Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

The Effect of Climate Variation on Infectious Diseases in Humans in New Zealand

Aleisha Brock

August 2011

The Effect of Climate Variation on Infectious Diseases in Humans in New Zealand

A thesis presented in partial fullfilment
of the requirements for the degree of
Masters in Veterinary Studies in Epidemiology
at Massey University, Palmerston North, New Zealand
under the supervision of
Dr. Jackie Benschop & Dr. Geoffrey Jones.
This work was sponsored by the
Institute of Environmental Science and Research Limited
as part of a Foundation for Research, Science and
Technology funded project (contract C03X0801).

Aleisha Brock

Submitted August 2011

Jeremiah 29:11-13

11 "*For I know the plans I have for you,*" declares the Lord,

"*plans to prosper you and not to harm you, plans to give you hope and a future.*

12 *Then you will call on me and come and pray to me,*

and I will listen to you.

13 *You will seek me and find me when you seek me with all your heart.*"

Abstract

The emergence and spread of infectious disease are a major issue associated with environmental change. Contributing to this is the effect climate variability and change may play in altering disease risk. The aims of this study were to investigate the association between climate and infectious diseases in humans throughout New Zealand from 1997 - 2007, then use the identified associations to project the burden of disease in 2015, 2040 and 2090 with respect to future climate change scenarios.

The four infectious diseases selected and investigated were campylobacteriosis, cryptosporidiosis, influenza hospitalisations and meningococcal disease. The association of weather variables and other confounders with the incidence risk (IR) of disease were explored using a quasi-Poisson generalized linear model, indicating that weather variables were significantly associated with disease risk. These results, along with expert opinion on epidemiological plausibility, were used to select confounders for the past association models.

Climate variation was associated with the IR of campylobacteriosis and cryptosporidiosis in New Zealand from 1997 - 2007. Campylobacteriosis notifications were found to be positively associated with the weekly absolute humidity. Additionally, campylobacteriosis notification risk factors were increasing beef and dairy density, intermediate and poor drinking water quality, being over 65 years of age and identifying with Pacific Island or Asian ethnicity. Protective factors were being less than 4 years of age and identifying with Maori ethnicity. Cryptosporidiosis notifications were found to be positively associated with the weekly average temperature and negatively associated with the weekly average rainfall. Risk factors for cryptosporidiosis notifications were poor drinking water quality, being less than 4 years of age and living in rural areas. Protective factors were identifying with Maori ethnicity and unknown drinking water quality.

Influenza hospitalisations and meningococcal disease notifications were not significantly associated with past climate variation. Identifying with Maori ethnicity was found to be a risk factor for influenza hospitalisations, with no protective factors identified. Risk factors for meningococcal disease notifications were an increasing social deprivation index (SDI) score,

being less than 4 years of age and identifying with Maori ethnicity. Identifying with Asian ethnicity was a protective factor for meningococcal disease.

The projection calculations of the change in disease incidence from the study period to 2015, 2040 and 2090 were carried out under a combination 3 Intergovernmental Panel of Climate Change (IPCC) climate scenarios (A2, B1 and A1B) and 12 downscaled global climate models. The projected change in campylobacteriosis and cryptosporidiosis suggested increases in the rate of notifications, and a small to no decrease in influenza hospitalisation and meningococcal disease notifications.

Acknowledgements

It is a pleasure to thank those who helped make this thesis possible.

To my supervisors, Dr. Jackie Benschop and Dr. Geoff Jones, I sincerely thank you for your help, guidance and encouragement in getting this work completed. With differing strengths and specialities, together you created a healthy work environment in which I increased in both knowledge and skills that I will continue to draw upon in the future. For this I am thankful.

With gratitude I thank my project leader, Dr. Dave Slaney of ESR for your guidance and encouragement. It has been a great honour to work with you and the FRST team.

During this thesis I have collaborated with many colleagues who I have great regard for, and I wish to thank all of those who have helped me with my work at both the EpiCentre and Statistics Department at Massey University, and those on the FRST project.

A special mention to Dr. Jonathan Marshall of Massey University, whose incredible computer coding skills helped make a seemingly impossible task less daunting. I'd also like to thank Professor Nigel French of Massey University for your help and guidance and Dr. Simon Spencer of Warwick University for the use of your Knorr-Held Richardson R-code. To Simon Verschaffelt of Massey University, thank you for always being willing to help with computer and programming dilemmas. To Dr. Andrew Tait of NIWA, thank you for your hard work producing the weather data and projections used in this thesis and for your guidance in using them. Thank you to all who helped with data sourcing and GIS coding.

A special thanks to Kate Holland for her contribution during her summer studentship and ongoing friendship. You helped fill our office with laughter and fun, and I wish you all the best with your future endeavours!

I sincerely appreciate the financial support from the Institute of Environmental Science and Research Limited (ESR) as part of a Foundation for Research, Science and Technology funded project (contract C03X0801).

I would like to thank the Bebarfald, Meade and Brock families for their unlimited support, love and encouragement. To Grandpa & Gran, thank you for your encouragement, prayers and foresight with financial provisions. To Nana & Grandad, thank you for your prayers and support. To Tom & Kate Brock, thank you for your love, support and numerous dinners. To my siblings, Jessica, Aimee, Joshua & Caleb, thank you for your support, laughs, prayers and help in procrastination - there has never been a dull moment! Lastly, and most importantly, I wish to thank my husband Ian Brock, and my parents Steven & Susan Meade for their love, support and prayers throughout the past two years. To them I dedicate this thesis.

"Men go abroad to wonder at the heights of mountains, at the huge waves of the sea, at the long courses of the rivers, at the vast compass of the ocean, at the circular motions of the stars, and they pass by themselves without wondering." (Saint Augustine, 354 – 430 AD)

Abbreviations

AH	Absolute humidity
AU	Area unit
AR	Autoregressive term
ARIMA	Autoregressive integrated moving-average model
ARMA	Autoregressive moving-average model
CAR	Conditional autoregressive models
C.I.	Credible Interval
CSF	Cerebrospinal fluid
DHB	District health board
ESR	Institute of Environmental Sciences and Research Ltd.
FRST	Foundation for Research, Science and Technology
GLM	Generalised linear model
HAIFA	The Health Analysis and Information for Action project
MeNZB	Meningococcal B strain vaccination for New Zealand
i.i.d	Identically and independently distributed
IPCC	Intergovernmental Panel on Climate Change
IR	Incidence risk
MA	Moving-average term
MAF	Ministry of Agriculture and Forestry
MCMC	Markov chain Monte Carlo
NIWA	National Institute of Water and Atmospheric Research Ltd.
NZCCC	The New Zealand Climate Change Centre
NZHIS	The New Zealand Health Information Service
OR	Odds ratio
PAR	Population at risk
PCA	Principle component analysis
PCR	Polymerase chain reaction
sARIMA	Seasonal autoregressive integrated moving-average model
SDI	Social deprivation index
VCS	Virtual climate station
VTEC	Verotoxin producing *Escherichia coli*

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

The emergence of infectious diseases and changing incidence of disease is a major issue associated with environmental change. Contributing to this is the effect climate variability and change may play in altering disease risk. Increasing knowledge and understanding of the effect of climate change on population health will need to play a vital role in shaping health policies throughout New Zealand to control climate-sensitive health determinants and outcomes [25, 53, 54, 68]. Based on both evidence and reasonable presumption, researchers conclude that climate change has already had an effect on the health and survival of people from various regions [25, 53, 68].

1.1.1 Climate Change

The climate change scenarios discussed in this thesis are obtained from the Intergovernmental Panel on Climate Change (IPCC) in 1996 [37]. There are four qualitative storylines forming the basis of the climate scenarios which branch into 6 scenarios. The three emission scenarios of interest in this work are the A2, B1 and A1B scenarios. The projection years (2014, 2040 and 2090) were driven by the availability of data under these climate scenarios.

The A2 climate scenario focuses on self-reliance and preservation of local identities, describing a very heterogeneous world based on a high population growth (15 billion by 2100), resulting in more extreme climate projections. The B1 climate scenario describes a convergent world, with the lowest population growth trajectory (8.7 billion by 2050, decreasing to 7 billion by 2100), resulting in conservative climate change projections. The A1B climate scenario is derived from the A1 scenario with an influence from B1, focusing on a balanced technological change in energy systems and economic growth.

For more information on the climate scenarios available, please see the IPCC Special Report on Emissions Scenarios [37].

1.1.2 Health in New Zealand

Notifiable diseases are those that require notification by medical practitioners to the local Medical Officer of Health. This information is collated in a computer system called EpiSurv, at the Institute of Environmental Sciences and Research (ESR).

Below is an overview of the most recent epidemiology status and burden of each disease in New Zealand, up to and including 2010. However, it must be noted that the study period of the data used in this thesis is 1997 - 2007.

Campylobacteriosis

Campylobacteriosis is generally a self-limiting disease caused by the *Campylobacter* bacterium, primarily via zoonotic transmission [58]. It is among the most common bacterial infections of humans and its symptoms include diarrhoea, sometimes bloody, cramps, fever and abdominal pain [58].

Burden of Disease

In New Zealand, there were 7346 cases of notified campylobacteriosis in 2010, similar to the 7177 notified cases in 2009 [60]. As shown in Figure 1.1 from the 2010 ESR Annual Surveillance Report, the number of cryptosporidiosis cases peaked in 2007, with a marked decrease thereafter [60]. This drop in the 2007/2008 incidence of campylobacteriosis has been attributed to the introduction of a range of voluntary and regulatory interventions in the poultry industry [67]. The HAIFA project study period incorporates this rise in campylobacteriosis incidence up to and including the rate in 2007, but does not take into account the effect of the poultry interventions, i.e. the drop seen from 2008 - 2010. This must be noted as a limitation during the projection calculation (Chapter 7) and the effect such interventions has on projection accuracy will be explored further in the Discussion (Chapter 8).

Under-reporting is thought to be an issue with campylobacteriosis in New Zealand, both from patients and differing District Health Board (DHB) practices. The true population rate of campylobacteriosis in New Zealand is likely to be higher than the notified cases suggest as only a small proportion of cases will seek medical attention and provide specimens for lab testing [7]. Under reporting was identified as a problem in England, where a multiplier of 7.6 was applied when estimating the true incidence from the number of nationally notified cases [7].

In 2010, age was recorded in 99.9% of the notified cases of campylobacteriosis in New Zealand, with the highest rate occurring in children from 1 to 4 years of age (314.4 per 100,000, 780 cases) followed by children less than 1 year (279.3 per 100,000, 178 cases) [60]. Males were found to have a higher rate of notified campylobacteriosis than females (190.9 cases per 100,000 and 144.3 cases per 100,000, respectively) [60]. Ethnicity was recorded for 88.9% of the notified cases of campylobacteriosis in New Zealand during 2010. European ethnicity was found to have the highest rate of campylobacteriosis notifications in New Zealand (206.2 per 100,000, 5556 cases) followed by Other (144.6 per 100,000, 49 cases), Asian (101.2 per 100,000, 345 cases), Maori (79.8 per 100,000, 451 cases) and Pacific Peoples (58.3 per 100,000, 132 cases) [60].

In 2010, the Taranaki and Hutt Valley District Health Boards (DHB) had the highest reported rates, with Taranaki reporting 84.1 per 100,000 population (47 cases) and Hutt Valley reporting 238.5 per 100,000 population (343 cases). The lowest rates were reported in Tairawhiti and Counties Manukau, with Tairawhiti reporting 79.6 per 100,000 population (37 cases) and Counties Manukau reporting 108.8 per 100,000 population (534 cases), as shown in Figure 1.2 [60].

Seasonality

Campylobacteriosis exhibits clear seasonality, with summer peaks and winter troughs [7, 32, 60]. In a study carried out in Denmark, they concluded that the maximum temperature 4 weeks prior to the reported infection was the best single predictor of campylobacteriosis incidence, explaining 68% of the variation [58]. A weak association between the peak of campylobacteriosis infections and a lag in temperature by 3 months was discussed by Baker et al. (2007), lending evidence to the suggestion that an increase in average temperature since the 1990's may have favoured *Campylobacter* survival [7]. Bi et al. (2008) conclude that the effect of weather on *Campylobacter* is different in varying cities in Australia, highlighting the need of further studies to increase the understanding of important epidemiological factors to control campylobacteriosis [13].

A study of the spatial and temporal variation of campylobacteriosis throughout New Zealand from 1993 to 2000, suggested that the seasonality of campylobacteriosis could be characterised using the following three groupings: rural North Island, urban North Island and urban South Island [32]. In the rural North Island has a relatively low summer incidence and low inter-seasonal variation is seen. Urban North Island has a relatively high summer incidence and a greater degree of seasonality than Rural North Island. The urban South Island experiences the highest summer incidence and in general the greatest seasonal variation [32]. These findings were supported by another study of the spatial distribution of campylobacteriosis in New Zealand from 1997 - 2005 [64].

Risk Factors

The risk factors associated with campylobacteriosis notifications in New Zealand during 2010 [60], from highest to lowest risk were as follows:

- contact with farm animals,
- food from retail premises,
- consuming untreated water,
- contact with faecal matter,
- recreational water contact,
- contact with other symptomatic people,
- contact with sick animals, and
- overseas travelling.

