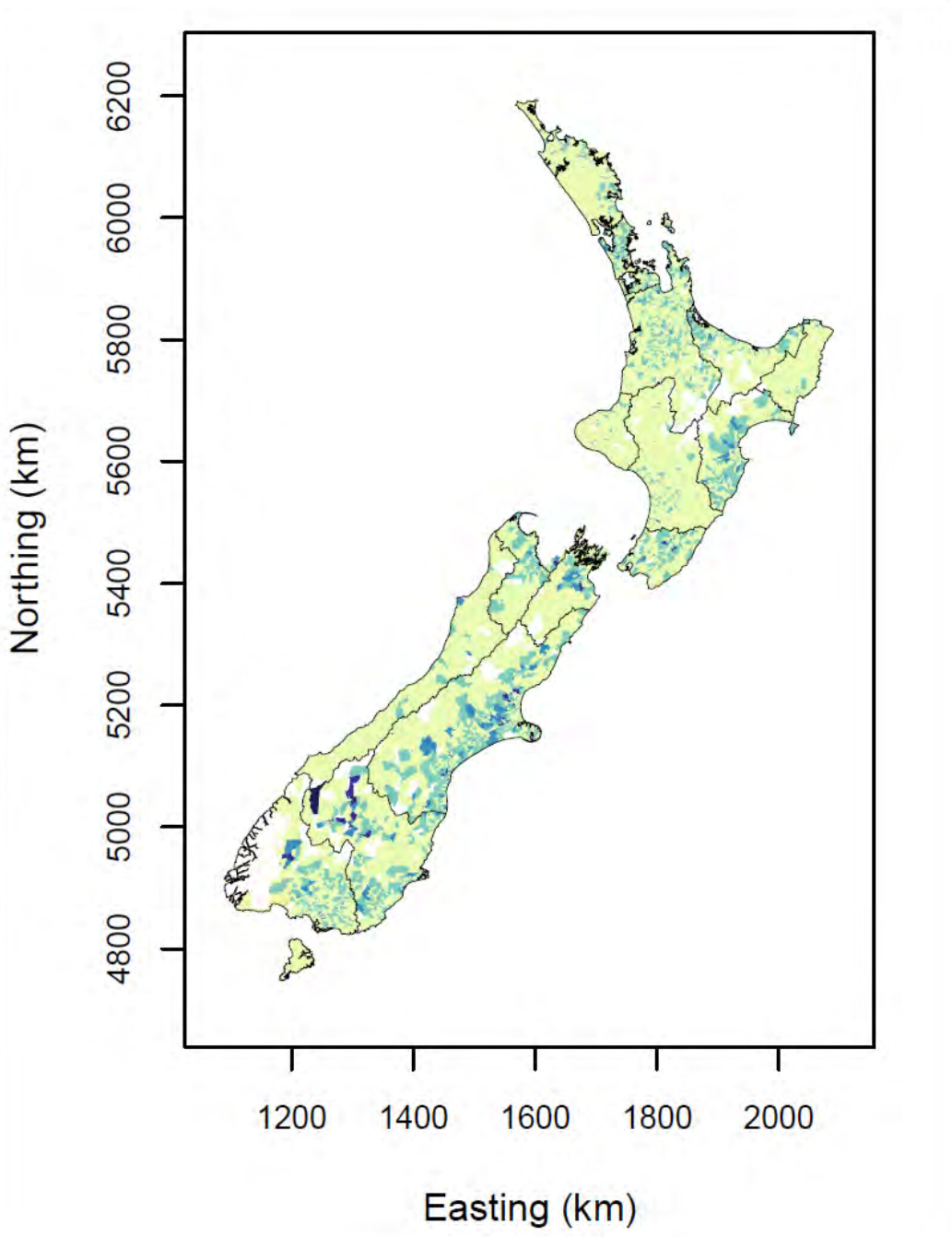

Figure A.38: Predicted number of reported *Salmonella* cases between 2000 and 2009 using Z-code, SDI, DHB, URProfile and Island as the predictors

Bibliography

[1] B. Adlam, K. Sexton, E. Lim, N. King, and R. Lake. Attribution of potentially foodborne enteric diseases: human salmonellosis. Technical report, Institute of Environmental Science & Research Limited, April 2010.

[2] H. Akaike. New Look at Statistical-Model Identification. *IEEE Transactions on Automatic Control*, AC19(6):716–723, 1974.

[3] M. R. Alley, J. H. Connolly, S. G. Fenwick, G. F. Mackereth, M. J. Leyland, L. E. Rogers, M. Haycock, C. Nicol, and C. E. Reed. An epidemic of salmonellosis caused by *Salmonella* Typhimurium DT160 in wild birds and humans in New Zealand. *N Z Vet J*, 50(5):170–6, 2002.

[4] M. G. Baker, C. N. Thornley, L. D. Lopez, N. K. Garrett, and C. M. Nicol. A recurring salmonellosis epidemic in New Zealand linked to contact with sheep. *Epidemiol Infect*, 135(1):76–83, 2007.

[5] M.G. Baker, E. Sneyd, and N.A. Wilson. Is the major increase in notified campylobacteriosis in New Zealand real? *Epidemiology and Infection*, 135(1):163–70, 2007.

[6] M. Callaghan, M. Azeem, G. Simmons, and I. Singsam. An outbreak of *Salmonella* Typhimurium phage type 160 in Auckland linked to an umu function. *New Zealand Public Health Report.*, 8 (6):44–45, 2001.

[7] R. G. Clark, S. G. Fenwick, C. M. Nicol, R. M. Marchant, S. Swanney, J. M. Gill, J. D. Holmes, M. Leyland, and P. R. Davies. *Salmonella* Brandenburg - emergence of a new strain affecting stock and humans in the South Island of New Zealand. *N Z Vet J*, 52(1):26–36, 2004.