The animal-related risk factors were supported by Gilpin et al. (2008) who discuss that *C. jejuni* isolates were found in the faeces of a range of animals in New Zealand, including chickens, ducks, dogs, cats, sheep, dairy and beef cattle [29, 30]. Additional risk factors associated with campylobacteriosis in New Zealand are the consumption of raw or uncooked poultry, handling uncooked poultry, the consumption of unpasteurized or contaminated milk and barbecuing [29, 32].

Cryptosporidiosis

Cryptosporidiosis is a condition caused by the *Cryptosporidium spp.*, causing diarrhoeal symptoms that can last anywhere between several days to several weeks.

Burden of Disease

In 2010, there were 954 cases of notified cryptosporidiosis, a significant increase from the 845 cases in 2009 [60]. The incidence in New Zealand is one of the highest reported rates in the world [46, 71]. As shown in Figure 1.3 from the 2010 ESR Annual Surveillance Report, the number of cryptosporidiosis cases peaked in 2001, with a marked decrease until 2004 [60]. The number of notifications then increased through to 2010. Under-reporting is thought to be an issue with cryptosporidiosis in New Zealand, both from patients and differing District Health Board (DHB) practices.

In 2010, age was recorded in 99.7% of the notified cases of cryptosporidiosis in New Zealand [60], with 54.9% of the cases occurred in children less than 15 years of age (522 cases). Children from the age of 1 – 4 years had the highest rate of cases (115.31 per 100,000, 286 cases), followed by children less than 1 year of age (53.4 per 100,000, 34 cases), with the lowest rate in the group who were 70 years and older (4.1 per 100,000, 16 cases) [60]. In New Zealand, males were found to have a higher number of gender-specific notified cryptosporidiosis cases than females (22.5 cases per 100,000, 482 cases and 21.0 cases per 100,000, 466 cases, respectively) in 2010 [60]. European ethnicity was found to have the highest rate of notifications in New Zealand (29.3 per 100,000, 790 cases) followed by Other (23.6 per 100,000, 8 cases), Maori (12.2 per 100,000, 69 cases), Asian (6.5 per 100,000, 22 cases) and Pacific Peoples (5.3 per 100,000, 12 cases) [60].

In 2010, the South Canterbury DHB had the highest reported rates (84.1 per 100,000 population, 47 cases) followed by Canterbury DHB (45.3 per 100,000, 230 cases), with the lowest rate in Hutt Valley DHB (6.3 per 100,000 people, 9 cases), as shown in Figure 1.4 [60]. The level of deprivation in each area has been shown to have an inverse relationship with the rate of infection [72].

Seasonality

Cryptosporidiosis exhibits clear seasonality, with peaks in spring and autumn [46, 60]. In a study carried out to explore the seasonal effects on human cryptosporidiosis in New Zealand, they concluded that the current and previous month's temperature were positively associated with cryptosporidiosis in summer and autumn [46]. The results of a study carried out in Brisbane, Australia suggested that weather variability possibly plays a significant role in *Cryptosporidium* transmission [34].

In New Zealand, the spring peak is larger than the autumn peak and found to be predominantly caused by *C. parvum*, while the peak in autumn is found to be predominantly caused by *C. hominis* [46]. The main calving and lambing season coincides with the spring peak where the bovine species dominates, leading researchers to believe it is associated strongly with farming practices [46, 48, 72]. Snel et al. (2009) discusses that areas in the North Island with increased cattle density were found to have increased rates of human cryptosporidiosis cases [72]. The South Island is found to have a higher spring peak than in the North Island, suggesting that this is attributable to an increased sheep density and that sheep may act as a source of infection [46, 72].

Risk Factors

The risk factors associated with cryptosporidiosis notifications in New Zealand during 2010 [60], from highest to lowest risk are as follows:

- contact with farm animals,
- recreational water contact,
- consuming untreated water,
- contact with faecal matter,
- contact with other symptomatic people,
- food from retail premises,
- contact with sick animals, and
- overseas travelling.

These risk factors have been identified in previous studies on human cryptosporidiosis [2, 46, 71, 72]. A rural study carried out in rural New South Wales, Australia, found a 137% increase in *Cryptosporidium* cases from 2004 – 2005, where the *C. parvum* strain was more prevalent than *C. hominis* [57]. They suggested that zoonotic transmission from cattle may explain these results. In areas of Australia where less intensive farming of cattle occurs (e.g. Western Australia and South Australia) a higher prevalence of *C. hominis* (82-83% and 72%, respectively) than *C. parvum* (17-18% and 28%, respectively) was found.

In 2008, Xiao and Feng reported that food or water contamination by cattle manure has been identified as the cause of several foodborne and waterborne outbreaks of cryptosporidiosis worldwide [81]. An association with contaminated water source was identified by Snel et al. (2009) in New Zealand, however no association was found with rainfall or river flow [71]. Duncanson et al. (2000) outlined the need for New Zealand to make the improvement to drinking water quality in small communities and rural areas a high priority in the hope of reducing the incidence of cryptosporidiosis in New Zealand [24].

Influenza Hospitalisations

Influenza is an acute viral respiratory disease that lead to 998 hospitalisations in New Zealand during 2010, where influenza was the primary diagnosis [60]. The rate in 2010 was lower than the 1517 hospitalisations in 2009, but higher than the 365, 316 and 464 hospitalisations in 2008, 2007 and 2006, respectively [59, 60]. The most common symptoms include chills, fever, sore throat, muscle pains, severe headache, coughing, weakness/fatigue and general discomfort. It is spread via person-to-person contact when the infected individual coughs, sneezes or breathes [51].

Burden of Disease

In Figure 1.5, the 2010 Annual Surveillance Report by ESR describes the trend of hospitalisations due to influenza in 2010 by week [60], where 95.4% of the hospitalisations occurred between June and October. Under-reporting is likely to occur as the data is obtained from hospital records, as very few people hospitalised with influenza are laboratory confirmed. Influenza vaccinations were available to the New Zealand population throughout the study period, 1997 to 2007, and taken into account in this project.

Seasonality

Influenza exhibits seasonality with a well defined epidemic almost every winter and a low level of disease existing throughout the year [51, 82]. In a study carried out in Hong Kong on the seasonal effects of influenza on mortality, the researchers conclude that seasonal factors possibly have an impact on the virulence of influenza, as well as viral transmission [82].

Risk Factors

Two transmission factors highlighted by Lofgren et al. (2007) were crowding and the use of indoor heating [51]. During crowding, the person-to-person spread is greatly enhanced by the infected subject being surrounded by a dense population of susceptible individuals. It was further discussed that the seasonality seen in influenza may be caused by changes in host behaviour. The use of indoor heating is another behavioural pattern that may increase the transmission of influenza viruses. During cooler seasons (decreased temperatures) an increase in indoor heating provides continuously recirculated air with low humidity, which is ideal for viral particles to remain suspended in the environment.

Meningococcal Disease

Meningococcal disease is a life threatening illness caused when encapsulated strains of *Neisseria meningitidis* invade the blood stream and often the meninges [63]. The transmission of the *Neisseria meningitidis* bacteria occurs when rhinopharynx secretions are emitted when an asymptomatic carrier or infected individual speaks or coughs [45, 63]. Only a small proportion of those infected develop the disease.

Burden of Disease

In 2010, New Zealand had 96 cases of meningococcal disease (2.4 per 100,000 population) with 84 laboratory confirmed [60]. As shown in Figure 1.6 from the 2010 ESR Annual Surveillance Report, the number of cases peaked during an epidemic in 2001 with a rate of 17.4 per 100,000 population, with the 2010 rate still higher than the pre-epidemic years (1989-1990) [60]. Under-reporting is unlikely to be a factor with meningococcal disease incidence throughout New Zealand due to the severity of the disease and requirement of urgent medical attention.

In 2010, the age, gender and ethnicity of each notified case of meningococcal disease was recorded in New Zealand [60]. The highest rate of notifications occurred in children less than 1 year (47.7 per 100,000, 27 cases) followed by children aged between 1 to 4 years (10.5 per 100,000, 23 cases) [60]. In New Zealand, males and females had similar notification rates of meningococcal disease in 2010 (2.4 cases per 100,000, 48 cases and 2.3 cases per 100,000, 48 cases, respectively) [60]. Pacific Peoples were found to have the highest rate of meningococcal disease notifications in New Zealand (7.5 per 100,000, 17 cases) followed by Maori (7.4 per 100,000, 42 cases) and European ethnicities (1.3 per 100,000, 35 cases) [60].

In 2010, the highest rates of meningococcal disease were recorded in the Hutt Valley (5.9 per 100 000 population, 8 cases) and Hawke's Bay DHBs (4.7 per 100 000, 7 cases) [60]. The Canterbury (1.5 per 100 000, 7 cases) and Auckland (1.7 per 100 000, 7 cases) DHBs had the lowest rates. There were no cases reported from the Whanganui, Wairarapa, and South Canterbury DHBs.

The populations identified in the ESR Annual Surveillance Report [60] to be at higher risk of meningococcal disease in New Zealand are consistent with the findings in other research [31, 43, 45]. From 1985 to 1986, New Zealand experienced a meningococcal disease serogroup A outbreak, with higher rates occurring in infants and young children, and those who identify with Maori and Pacific Island ethnicities [43]. These populations were again observed to have a higher risk of disease in the 1991 meingococcal disease serogroup B epidemic, with over half the cases occurring in children less than 5 years old [31, 43]. Outbreaks of meningococcal disease are also reported to be associated with overcrowding [5, 6] and occur in partly closed communities, such as schools [63].

In 2004, the New Zealand Ministry of Health rolled out the National Meningococcal B Immunisation Programme which aimed to vaccinate 90% of people from 6 months to 19 years old [17, 42]. The programme consisted of three doses of the MeNZB vaccination administered between July 2004 and June 2006. Data describing the MeNZB vaccine coverage was not obtained for the HAIFA project. The New Zealand wide coverage was not achieved until the end of our study period (2007). Due to the extended roll out period, there was no complete spatial or temporal dataset for New Zealand. More information on the MeNZB vaccine and coverage can be obtained from the New Zealand Ministry of Health [17].

Seasonality

In a study carried out in Auckland, the effects of meteorological conditions on meningococcal disease were explored [49]. Meningococcal disease was found to vary with season, increasing with high humidity and cooler temperatures, and decreasing with prolonged periods of heavy rain.

Risk Factors

A study carried out on the households of meningococcal disease patients in Auckland, New Zealand from October 1996 to September 1998 [70] identified risk factors associated with the carriage of *Neisseria meningitidis* bacteria as:

- living in a smoking household,
- being of Maori ethnicity,
- being male, and
- those aged 0 - 4 years or 15 - 24 years compared to those over 25 years of age.

Individuals with a cough had a slightly increased risk (OR 1.3) of being a carrier, whereas individuals with a runny nose associated with lower risk of carriage (OR 0.7) [70].

Household crowding (or household density) was identified to be strongly associated with the incidence of meningococcal disease (OR 10.7, $p < 0.0001$), having adjusted for other confoundering factors, during a case-control study carried out in Auckland from May 1997 to March 1999 [6]. Baker et al. (2000) conclude that measures reducing household overcrowding could greatly reduce the incidence of meningococcal disease in Auckland children [6].

1.2 The Health Analysis and Information for Action (HAIFA) Project

This work was carried out as part of my Masters Study at the Turitea Campus of Massey University, New Zealand. The funding and resources for this thesis were provided by the Population and Health Department at the Environmental Sciences and Research Ltd (ESR) as part of a joint-collaboration project called The Health Analysis and Information For Action project (HAIFA).

HAIFA is a Foundation for Research, Science and Technology (FRST) funded multi-disciplinary project which is currently developing a resource system for reducing New Zealand's vulnerability to human health impacts of climate variation and change. System development has occurred through three inter-related objectives: Objective 1: Identify and Model Indicator Diseases; Objective 2: HAIFA System Development; and Objective 3: Identify Responses and Adaptive Strategies. This thesis comprises work completed as part of Objective 1.

The Advisory Panel are representatives from various organisations throughout New Zealand, Australia and Europe, such as Medical Officers of Health, Maori and Pacific Islander representatives, World Health Organization and the New Zealand Ministry of Agriculture and Forestry (MAF) Biosecurity.

1.3 Purpose and Scope of this Thesis

The aims of this thesis were to investigate the association of climate variation with four indicator diseases throughout New Zealand from 1997 - 2007 at a fine spatial and temporal level. Each disease chosen to be investigated was selected as an "indicator" disease based on a number of criteria (see Chapter 4) and in particular ones that would likely be affected by climate change. These associations were then used to produce projection calculations for the percentage change in incidence risk of these diseases in the years 2015, 2040 and 2090. Through the HAIFA project, the projected percentage change in incidence risk will be made available to policy makers through the use of a web-based tool, the details of which are beyond the scope of this thesis.

An overview of the statistical model approaches considered in this thesis are outlined in Chapter 2, with a strong focus on techniques used in this study. Firstly, the data format and necessary considerations are highlighted, then the chapter continues with an overview of statistical modelling approaches considered for this thesis. The modelling techniques described strongly focus on Poisson regression and the Knorr-Held Richardson Model.

Chapter 3 describes the data obtained by ESR from various organisations and used in the HAIFA project. A feature of the data are the large number of covariates, each with a spatial and temporal correlation structure.

Chapter 4 describes the disease selection process in which four indicator diseases were chosen to represent the effect of climate on infectious diseases throughout New Zealand. These diseases were selected from an original list of 52 notifiable diseases in New Zealand. The inclusion of non-notifiable seasonal influenza hospitalisations was advised by the FRST Advisory Panel. Fourteen of the 53 diseases were selected on the HAIFA project as possible indicator diseases and tested for sufficient incidence risk and seasonal behaviour, resulting in seven diseases identified as eligible indicator diseases. In the final step of selection, the four indicator diseases chosen and endorsed by the FRST Advisory Panel were campylobacteriosis, cryptosporidiosis, influenza hospitalisations and meningococcal disease.

Chapter 5 explores the past association of weather variables and other possible confounders with the IR of each indicator disease during a screening process. The modelling technique used was a quasi-Poisson generalized linear model (GLM).

Chapter 6 describes the method used to model the effect of past climate variation on the four indicator diseases in New Zealand from 1997 - 2007, at a fine spatial and temporal level (574 weekly measurements for each of the 11,491 virtual climate station grids). Expert opinion on epidemiological plausibility, alongside the screening results from Chapter 5, were used to select the confounding variables for each model.

Chapter 7 describes the method used to project the percentage change in incidence risk (IR) of each indicator disease in New Zealand for the years 2015, 2040 and 2090. This was carried out using a combination of three Intergovernmental Panel of Climate Change (IPCC) emission scenarios (A2, B2 and A1B) and 12 downscaled global climate models.

This document describes the data and methodology used to investigate the past associations of climate variation on disease incidence in New Zealand and project the disease burden for 2015, 2040 and 2090, along with their limitations. In the final chapter, we report and discuss the results, including a discussion on what these results suggest about infectious disease incidence in New Zealand in the future.

A flow chart summarising the chapters contained within this thesis is provided in Figure 1.7.

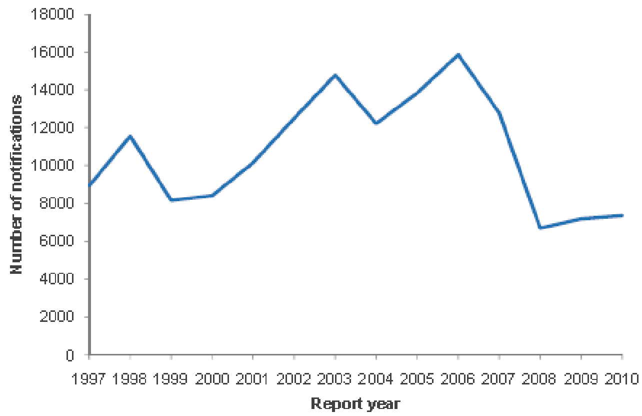


Figure 1.1: Campylobacteriosis time series of the number of notified cases from 1997 - 2010 [60].

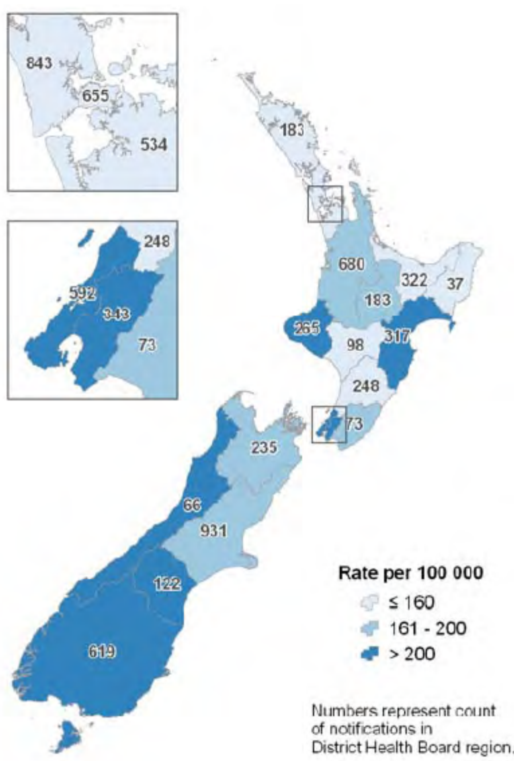


Figure 1.2: Campylobacteriosis rate per 100,000 by District Health Board in 2010 [60].

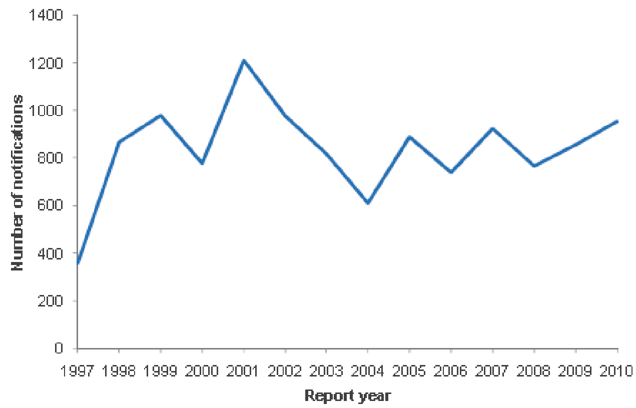


Figure 1.3: Cryptosporidiosis time series of the number of notified cases from 1997 - 2010 [60].

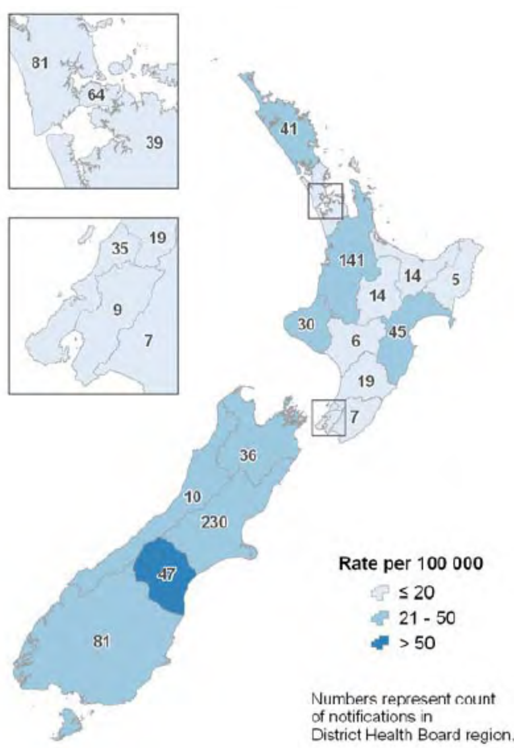


Figure 1.4: Cryptosporidiosis rate per 100,000 by District Health Board in 2010 [60].

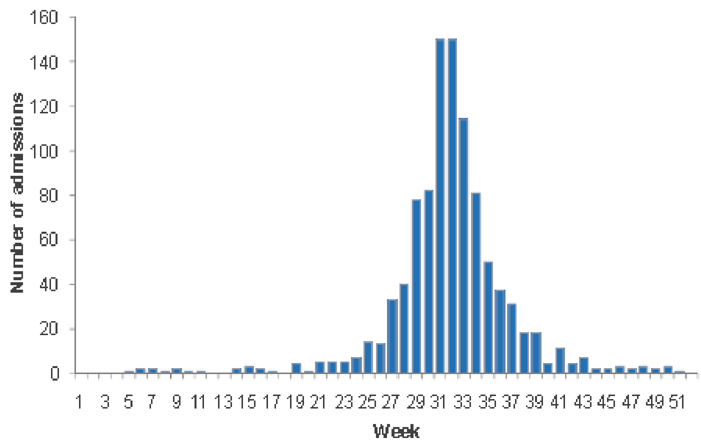


Figure 1.5: Influenza hospitalisations by week discharged in 2010 [60].

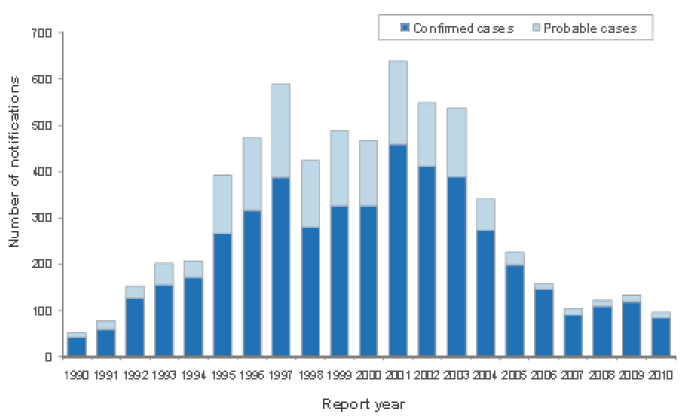


Figure 1.6: Meningococcal disease notifications by year, 1990 - 2010 [60].

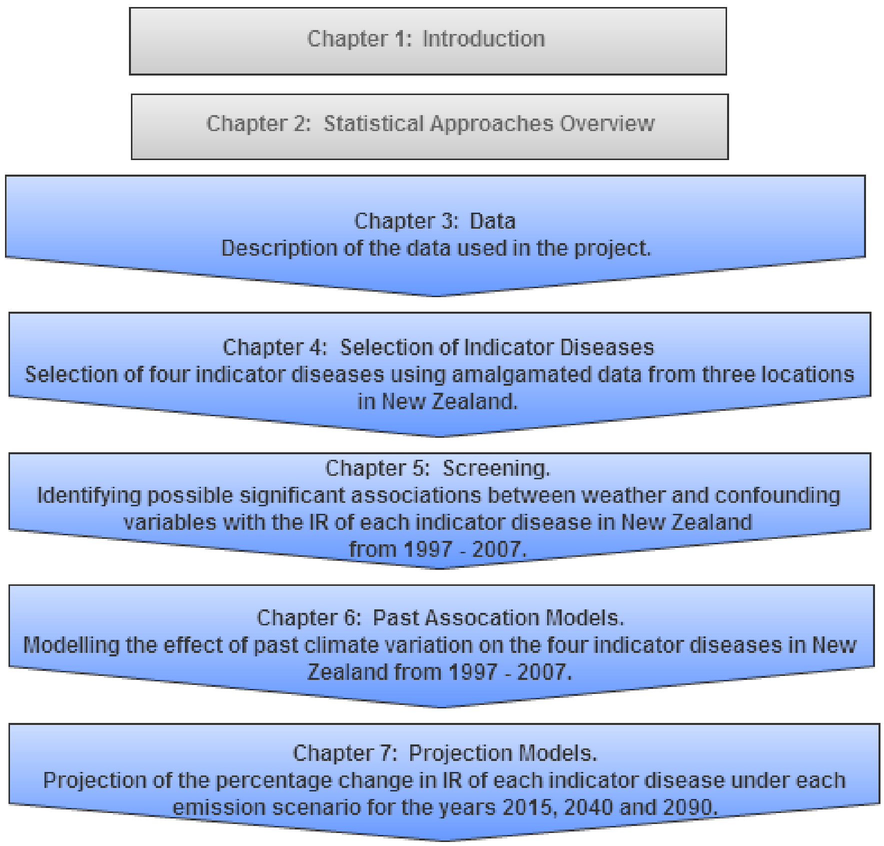


Figure 1.7: An outline of the chapters contained within this thesis.

Chapter 2

Statistical Approaches Overview

2.1 Introduction

An overview of the statistical model approaches currently available for modelling and projecting the effect of past climate on infectious disease is outlined below, with a strong focus on techniques used in this thesis.

The overview of statistical model approaches in this section focus on methods for screening covariates, techniques to produce a past association model and projection calculation within a Bayesian framework, with reference to past research carried out on similar data format and diseases. The statistical models used will need to take into account correlation both spatially and temporally.

Due to the complex nature of the data used in this thesis, it is important to start by describing the data format and considerations required for statistical methods. The data format is comprised of 11,491 5km x 5km grids throughout New Zealand, each containing a weekly time series spanning eleven years of disease count, weather and other possible confounding variables. The data spans from January 1997 to December 2007.

The outcome variable, the count of diseases (Y_{it}) in each grid (i) during week (t), is assumed to be Poisson distributed, with a mean (λ) and variance (λ). The explanatory variables are weather and other possible confounders which will need to be adjusted for (such as rurality and gender).

A number of considerations need to be taken into account when modelling this type of data. These include spatial and temporal correlation in the disease counts; seasonality exhibited in both outcome and explanatory variables; and collinearity between the explanatory and possible confounders. These will be discussed in the modelling approaches outlined below.

This chapter concludes with a discussion of visualisation techniques used to present the projected percentage change in IR for each indicator disease, under each climate scenario (A2, A1B, B1) for 2015, 2040 and 2090.

2.2 Statistical Models

There are various modelling techniques that could be used to model data with both spatial and temporal structure. It is not my intention to provide a review of all techniques available, but an overview of the techniques considered for this work. For further reading on other statistical methods available, please refer to the following research and review articles by Afifi et al. (2007) [1], Berk and MacDonald (2008) [10], Fisman (2007) [26], Lindsey (1999) [50] and Schmidt and Pereira et al. (2011) [65].