[8] R. G. Clark, S. Swanney, C. M. Nicol, M. Leyland, and S. G. Fenwick. *Salmonella* Brandenburg - emergence of a variant strain on a sheep farm in the South Island of New Zealand. *N Z Vet J*, 51(3):146–7, 2003.

[9] R. Clarke and P. Tomlinson. Salmonella brandenburg: changing patterns of disease in southland province, new zealand. *N Z Med J*, 117(1205):U1144, 2004.

[10] J. H. Connolly, M. R. Alley, G. J. Dutton, and L. E. Rogers. Infectivity and persistence of an outbreak strain of *Salmonella enterica* serotype Typhimurium DT160 for house sparrows (*Passer domesticus*) in New Zealand. *N Z Vet J*, 54(6):329–32, 2006.

[11] T.M. Davies, M.L. Hazelton, and J.C. Marshall. sparr: Analyzing spatial relative risk using using fixed and adaptive kernel density estimation in R. *Submitted for publication.*, 2010.

[12] P. Diggle, L. Knorr-Held, B. Rowlingson, T-L. Su, P. Hawtin, and T.N. Bryant. On-line monitoring of public health surveillance data. In R. Brookmeyer and D.F. Stroup, editors, *Monitoring the health of populations*. Oxford University Press, 2002.

[13] P.K. Dunn and G.K. Smyth. Randomized quantile residuals. *J. Comput. Graph. Statistics*, 5:236–244, 1996.

[14] ESR. Disease outbreak manual. Technical report, Institute of Environmental Science and Research Limited, April 2002.

[15] P J Fedorka-Cray, J T Gray, and Wray C. Salmonella *in Domestic Animals*. CABI Publishing, New York, 2000.

[16] S. Jackman. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. Technical report, Stanford University, 2010. R package version 1.03.6.

[17] J. I. Kerslake and N. R. Perkins. *Salmonella* Brandenburg: case-control survey in sheep in New Zealand. *N Z Vet J*, 54(3):125–31, 2006.

[18] N. King and R. Lake. Review of notified salmonellosis outbreak data as a source of information for attribution. Technical report, Institute

of Environmental Science and Research Limited, Christchurch Science Centre, 2007.

[19] L. Knorr-Held and S. Richardson. A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Appl. Statist.*, 52(2):169–183, 2003.

[20] J.C. Marshall, S. Spencer, and N.P. French. SCIG-MAS-001 Final report for New Zealand Food Safety Authority: Development and application of new tools for the analysis of *Campylobacter* surveillance data: identifying the spatial and temporal determinants of raised notifications in New Zealand. Technical report, Massey University, August 2009.

[21] L. McCallum, M. Torok, M. T. Dufour, A. Hall, and G. Cramp. An outbreak of *Salmonella* typhimurium phage type 1 associated with watermelon in Gisborne, January 2009. *N Z Med J*, 123(1322):39–45, 2010.

[22] P. Mullner, G. Jones, A. Noble, S. Spencer, S. Hathaway, and N.P. French. Source Attribution of Food Borne Zoonosis in New Zealand: A Modified Hald Model. *Risk Anal*, 29:970–984, 2009.

[23] D. Petkov, J. Collins-Emerson, and NP French. Characterisation of salmonella. *report to MAF Biosecurity New Zealand*, 2010.

[24] M. Price-Carter, P. Roy-Chowdhury, C. E. Pope, S. Paine, D. E. Lisle GW, D. M. Collins, C. Nicol, and P. E. Carter. The evolution and distribution of phage ST160 within *Salmonella enterica* serotype Typhimurium. *Epidemiol Infect*, pages 1–10, 2010.

[25] G. Simmons, V. Hope, G. Lewis, J. Whitmore, and W. Gao. Contamination of potable roof-collected rainwater in Auckland, New Zealand. *Water Res*, 35(6):1518–24, 2001.

[26] S. J. Snel, M. G. Baker, V. Kamalesh, N. French, and J. Learmonth. A tale of two parasites: the comparative epidemiology of cryptosporidiosis and giardiasis. *Epidemiology and Infection*, 137(11):1641–1650, NOV 2009.

[27] S. Spencer, J. Marshall, R. Pirie, D. Campbell, and N.P. French. The detection of spatially localised outbreaks in campylobacteriosis notification

data. *Spatial and Spatio-temporal Epidemiology*, August 2011 accepted subject to minor ammendments.

[28] C. N. Thornley, G. C. Simmons, M. L. Callaghan, C. M. Nicol, M. G. Baker, K. S. Gilmore, and N. K. Garrett. First incursion of *Salmonella enterica* serotype typhimurium DT160 into New Zealand. *Emerg Infect Dis*, 9:493–5, 2003.

[29] L. E. Unicomb, G. Simmons, T. Merritt, J. Gregory, C. Nicol, P. Jelfs, M. Kirk, A. Tan, R. Thomson, J. Adamopoulos, C. L. Little, A. Currie, and C. B. Dalton. Sesame seed products contaminated with *Salmonella*: three outbreaks associated with tahini. *Epidemiol Infect*, 133:1065–72, 2005.

[30] N.A. Wilson and M.G. Baker. A systematic review of the aetiology of human campylobacteriosis in New Zealand. Technical report, Food Safety Authority of New Zealand, 2009.

[31] T. L. Wong, C. Nicol, R. Cook, and S. MacDiarmid. *Salmonella* in uncooked retail meats in New Zealand. *J Food Prot*, 70(6):1360–5, 2007.

[32] A. Zeileis, C. Kleiber, and S. Jackman. Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8), 2008.

List of Figures

List of Tables