Collinearity

The methods outlined in the overview of statistical analysis below can be affected by collinearity between explanatory variables. Collinearity refers to the case when predictors are highly correlated. It often occurs with time-series data, where different variables (e.g. temperature and humidity) tend to rise and fall together. The methods of identifying and dealing with collinearity are well documented, with possible methods including factor analysis, principal component analysis (PCA) or only including one variable exhibiting collinearity within the statistical analysis. In a study examining the seasonal and temporal variation in the notification rates of campylobacteriosis in New Zealand, PCA was used to eliminate correlation between seasonal rates of campylobacteriosis by collapsing four seasonal crude rates into two principal components [32].

2.2.1 Poisson Regression

Poisson regression is used to model rare events in the form of count data, where the outcome variable (Y) is Poisson distributed, with the mean (λ) equal to the variance (λ), i.e. $Y \sim Pois(\lambda)$. Generalized linear models (GLM) are a framework by which Poisson regression modelling can be analysed, developed in the early 1970s [83]. GLMs are an extension of the linear regression model, able to handle larger classes of distributions for the outcome variable (Y_t) and allow a time series approach [1, 9, 50, 66].

In the case of our time series data, the use of Poisson regression assumes that the outcome variable (the natural log of the expected disease count) can be modelled by a linear combination of weather covariates and other confounders, see Equation 2.1 [66]. Log transformations are used for data that is positive and continuous (measured on an interval scale) [41]. Taking the natural logarithm transforms skewed data into a more symmetrical distribution [1]. For the number of disease counts in each week (t), $Y_t \sim \text{Poisson}(\mu_t)$ and

$$\ln(\mu_t) = \beta_0 + \beta_1 X_{1t} + ... + \beta_k X_{kt} \tag{2.1}$$

where $\ln(\mu_t)$ denotes the natural logarithm transformation, i.e. $\log_e$, of the expected number of disease count in week t; β_0 denotes the constant; β_k denote the predicted coefficients; and $X_{1t}, ..., X_{kt}$ denote weather covariates and other confounders in week t.

Other assumptions in Poisson regression are that the changes in the combined effects of the explanatory variables (weather and confounders) act multiplicatively on the rate of disease, the variance is equal to the mean number of cases at each level of the covariates and that each occurrence of disease is independent. When these assumptions are broken, overdispersion and an adjustment for correlation structure can be used to adjust for this, as described later in the chapter.

The addition of an offset term can be used to produce an outcome variable representing the rate that an event occurs. This proves to be particularly useful when modelling the counts of disease as it models the incidence risk of disease instead of the raw count, taking into account differing population sizes. The term $\log(PAR_t)$ is added to the right-hand side of the equation, producing model 2.2.

$$\ln(\mu_t) = \beta_0 + \beta_1 X_{1t} + ... + \beta_k X_{kt} + \log(PAR_t) \tag{2.2}$$

where PAR_t denotes the population at risk at time t.

Overdispersion

Overdispersion occurs when the assumption that the variance is equal to the mean number of cases at each level of the covariates is violated. The use of a quasi Poisson or negative binomial model can be used to allow for overdispersion [1, 4, 10, 38, 83]. The use of a negative binomial model was used to explore the association between cryptosporidiosis and average rainfall to evaporation ratio in New Zealand between 1965 and 2006 [15].

Lagged Values

The addition of lagged explanatory variables allows for the delayed effect of explanatory variables on the outcome variable (disease) at a time period [3, 7, 23, 58]. For example, this could allow for the lagged relationship identified with maximum temperature and campylobacteriosis [7, 58]. There are numerous ways of including the effects of lag into a Poisson regression, including the addition of moving averages or a geometrically distributed lag model [66].

The addition of moving averages would allow the contribution of multiple weeks of weather variables or other confounders to be estimated on the incidence of disease in New Zealand. A geometrically distributed lag model is described to be the best approach, as it allows the influence of each lag to decline with time [66]. This is achieved by fitting a polynomial function to the pattern of effects, shown in Equation 2.3, where the following is assumed:

$$\ln(\mu_t) = \text{covariates} + \beta(X_t + \alpha X_{t-1} + \alpha^2 X_{t-2} + \alpha^3 X_{t-3} + ...) \tag{2.3}$$

Seasonality

The diseases of interest in this thesis and weather variables exhibit seasonality [49, 82]. If unaccounted for, a spurious association might appear purely because both the outcome and explanatory variables are seasonal. The inclusion of season or month as a factor, or sine and cosine terms have been used to adjust for seasonality [13, 34, 46, 49, 82]. The most common inclusion of these terms are to the right hand side of the equation.

Correlation Structure

The Poisson regression model can also be extended to include spatial and/or temporal correlation adjustment terms. By not adjusting for correlation structure within the data, the Poisson distribution assumption that the occurrence of each case of disease is independent will be violated. Methods to adjust for spatial correlation include conditional autoregressive models (CAR) or intrinsic CAR models [11, 21]. The adjustment of temporal correlation includes methods such as the inclusion of lagged values, time-varying coefficients or transition models [44, 66, 82].

Time Series Poisson Model

A time series Poisson model was applied to cryptosporidiosis data from Brisbane [34]. It was assumed that the incidence of cryptosporidiosis had a Poisson distribution,

$$Y_t \sim \text{Poisson}(\mu_t); \mu_t = \exp(\beta_0 + \sum_{t=1}^{N} \beta_i X_{ti})$$

where Y_t is the number of cryptosporidiosis cases in month t; β_0 denotes the intercept; β_i denotes the effect of the weather covariates X_t on the number of cases of cryptosporidiosis during month t.

The impact of seasonality and trend was controlled for via decomposition of the incidence of cryptosporidiosis (Y_t) into seasonal (S_t) and trend (T_t) series (see Equation 2.4).

$$Y_t = T_t + S_t + E_t \tag{2.4}$$

where Y_t is the original series; T_t the trend series; S_t the seasonal series; and E_t the error term.

2.2.2 Time Series Analysis

A time series is a sequence of data points measured at successive periods of time, resulting in a longitudinal dataset. Time series analysis comprises of methods to extract meaningful statistics and forecasts from time series data. A popular form of time series analysis is an autoregressive moving-average (ARMA) model, denoted as an $ARMA(p,q)$ model, where p represents the order of autoregressive parameters and q represents the order of moving average parameters (Equation 2.5).

$$Y_t = \alpha + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{2.5}$$

where Y_t is the time series data of the count of disease in week t; α denotes the intercept; $\sum_{i=1}^{p} \varphi_i Y_{t-i}$ denotes the autoregressive term (AR) of order p; $\sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$ denotes the moving average term (MA) of order q; and ε_t denotes white noise [69].

These models assume the data are stationary, where the distribution of the count of disease in week (Y_t) is unconditional on other values of Y and assumed to be identical in all weeks (t). This means that the probability of a case disease should not vary with time, with no increasing or decreasing trend in the number of cases (i.e. the number of cases varies at a constant rate from the expected number). Stationarity is tested for in the residuals (ε_t) which should be identically and independently distributed (i.i.d), or in other words sampled from a normal distribution with zero mean and constant variance, i.e. $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

There are two common extensions to the ARMA model which adjust for non-stationarity, these include an integration term and/or seasonal term. Wavelet analysis is another method in which non-stationarity is characterized and estimated as a function of time [16].

The addition of an integration term uses the specified order of differencing to remove trends or other non-stationary behaviour as identified by an autocorrelation function (ACF) plot of the data. This model is named an autoregressive integrated-moving average (ARIMA) model, denoted $ARIMA(p, d, q)$, where p represents the number of autoregressive parameters, d represents the order of differencing and q represents the number of moving average parameters (Equation 2.6). The differencing operator of order d, Δ^d, is equal to $(1 - L)^d Y_t$. Here L is an operator that shifts Y_t back by one time period, e.g. for $d = 1$, $\Delta^d Y_t = (1 - L)^1 Y_t = Y_t - Y_{t-1}$.

$$\Delta^d Y_t = \alpha + \sum_{i=1}^{p} \varphi_i \Delta^d Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{2.6}$$

where $\Delta^d Y_t$ represents the differencing operator on the series Y_t; α denotes the intercept; $\sum_{i=1}^{p} \varphi_i \Delta^d Y_{t-i}$ denotes the autoregressive term (AR) of order p; $\sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$ denotes the moving average term (MA) of order q; and ε_t denotes white noise, assumed to be i.i.d.

Alternatively, the ARIMA model can be described using the following notation,

$$\phi(L)\Delta^d Y_t = \theta(L)\varepsilon_t$$

where $\phi(L)$ denotes the autoregressive term (AR) of order p; $\theta(L)$ denotes the moving average term (MA) of order q; and ε_t denotes white noise, assumed to be i.i.d.

Seasonality

Seasonality can be adjusted for via seasonal differencing, for s observations per period, denoted by a $sARIMA(p, d, q) \times (P, D, Q)_s$ model (Equation 2.7). In the equation, P represents the seasonal autoregressive term ($\Phi_s(L)$) of order P, D represents the number of seasonal differences, and Q represents the seasonal moving-average term ($\Theta_s(L)$) of order Q. The seasonal differencing operator of order D, Δ_s^D, is equal to $(1 - L^s)^d Y_t$. Here L is an operator that shifts Y_t back by one seasonal period, e.g. for $D = 12$ representing 12 months, $\Delta_s^D Y_t = (1 - L^{12})Y_t = Y_t - Y_{t-12}$.

$$\Phi(L)\Delta^d \Phi_s(L)\Delta_s^D Y_t = \Theta(L)\Theta_s(L)\epsilon_t \tag{2.7}$$

where

$$\Phi_s(L) = 1 - \phi_{s,1}L^s - \phi_{s,2}L^{2s} - ... - \phi_{s,P}L^{Ps}$$

and

$$\Theta_s(L) = 1 + \theta_{s,1}L^s + \theta_{s,2}L^{2s} + ... + \theta_{s,Q}L^{Qs}$$

Here Δ_s^D denotes the order of seasonal differencing and ϵ_t denotes the error term, assumed to be white noise (i.e. i.i.d).

Forecasting

Forecasting is frequently carried out from stationary ARMA models, or derivatives of the model. In a study on the effect of climate variation on salmonellosis transmission in Adelaide, South Australia, standard Poisson regression and sARIMA models were compared, along with other models [84]. The researchers concluded that the sARIMA model was better in both goodness-of-fit and forecasting ability, over the other regression models applied (tested using the mean standard error). This forecasting ability was echoed in a paper on modelling of weather variability and the incidence of cryptosporidiosis in Brisbane from January 1996 to December 2004 [34]. Once again, the researchers concluded that the sARIMA model had a better predictive ability than the Poisson regression model, for up to three years. The study used two study periods: period one spanned January 1996 - December 2003; period two spanned January to December 2004. Study period one was used to produce each model and period two was then used to validate the forecasting ability of the models, using the AIC to validate the goodness of fit. Further data was used to validate the models for 2 and 3 year periods.

For more information and discussion on time series analysis and forecasting, please refer to DeGooijer and Hyndman (2006) [22] and Fisman (2007) [26].

2.2.3 Hierarchical Bayesian Model for Space-Time Data

Bayes' Theorem is a way of manipulating conditional probabilities allowing new information to update a prior system of beliefs. This relationship is shown in Equation 2.8, where the conditional probability of the event y given the event x is given by

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{2.8}$$

Bayesian analysis is a form of parameter estimation where the model uses both prior distributions and likelihood functions to produce posterior probability distributions. It is based on the theory that the addition of prior beliefs with new information provided by emerging research produces new and improved beliefs [28]. These posterior distributions can be used in hypothesis testing since they can be interpreted to give the probability that the true value of a parameter is greater than (or less than) zero.

A hierarchical Bayesian framework decomposes highly dimensional and complex data into a series of simpler conditional levels [18]. As with Bayesian analysis, it uses both prior and posterior probability distributions in a hierarchical fashion [28]. These models are especially useful in situations where there is both a spatial and temporal correlation structure within the data. The use of hierarchical models allow aspects within the distribution of grid i to be estimated through the use of hyper-prior distributions. Information for the hyper-prior distributions are obtained through prior knowledge or expert opinion.

Markov chain Monte Carlo (MCMC) algorithms are used to estimate the posterior distribution of the parameters of interest in Bayesian analysis, using random numbers to generate a sample that is conditioned on the observed data. The algorithm produces a Markov chain of random variables $(X_0; ...; X_n)$ that converge to the posterior distribution. The number of iterations required for the sequence to converge to a stationary distribution with an acceptable error is the main problem in MCMC updating.

Knorr-Held Richardson Model

In the past, space-time data was analysed using models initially developed for just spatial or just temporal data [35]. However, a Bayesian hierarchical model was developed by Leonhard Knorr-Held and Sylvia Richardson to analyse space-time surveillance data on meningococcal disease incidence in France [45]. This model contains latent parameters that capture temporal, seasonal and spatial trends in disease incidence for endemic periods, by decomposing the log-relative-risks of disease ($\log(\lambda_{it})$) into three components, as seen in Equation 2.9.

$$\log(\lambda_{it}) = r_t + s_t + u_i \tag{2.9}$$

where λ_{it} is assumed to be Poisson distributed with mean $e_{it}\lambda_{it}$, where e_{it} is the number of expected cases; r_t denotes the trend component; s_t denotes the seasonal component; and u_i denotes the spatial component.

As described by Knorr-Held and Richardson, Gaussian Markov random-field prior distributions were used for all three components, where within the Bayesian framework, the temporal components (s_t and r_t) used Markov processes of arbitrary order with Gaussian increments. A gamma distributed hyperprior was assigned to the precision parameter of each component (κ_r, κ_s and κ_u), estimated from the French meningococcal disease data.

The trend component (r_t) assumed a Gaussian second-order random walk prior (Equation 2.10), i.e. the change in log-relative-risk from week t to week $t+1$ will be similar to the change in risk from week $t-1$ to week t [52]. Knorr-Held and Richardson discuss that their model is only appropriate for short-term prediction, due to the cubic smoothing spline joint distribution for the trend component, r_t [45].

$$r_t|\text{past} = 2r_{t-1} - r_{t-2} + \varepsilon_t \tag{2.10}$$

where $\varepsilon_t \sim N(0, \frac{1}{\kappa_r})$ and assumed to be i.i.d; and flat priors assumed for r_{t-1} and r_{t-2}.

For the seasonal component (s_t), Knorr-Held and Richardson assumed the components add up to Gaussian white noise throughout 12 months, see Equation 2.11.

$$s_t|\text{past} \sim N(-s_{t-11} - ... - s_{t-1}, \frac{1}{\kappa_s}) \tag{2.11}$$

where κ_s is an unknown precision parameter.

In regards to the spatial component (u_i), Knorr-Held and Richardson assumed similar parameter values for u_i and u_j, where adjacent departments ($i \sim j$) are similar. The unknown precision parameter κ_u determines the strength of similarity between the departments. See Equation 2.12.

$$p(\mathbf{u}|\kappa_u) \propto \exp\{-\frac{\kappa_u}{2}\sum_{i\sim j}(u_i - u_j)^2\} \tag{2.12}$$

Predictions

The use of hierarchical models for prediction purposes have been explored and compared in numerous papers [18, 35, 45]. A hierarchical Bayesian model was used to forecast ozone levels in Lake Michigan during spring-summer 1999 [55]. The inclusion of forecasted meteorological data improved the predictions of ozone levels and the researchers concluded that the model was successful at predicting the onset of high ozone levels periods, but more work was required to accurately identify departures from these periods.

The predictive abilities between the use of a spatio-temporal Bayesian maximum entropy (BME) model and a transformed hierarchical Bayesian space-time interpolation model were compared in predicting precipitation in Pakistan during the monsoon period [35]. As described in the paper, BME utilises a highly informative prior distribution to describe physical knowledge about natural processes. The hierarchical Bayesian space-time interpolation model produces space-time predictions completely characterised by probability density functions as it utilises a hierarchical framework that takes into account the uncertainty of the mean and covariance models. The researchers concluded that the transformed hierarchical Bayesian space-time interpolation model provided more accuracy and lower prediction error compared to the BME model.

2.3 Visualisation of Data

The choice of visualisation technique used in presenting the projected percentage changes in IR of each indicator disease for the years 2015, 2040 and 2090, at the fine spatial level (11,491 VCS grids) will need to be carefully considered. Maps can enable researchers to generate epidemiologically plausible hypothesis for the mechanisms underlying a projected increase or decrease in percentage IR of disease for each region within New Zealand.

2.3.1 Disease Mapping

Disease mapping is a graphical method of showing spatial distributions of disease incidence within a population. It can be used to generate hypothesis about transmission, risk factors and disease aetiology [12, 33]. Multivariate disease mapping gains information from the joint distributions of diseases mapped for the same graphical region [39].

Disease mapping was used to investigate a link between environmental exposures to endocrine distributing chemicals and prostate cancer in UK [12]. The spatial variation of the rates of multiple diseases with common risk factors were analysed using joint disease mapping [33]. The researchers concluded that the joint modelling of oral, oesophagus, larynx and lung cancer in Germany was a useful and valuable extension compared to an individual analyses of each disease.

For a more detailed introduction and discussion of disease mapping, please see Lawson and Williams (2001) [47].

2.3.2 Visualisation Techniques

Visualisation techniques are required to describe the spatial distribution of the projections. The techniques used in this body of work were choropleth maps. A choropleth map is a thematic map used to visualise a statistical variable of interest spatially. A geographical area is dividing into levels of the statistical variable and the magnitude is indicated by the use of patterns or colour.

There are many issues that can arise when using choropleth maps. These include the use of raw data values (such as population) rather than mapping the density (achieved by normalisation of data) or the allocation of colour shading scales to highly skewed data. Another issue occurs when the colours are similar in hue or intensity, causing the reader's perception of the value represented in one area to be affected by the colours in the surrounding areas. This is compounded by large regions dominating the display, where boundaries selected may produce patterns that produce the illusion of an underlying spatial distribution that can result in misleading spatial conclusions of the attribute of interest.

For a more detailed introduction and discussion of visualisation techniques, please see Kardos (2005) [40].

2.4 Conclusions

Individually modelling the effects of past climate variation on the four indicator diseases will be complex and may require different modelling methods for each disease.

After careful deliberation, there were reservations experienced with applying an sARIMA model to our data, as the methods of judging non-stationarity with Poisson distributed data that is known to be both spatially and temporally correlated were unknown to the modelling team. The GLM framework appeared to be a more appropriate method for the data format we were provided.

The methods for identifying collinearity between weather variables are well documented and readily available. The seasonality of both climatic and disease variables have been adjusted for in multiple studies, including the addition of sine and cosine terms or seasonal factors included in the model.

The spatial and temporal correlation will require careful attention, with the hierarchical Bayesian model developed by Leonhard Knorr-Held and Sylvia Richardson [45] a useful starting block to model the past association of climate on disease in New Zealand.

Chapter 3

Data

3.1 Introduction

The data sourced for this work was provided by ESR through the HAIFA project which is funded by the Foundation for Research, Science and Technology. The organisations that provided the data and format in which it was provided is outlined below.

3.2 Weather Data

A stepwise process was carried out by National Institute of Water and Atmospheric Research Ltd (NIWA) to prepare the weather data.

1. The actual weather measurements were recorded at climate stations located throughout New Zealand (approximately 500 sites recorded rainfall, 150 sites recorded temperature and humidity, 120 sites recorded wind and 100 sites recorded solar radiation). The data were collected by NIWA and MetService, and archived in the NIWA National Climate Database.

2. Data from the climate stations were then used to extrapolate daily measurements at 11,491 5km x 5km thiessen polygons grids, the centres of which are regarded as virtual climate stations (VCS), using a "second order derivative trivariate thin plate smoothing spline spatial interpolation model" [74]. The 11,491 VCS grids are shown in Figure 3.1, which was supplied by ESR.

3. Following this, ESR converted the daily weather measurements into weekly measurements, whereby the average weekly measurements were derived from the average of the average daily measurements; the minimum weekly measurements were derived from

the minimum of the minimum daily measurements; and the maximum weekly measurements were derived from the maximum of the maximum daily measurements. The weekly weather variables used in this body of work are shown in Table 3.1, consisting of weekly data for the eleven year study period (574 weeks) for 11,491 grids spanning New Zealand.

Weather Variable	Units	Description	Weekly Measurements	Mean (s.d.)
Rainfall	mm	24-hour rainfall from 9am	Minimum	0.0611 (0.559)
			Average	5.05 (7.64)
			Maximum	20.0 (28.3)
Rain Days	count	The number of days it rains in a week		2.54 (1.81)
Temperature	°C		Minimum	2.56 (4.68)
			Average	11.0 (4.50)
			Maximum	19.0 (5.15)
Wind Speed	m/s	Wind speed at 10m above ground level	Minimum	2.14 (1.10)
			Average	3.74 (1.53)
			Maximum	5.81 (2.37)
Absolute Humidity	kg/m^3		Minimum	0.00649 (0.00183)
			Average	0.00792 (0.00168)
			Maximum	0.00901 (0.00131)
Radiation	MJ/m^2	24-hour global solar radiation	Minimum	7.29 (5.01)
			Average	13.5 (6.65)
			Maximum	18.7 (8.59)

Table 3.1: Description of the weekly weather variables, their units, within variable measurements, and the mean and standard deviation from 1997 - 2007.

3.3 Disease Data

The final 18 diseases considered as possible indicator diseases, representative of the current behaviour and burden of disease on the population, were derived from an original list of 52 notifiable diseases. Notifiable diseases are diseases requiring notification by medical practitioners to their local Medical Officer of Health, under the Health Act 1956 and the Tuberculosis Act 1948 (for current list of notifiable diseases see `http://www.moh.govt.nz/moh.nsf/wpg_index/About-notifiable+diseases`).

The first step in deriving the 18 diseases was to remove all "imported" diseases (e.g. dengue fever, malaria, Q Fever), leaving only diseases established in New Zealand. The imported diseases became a separate group investigated by other members of the HAIFA project [75]. The next step was to remove all vaccine preventable diseases (e.g. measles, mumps), sexually transmitted infections, blood borne diseases (e.g. AIDS), diseases with less than 10 notified cases per year (e.g. Creutzfeldt-Jakob disease, *Enterobacter sakazakii* invasive disease), and those newly added to the notification schedule (e.g. invasive pneumoccocal disease). Following this process 17 diseases were left. The final step involved input from the Advisory Panel on the HAIFA project, who recommended the inclusion of non-notifiable seasonal influenza hospitalisations as the eighteenth disease due to its public health significance. The Advisory Panel on the HAIFA project are representatives from various organizations throughout New Zealand, Australia and Europe whose role it was to ensure considerations were taken for the overall benefit of the New Zealand population within the project. Data for influenza hospitalisations was obtained from the New Zealand Health Information Service and consisted of cases of seasonal influenza recorded in hospital discharge data. Four of the 18 diseases were excluded from this analysis due to the absence of case specific data (norovirus, rotavirus, *Vibrio parahaemolyticus* and *Clostridium perfringens* food intoxication).

The 14 diseases selected for consideration as an indicator disease in the HAIFA project were: campylobacteriosis, cryptosporidiosis, giardiasis, influenza hospitalisations, legionellosis, leptospirosis, listeriosis, meningococcal disease, rheumatic fever, salmonellosis, shigellosis, tuberculosis, yersiniosis and verotoxin producing *Escherichia coli* (VTEC). The weekly count of each disease, both confirmed and probable cases, from 1997 through to 2007 inclusive throughout New Zealand, was obtained from the ESR EpiSurv system. The influenza hospitalisations data was provided by the New Zealand Health Information Service (NZHIS), originating from routine hospital discharge data. Only the notifications of each disease in which the individual was in New Zealand for the duration of the estimated incubation period were included in this analysis. In the situation where it was known that an individual was overseas during part of the incubation period, the case was omitted from the data set.

In work carried out by ESR, the weekly counts of disease (both confirmed and probable cases) were matched to the 11,491 VCS grids defined by NIWA, spanning the 11 year study period. Confirmed and probable cases for the 14 diseases were included to try and capture the true incidence of the diseases in New Zealand (including imported cases).

A summary of the process for the selection of diseases on the HAIFA project is summarised in Figure 3.2.

3.4 Demographic Data

The following demographic variables have been identified to possibly confound the relationship between climate variation and the IR of infectious disease in New Zealand, and may need to be adjusted for. Past research highlighted differences in some or all of the risk of indicator diseases corresponding to these variables, as outlined in Chapter 1.1.2.

3.4.1 Population at Risk, Age and Gender

Demographic data on the count of the population at risk (PAR), proportion of each age group and gender structures were provided by Statistics New Zealand for each area unit (AU) spanning New Zealand. The known population in each area unit from the New Zealand census taken every five years was linearly interpolated to produce data spanning the 11 year study period, in work carried out by ESR. The data for 1997 - 1998 was linearly interpolated from the 1996 Census year; 1999 - 2003 data was linearly interpolated from the 2001 Census year; and the 2004 - 2007 data was linearly interpolated from the 2006 Census year. The AU from Statistics New Zealand and VCS grids from NIWA were overlaid and intersected to assign each VCS grid the corresponding measurement for each variable, as seen in Figure 3.3 obtained from ESR. The count of PAR in each VCS grid in 2007 is shown in Figure 3.4.

3.4.2 Ethnicity

Ethnicity structures were also provided by Statistics New Zealand for each AU spanning New Zealand. The proportion in each VSC grid was linearly interpolated from the known population in each AU taken from the New Zealand census taken every five years, in work carried out by ESR. The ethnicity data for 1997 - 1998 was linearly interpolated from the 1996 Census year; 1999 - 2003 ethnicity data was linearly interpolated from the 2001 Census year; and the 2004 - 2007 ethnicity data was linearly interpolated from the 2006 Census year.

Ethnicity is defined as the ethnic group(s) people feel they identify with or feel they belong to [76]. It is self-perceived and not mutually exclusive as people can and do identify with more than one ethnicity. The five groups provided were as follows:

- European or Other Ethnicity (including New Zealander)
- Maori
- Pacific Peoples
- Asian
- Middle Eastern/Latin American/African.

The ethnic groups that people identify with can change over time. For this reason, caution must be exercised when interpreting results involving ethnicity [76]. A Classification Code Hierarchy that specifies which ethnic group falls under which of the five ethnic groups provided can be located from the Statistics New Zealand website (`http://www.stats.govt.nz/`).

3.4.3 District Health Board

There are twenty-one District Health Boards (DHB) throughout New Zealand. Information on DHBs was provided from the New Zealand Ministry of Health. DHBs have existed in New Zealand since 2001, and did not change from 2001 to 2006. For this reason, ESR used the boundaries from the 2006 DHB data to identify the DHB each VCS grid belonged to. For VCS grids with multiple DHBs, the VCS grid was assigned to the DHB which occupies the maximum area within the grid. A list of the New Zealand DHBs is supplied in Table 3.2.

DHB Code	DHB Name
1	Northland
2	Waitemata
3	Auckland
4	Counties Manukau
5	Waikato
6	Lakes
7	Bay of Plenty
8	Tairawhiti
9	Taranaki
10	Hawke's Bay
11	Whanganui
12	Mid Central
13	Hutt Valley
14	Capital & Coast
15	Wairarapa
16	Nelson-Malborough
17	West Coast
18	Canterbury
19	South Canterbury
20	Otago
21	Southland

Table 3.2: List of New Zealand District Health Boards.

3.4.4 Social Deprivation Index

Social Deprivation Index (SDI) is a measure of deprivation on an ordinal scale from 1 – 10, developed at the Department of Public Health at the University of Otago, Wellington. A deprivation score of 1 represents least deprived, where a deprivation score of 10 represents most deprived. The original data was provided to ESR at the meshblock level and deprivation scores were allocated to the grids used in this work using a weighted average of New Zealand Deprivation score values, as suggested in the NZDep2006 Manual [80]. The weighted average for each grid-week (it) spanning the 11 year study period, was assigned a NZDep score from the closest census year using equation 3.1. The grid-weeks in 1996 - 1998 were assigned the

NZDep grid score from the 1996 census; the 1999 - 2003 grid-weeks were assigned the NZDep grid score from the 2001 census; and the 2004 - 2007 grid-weeks were assigned the NZDep grid score from the 2006 census.

$$WeightedAverage_{it} = \frac{\sum NZDepscore_{it} \times PopulationNumber_{it}}{\sum PopulationNumber_{it}} \tag{3.1}$$

ESR then calculated the percentile of the Deprivation score and used it to create the ordinal score of Deprivation from 1 to 10 for each VCS grid. VCS grids which were predominately oceanic, sea inlets or river estuaries and contain very few people in total were omitted from the index and given a deprivation score of 0.

A summary of the weekly demographic variables is supplied in Table 3.3.

Demographic Variable	Description	Measurements
Age	Proportion of those less than 4 years of age and greater than 65 years in each grid.	Proportion
District Health Board (DHB)	The DHB each grid falls into.	Index
Ethnicity	Those who identify with being 'European or Other Ethnicity (including New Zealander)', Maori', 'Pacific', 'Asian' and/or 'Middle Eastern /Latern American/African' in each grid.	Proportion
Gender	Proportion male and female in each grid.	Proportion
Population at Risk (PAR)	The count of the population residing in each grid.	Count
Social Deprivation Index (SDI)	The social deprivation of each grid on a scale of 0 - 10.	Index

Table 3.3: Description of demographic variables.

3.5 Environmental Data

The following environmental variables have been identified to possibly confound the relationship between climate variation and the IR of infectious disease in New Zealand, and may need to be adjusted for. Past research highlighted variables associated with these environmental confounders as risk factors for campylobacteriosis and cryptosporidiosis transmission, as outlined in Chapter 1.1.2.

3.5.1 Animal Density

Animal density data was originally supplied by Agribase at the farm address level. By allocating each farms centroid to the VCS grids used in this work, Massey University calculated the count of each animal per grid of dairy cows, beef and sheep, producing the overall count of animals per VCS grid. The density of dairy cows, beef and sheep in each VCS grid was then calculated by dividing the count by the area (km^2) of the corresponding grid.

3.5.2 Land Use

ESR obtained land use data from the New Zealand Ministry for the Environment. The land use data was extracted from LCBD2 data collected in summer 2001/2002 and was assumed to be consistent over the study period from 1997 to 2007. The original LCBD2 classifications were then regrouped into eight categories listed in Table 3.4, supplied as the percentage of each type of land use in each VCS grid.

3.5.3 Drinking Water Quality

Information on drinking water quality was produced from yearly drinking water quality zone data obtained from the ESR Water Programme. ESR used both the zone code and protozoa compliance to construct a scoring system for drinking water quality in each grid. A VCS grid score of 0 indicates good drinking water quality (complied); a VCS grid score of 1 indicates intermediate drinking water quality (inadequately monitored); a VCS grid score of 2 indicates poor drinking water quality (non-compliant and either not monitored or contained *E.coli*); and a VCS grid score of 3 indicates the drinking water quality is unknown.

3.5.4 Rurality

The rurality of each VCS grid was identified by ESR using data from Statistics New Zealand. Grids with the scores ‘main urban areas’, ‘satellite urban areas’ or ‘independent urban areas’ were classified as ‘Urban’. Grids with the scores ‘rural areas with high urban influence’, ‘rural areas with moderate urban influence’, ‘rural areas with low urban influence’ and ‘highly rural/remote areas’ were classified as ‘Rural’. VCS grids with the classification ‘area outside urban/rural profile’ were mainly grids that comprised of inland water or water inlets and were classified as either rural or urban according to the classification of their immediate neighbours.

A summary of the weekly environmental variables provided is supplied in Table 3.5.

3.6 Other Variables

3.6.1 Influenza Vaccination Coverage

Influenza vaccination coverage was provided to ESR by a variety of sources including the Ministry of Health, ESR and Dr. Lance Jennings at Canterbury Health Laboratories, Christchurch. Each data provider had different formats for the vaccination coverage. For the study years 1997 - 2003, the total vaccination coverage throughout New Zealand was provided, resulting in VCS grids assigned the same vaccination rate based on the total New

Zealand population. In 2004 - 2007, individual DHB data was available and the VCS grids were assigned accordingly. The data was then transformed into the vaccination coverage per 100,000 people, and each VCS grid assigned a rate proportional to the population at risk in each grid.

The influenza vaccination coverage has been identified to possibly confound the relationship between climate variation and the IR of infectious disease in New Zealand, and may need to be adjusted for. Vaccination is hoped to achieve a reduction in the number of susceptible individuals in a population, having a direct effect on the number of cases expected each year.

3.6.2 School Holidays

Information on school holiday dates from 1997 – 2007 inclusive, were collected from the New Zealand Ministry of Education. The variable 'holiday week' is classified as a week that contains at least one day of the week that is a holiday. For a week containing a holiday, 'holiday week' is given the value 1, otherwise the value 0.

The variable school holiday has been identified to possibly confound the relationship between climate variation and the IR of infectious disease in New Zealand, and may need to be adjusted for. In past research, outlined in Chapter 1.1.2, variables associated with recreational behaviour were highlighted as risk factors for the indicator diseases. These included contaminated swimming water and barbecuing for campylobacteriosis and cryptosporidiosis, and overcrowding for influenza hospitalisations and meningococcal disease.

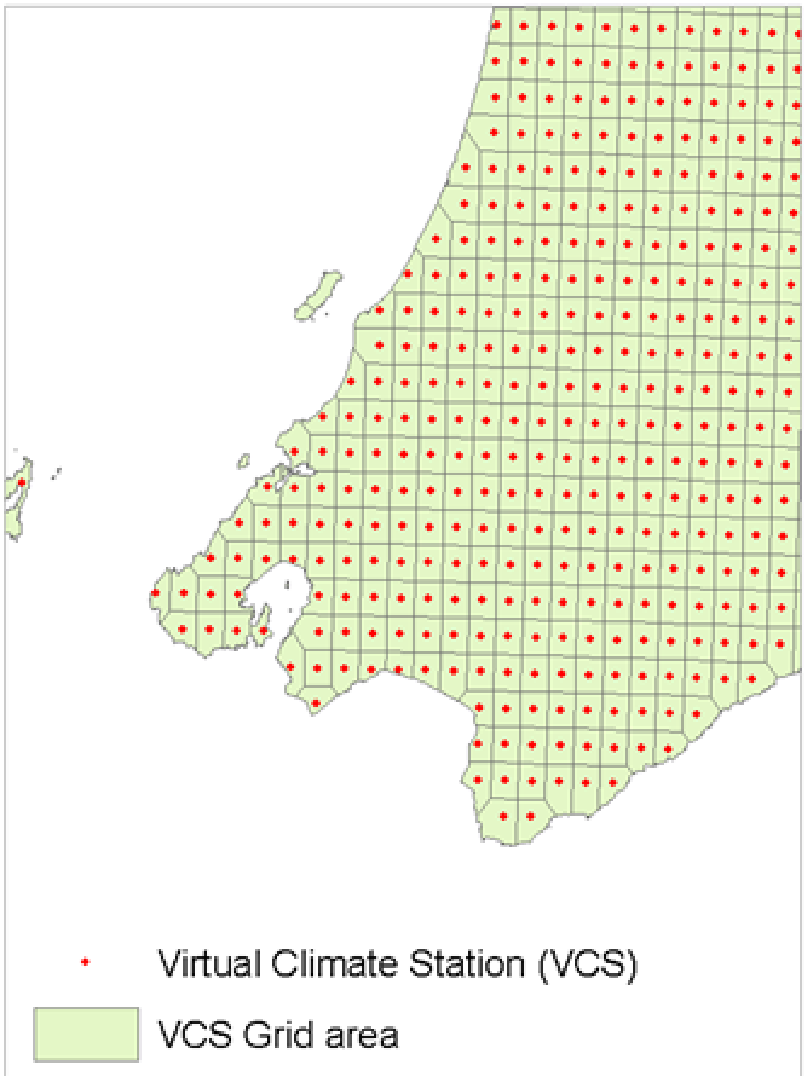


Figure 3.1: Extract of thiessen polygon grids based on the virtual climate stations (VCS).

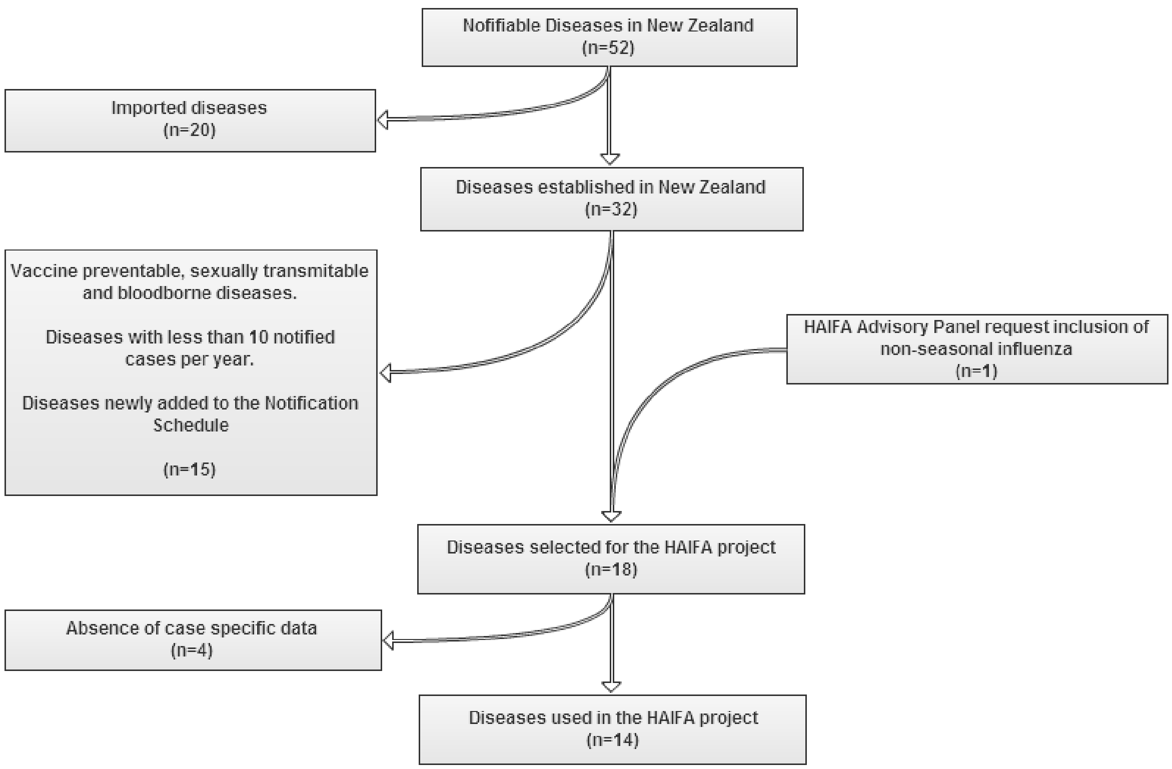


Figure 3.2: Flow Chart of the disease selection process used in the HAIFA project.

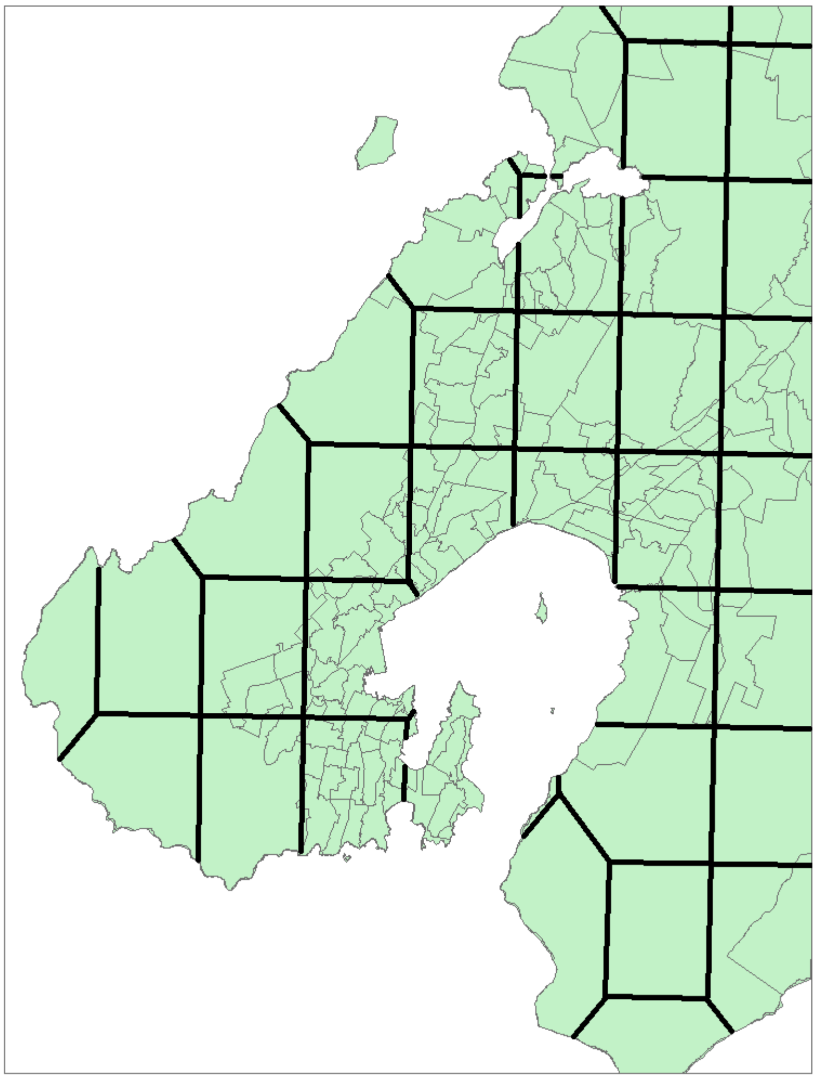


Figure 3.3: An example of virtual climate station (VCS) polygons (bold line) intersected with area units (AU) (fine line) supplied from Statistics New Zealand.

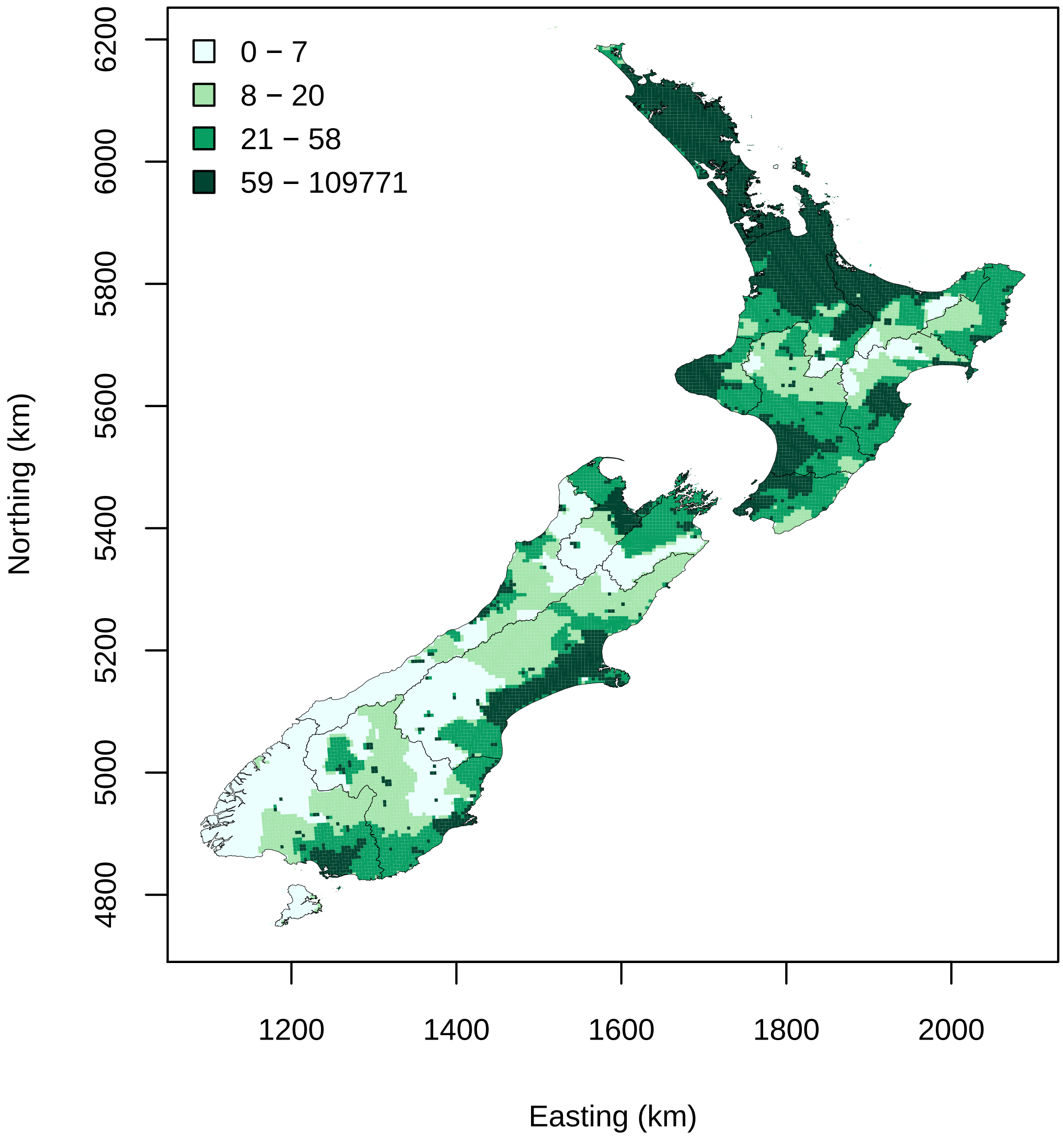


Figure 3.4: The count of the population at risk (PAR) in each virtual climate station (VCS) grid throughout New Zealand in 2007.

Land Use Index	Land Use Description	LCBD2 Class	LCBD2 Description
1	Artificial surfaces	1	Built-up Area
		2	Urban Parkland/ Open Space
		3	Surface Mine
		4	Dump
		5	Transport Infrastructure
2	Bare Surfaces	10	Coastal Sand and Gravel
		11	River and Lakeshore Gravel and Rock
		12	Landslide
		13	Alpine Gravel and Rock
		14	Permanent Snow and Ice
		15	Alpine Grass-/Herbfield
3	Water Bodies	20	Lake and Pond
		21	River
		22	Estuarine Open Water
4	Cropland	30	Short-rotation Cropland
		31	Vineyard
		32	Orchard and Other Perennial Crops
5	Grassland	40	High Producing Exotic Grassland
		41	Low Producing Grassland
		43	Tall Tussock Grassland
		44	Depleted Tussock Grassland
6	Sedge/Saltmarsh	45	Herbaceous Freshwater Vegetation
		46	Herbaceous Saline Vegetation
		47	Flaxland
		70	Mangrove
7	Scrub	50	Fernland
		51	Gorse and Broom
		52	Manuka and or Kanuka
		53	Matagouri
		54	Broadleaved Indigenous Hardwoods
		55	Sub Alpine Shrubland
		56	Mixed Exotic Shrubland
		57	Grey Scrub
8	Forest	60	Minor Shelterbelts
		61	Major Shelterbelts
		62	Afforestation (not imaged)
		63	Afforestation (imaged, post LCDB 1)
		64	Forest Harvested
		65	Pine Forest - Open Canopy
		66	Pine Forest - Closed Canopy
		67	Other Exotic Forest
		68	Deciduous Hardwoods
		69	Indigenous Forest

Table 3.4: A summary of the regrouping of LCBD2 land use definitions.

Environmental Variable	Description	Measure
Animal Density	Number of dairy cows, beef and sheep per km^2 in each grid.	Density
Land Use (LCBD2)	Majority land use in each grid: Artificial Surfaces (1), Bare Surfaces (2), Water Bodies (3), Cropland (4), Grassland (5), Sedge/Marsh (6), Scrub (7), Forrest (8).	Index
Drinking Water Quality	Drinking water quality in each grid: Good Quality (0), Intermediate quality (1), Poor Quality (2), Unknown (3).	Index
Rurality	Grid predominantly ubran or rural	Index

Table 3.5: A description of environmental variables.

Chapter 4

Selection of Indicator Diseases

4.1 Introduction

The purpose of this chapter is to summarise the selection of indicator diseases used in assessing the past and future impacts of climate on the incidence of influenza hospitalisations and notified infectious diseases in New Zealand. Firstly, the data and methodology used to assess the seasonality of disease incidence is described, followed by the method used for the selection of diseases to be taken forward into the modelling processes, in order to identify past associations with climate variability and project the incidence risk of disease in New Zealand. It was reasoned that diseases which exhibit seasonal patterns are more likely to be associated with weather variables than those that do not. A summary of the step-wise approach for selection of the four indicator diseases carried out in Chapter 3.3 and this chapter is shown in Figure 4.1.

4.2 Method

4.2.1 Data

For the process of indicator disease selection, data from three locations in New Zealand (Canterbury, Manawatu and Waikato) were used, selected based on their characteristics for each of the disease, climatic and demographic variables (Figure 4.2). The three locations were also selected to represent the north-to-south climate gradient present in New Zealand. All three study regions have a mixture of rural and urban areas, with the Canterbury study region selected as it is predominantly sheep farming; the Manawatu study region was chosen for its beef and dairy farming; and the Waikato study region was chosen as is predominantly dairy farming. The data was initially provided in the format of 81 (i.e. nine by nine)

0.05°latitude/longitude thessian grids, each surrounding a virtual climate station (VCS). The data comprised of weekly counts of each disease and weather measurements corresponding to each grid. The fine spatial resolution of the original data was not required for the selection of diseases hence the 81 grids in each location were amalgamated. The disease, weather and demographic time series in each VSC grid were amalgamated over the corresponding location (i) for each week (t), denoted by location-year-week (LYW).

All data manipulations and statistical analyses in this chapter were performed using R version 2.9.2 [61].

Weather Data

The weekly weather measurements were amalgamated by taking the average weather measurement over the corresponding location-year-week (LYW). The weekly weather variables used in this analysis are shown in Table 3.1.

Disease Data

The 14 diseases pre-selected for consideration as indicator diseases in the HAIFA project were: campylobacteriosis, cryptosporidiosis, giardiasis, influenza (hospitalisations), legionellosis, leptospirosis, listeriosis, meningococcal disease, rheumatic fever, salmonellosis, shigellosis, tuberculosis, yersiniosis and verotoxin producing *Escherichia coli* (VTEC) (see Chapter 3.3). The weekly count of each disease, both confirmed and probable cases, from 1997 through to 2007 inclusive throughout New Zealand were amalgamated by summing the disease counts over the corresponding location-year-week (LYW). The incidence risk (IR) of each disease was calculated in order to adjust for different population sizes (PAR_{it}) over the three locations as well as the growing population throughout New Zealand from 1997 - 2007. The following relationship was used to calculate the IR of disease for each location (i) and the average IR of disease throughout the three locations at week (t):

$$IR_{it} = \frac{DiseaseCount_{it}}{PAR_{it}} \tag{4.1}$$

Demographic Data

The population at risk (PAR) data was amalgamated by adding the populations for each study region over the location-year-week (LYW).

4.2.2 Exploratory Data Analysis of Diseases

The aim of this section was to explore the annual IR of each disease and any seasonal patterns to decide which diseases are to be included in the list of possible indicator diseases. The temporal trend was visually explored by inspecting the time series of the (untransformed) IR of each disease, using the loess smoothed [20] average IR over all three regions and loess smoothed IR stratified by study region. This stratification enables us to identify any differences in the IR of each disease between the study regions. For diseases that had an annual IR < 5 per 100,000 population over all three study regions, the time series could not effectively be loess smoothed or stratified by region, so were therefore analysed using the raw IR of disease throughout the three regions.

Raw and Smoothed Time Series

The combined weekly incidence risk (IR) of each disease over the three study regions of interest were presented as time-series graphs to visually explore trend and seasonality. Loess smoothing [20] was applied to the raw time series to emphasise the major features while reducing distraction from random variation [14]. The degree of smoothing was selected to emphasise seasonality. If there were sufficient cases (i.e. a combined annual IR ≥ 5 per 100,000) the time series were then stratified by location to investigate the spatial variation at a finer resolution (refer to Figures 4.6 - 4.16). In some of the loess smoothed plots, extreme positive and negative values are observed at the beginning and end of the time series and are a limitation of smoothing methods (e.g. Figures 4.7, 4.16).

Assessing Seasonality

To identify the presence of seasonality, box plots grouped by week were examined e.g. a summary of week 1 for all years of data. Following this, cyclicity was identified by plotting cumulative periodograms [19]. This involves fitting sinusoidal waves with a discrete set of frequencies (Fourier frequencies) to the time series of the average IR over the three regions of interest for each disease. A cycle is identified at the point the trace deviates outside the 95% confidence lines.

Disease Exclusion Criteria

To model the relationship between the weather variables and the IR of infectious diseases in future chapters, it was decided that the indicator diseases selection should not have sparse counts of disease and should exhibit seasonality. Therefore, in the three study regions, it was decided that if

1. the mean annual IR of disease was less than 5 per 100,000 population from 1997 - 2007 throughout all three locations, or

2. the disease showed no seasonal pattern in the IR of notification of disease throughout the year

the disease would be excluded from the list of possible indicator diseases to be used in the HAIFA project, and hence the rest of the work carried out in this thesis.

4.3 Results

4.3.1 Exploratory Data Analysis

Weather Data

A visual appraisal of the weather variables (average rainfall, temperature, wind speed, absolute humidity, radiation and the number of days it rained during each week) showed they all exhibited a cyclic pattern. The mean and standard deviation of each weather variable over the study period, 1997 - 2007, for each location is summarised in Table 4.1.

Weekly Weather Variable (Units)	Location	Mean (1997 - 2007)	Standard Deviation (1997 - 2007)
Average rainfall (mm)	Waikato	3.25	3.21
	Manawatu	2.81	2.87
	Canterbury	1.79	2.25
Average number of days of rain (count)	Waikato	2.92	1.71
	Manawatu	2.78	1.63
	Canterbury	2.03	1.32
Average temperature (°C)	Waikato	14.0	3.66
	Manawatu	13.0	3.61
	Canterbury	11.8	4.07
Average wind speed (m/s)	Waikato	2.82	0.801
	Manawatu	4.23	1.12
	Canterbury	3.50	0.790
Average absolute humidity (kg/m^3)	Waikato	0.00912	0.00116
	Manawatu	0.00862	0.00130
	Canterbury	0.00768	0.00164
Average radiation (MJ/m^2)	Waikato	14.0	5.97
	Manawatu	13.7	6.30
	Canterbury	13.4	6.74

Table 4.1: The average measurement and standard deviation of each weather variable from 1997 - 2007 in the Canterbury, Manawatu and Waikato study regions.

The average weekly rainfall in the Canterbury study region (1.79 mm) was found to be less than that in both the Waikato and the Manawatu study regions. The average weekly rainfall in the Waikato (3.25 mm) and Manawatu (2.81 mm) study regions were similar (Figure 4.3a). The Waikato study region tended to have the highest average rainfall, with a predominant peak during winter of 1998.

Over the 11 year period, Canterbury received the least average number of days of rain per week (2.03 days) of the three regions, while in general Waikato (2.92 days) received the most (Figure 4.3b). Therefore, with regard to average weekly rainfall and the number of days of rain per week, the Canterbury study region presents a different pattern, whereas the Waikato and Manawatu study regions behave more similarly.

There were spatial differences in relation to the average temperature over the three regions: we see that regardless of season, the Waikato region tended to have the highest average temperature (14.0 °C) followed by Manawatu (13.0 °C) and then Canterbury (11.8 °C) (Figure 4.4a). In general, the average temperature between regions differed by a couple of degrees.

Again, there was evidence of spatial differences between the three regions when we examined the average wind speed (Figure 4.4b). Manawatu was generally the windiest throughout the study period (4.23 m/s) while Waikato was the calmest (2.82 m/s). The average wind speed in each study region showed a slight downward trend throughout the 11 year period from 1997 to 2007.

The highest average absolute humidity throughout the three study regions occurred in summer and the lowest in winter (Figure 4.5a). Waikato was found to have the highest all year round average absolute humidity (0.00912 kg/m^3), followed by Manawatu (0.00862 kg/m^3) then Canterbury (0.00768 kg/m^3).

Spatial differences in the average radiation levels during summer and winter were observed over the 11 year period (Figure 4.5b).

Disease Data

The combined number of times each disease was notified and the mean annual IR, for the period of 1997 - 2007 inclusive, is shown in Table 4.2 for the Canterbury, Manawatu and Waikato study regions. The diseases with an annual average IR $<$ 5 per 100,000 population were VTEC (IR: 2.74 per 100,000), leptospirosis (IR: 1.76 per 100,000), legionellosis (IR: 1.55 per 100,000), shigellosis (IR: 1.03 per 100,000), rheumatic fever (IR: 1.00 per 100,000) and listeriosis (IR: 0.42 per 100,000).

Disease	Total Count 1997 – 2007	Mean Annual Incidence Risk per 100,000
Camylobacteriosis	18104	295.52
Salmonellosis	2166	35.36
Giardiasis	2052	33.50
Cryptosporidiosis	1582	25.82
Influenza	1071	17.48
Yersiniosis	660	10.77
Meningococcal Disease	418	6.82
Tuberculosis	343	5.60
Verotoxin Producing Escherichia coli (VTEC)	168	2.74
Leptospirosis	108	1.76
Legionellosis	95	1.55
Shigellosis	63	1.03
Rheumatic Fever	61	1.00
Listeriosis	26	0.42

Table 4.2: The combined disease counts and annual incidence risk (per 100,000) from 1997 - 2007 in the Canterbury, Manawatu and Waikato study regions.

Demographic Data

The population at risk (PAR) residing in each location interpolated from the census data collected every 5 years, changed over the years from 1997 - 2007. There was an increase in population size for each study region with the largest PAR found in the Canterbury region, followed by the Waikato and then Manawatu regions. In 1997, the Canterbury study region was estimated to have a population size of 239,980 people increasing to 277,236 in 2007. (Table 4.3)

4.3.2 Exploratory Data Analysis of Diseases

Raw and Smoothed Time Series

Campylobacteriosis

The time series of the average IR of campylobacteriosis over all three study regions showed a cyclic increase in average IR during the summer period of each year (Figure 4.6a). There were also peaks in the average IR during the winter of 1999, 2004 and 2006; however these were much smaller in magnitude in comparison to the summer peaks. There appeared to be a slight decrease in trend from 1997 - 2001, then an increase in the average IR, from 2001 - 2007. The summer peak in the IR of campylobacteriosis stratified by region was evident in all three study regions, however the study region in Canterbury appeared to have a delayed peak in 1998, 2000, 2002 and 2007 in comparison to the other regions (Figure 4.6b). The IR of campylobacteriosis in the Manawatu study region had winter seasonal peaks occurring in

Year	Population at Risk (PAR)			
	Canterbury	Manawatu	Hamilton	All Locations
1997	239980	108241	178124	526345
1998	241943	107994	180684	530620
1999	243394	107990	182724	534108
2000	245046	107708	184400	537155
2001	246846	107855	185601	540302
2002	252075	108649	189616	550339
2003	258469	109874	194202	562545
2004	264015	110790	198503	573307
2005	268344	111220	202144	581708
2006	273170	111895	205903	590967
2007	277236	112298	209127	598660
Average	255502	109501	191912	556914

Table 4.3: The population at risk (PAR) from 1997 - 2007 in the Canterbury, Manawatu and Waikato study regions, over the period 1997 - 2007.

1998, 1999, 2002 and 2004 - 2007. The IR in the Manawatu study region was consistently lower than the IR in both the Canterbury and Waikato study regions. The Waikato and Manawatu regions followed the trend shown in the average IR, with a slight decrease in trend from 1997 - 2001, and an increase in trend from 2001 - 2007. The Canterbury study region showed a fairly consistent IR of campylobacteriosis from 1997 - 2005, with an increase in IR from 2006 - 2007.

Cryptosporidiosis

The time series of the average IR of cryptosporidiosis over all three study regions showed a cyclic increase in average IR during autumn and spring each year, except in 2003, 2004 and 2006 where a peak in IR was only observed in spring (Figure 4.7a). There were larger increases in average IR of cryptosporidiosis during the spring of 1998, 2001 and 2003. There appeared to be no apparent increasing or decreasing trend in the average IR throughout the period 1997 - 2007 inclusive. The IR of cryptosporidiosis stratified by region showed the Canterbury study region had a lower IR of cryptosporidiosis throughout the period 1997 - 2007 in comparison to the Waikato and Manawatu study regions (Figure 4.7b). All three regions showed cyclicity representative of that seen in the average IR (Figure 4.7a). The Waikato region showed peaks in IR during spring each year, with additional peaks in IR during the autumn of 1998, 1999, 2005 and 2007. The Canterbury and Manawatu study regions showed similar temporal trends, however the spikes per year were larger in Manawatu than in Canterbury.

Giardiasis

The visual appraisal of the time series for the average IR of giardiasis over all three study regions showed seasonal peaks in late summer and spring of each year, except in 1997, 2006 and 2007 (Figure 4.8a). The overall trend appeared to gradually decrease from 1997 - 2007. The IR of giardiasis stratified by region showed similar seasonal patterns, with peaks in late summer and spring (Figure 4.8b). There appeared to be a slight negative parabolic trend in both the Manawatu and Canterbury study regions. The Waikato study region showed a steady decrease in IR from 1997 - 2007.

Influenza Hospitalisations

The time series of the average IR of influenza hospitalisations over all three study regions showed a seasonal peak in IR during winter from 1997 - 2007 (Figure 4.9a), with larger spikes in 1999, 2001 and 2002 compared to other years. There appeared to be no trend in the average IR from 1997 - 2007. The IR of influenza hospitalisations stratified by region showed that the Canterbury study region has seasonal peaks each year during winter, whereas the Waikato study region only has winter peaks in 1999, 2002, 2004 and 2006 (Figure 4.9b). The IR of influenza hospitalisations in the Manawatu study region did not appear to stray from approximately zero, which may be due to the smoothing used.

Legionellosis

The time series of the raw (unsmoothed) IR of legionellosis over all three study regions appeared to cluster in summer 1999 and 2001, then again in winter 2005 (Figure 4.10a). There did not appear to be a trend in the IR of legionellosis throughout the study period.

Leptospirosis

The time series of the raw (unsmoothed) IR of leptospirosis over all three study regions appeared to be consistent, with no apparent pattern or seasonality (Figure 4.10b). There did not appear to be a trend in the IR of leptospirosis from 1997 - 2007 inclusive.

Listeriosis

The time series of the raw (unsmoothed) IR of listeriosis over all three study regions indicate that cases of listeriosis were spurious with clusters of cases occurring in winter 1998, 1999, 2002 and 2007. Summer clusters were observed in 2000 - 2004 and 2006 (Figure 4.11a). There appeared to be an no visual trend in the average IR throughout the period 1997 - 2007 inclusive.

Meningococcal Disease

The time series of the average IR of meningococcal disease over all three study regions showed seasonal peaks in winter, with additional peaks in summer of 1998, 2001, 2005 and 2006 (Figure 4.12a). There appeared to be a general increase in IR trend from 1997 - 2001 followed by a decrease in trend from 2002 - 2007. A larger peak was observed in the winter of 2001. The IR of meningococcal disease stratified by study region showed the IR was highest in the Waikato region, followed by the Manawatu region then Canterbury (Figure 4.12b). The peaks in average meningococcal disease IR appear to be driven by the IR in the Waikato study region. The larger peak in 2001 was apparent in both the Waikato and Manawatu study regions. The IR in the Canterbury study region only deviated from approximately zero in winter 2007. The Waikato study region showed an increase in trend from 1997 - 2001 followed by a decrease in trend in the average IR from 2002 - 2007. There appeared to be no evident trend in the Canterbury and Manawatu study regions.

Rheumatic Fever

The time series of the raw (unsmoothed) IR of rheumatic fever over all three study regions showed no apparent seasonality where the IR indicates sporadic cases between 1997 and 2004, with very few cases during 2004 - 2007 (Figure 4.11b). There appeared to be no trend in the average IR throughout the period 1997 - 2007 inclusive.

Salmonellosis

The time series of the average IR of salmonellosis over all three study regions showed summer peaks each year (Figure 4.13a). Smaller peaks were present in the spring of 2002 - 2006 inclusive. There appeared to be a decreasing trend in the average IR from 1997 - 2007. The IR of salmonellosis stratified by region showed very similar behaviour throughout the three study regions of interest (Figure 4.13b). The decreasing trend observed in the average IR was also represented when stratified by region.

Shigellosis

The raw (unsmoothed) time series of the IR of shigellosis over all three study regions showed a sporadic IR with no visible cyclicity. There were two larger peaks in summer and spring 2004 (Figure 4.14a). There appeared to be no evident trend in the average IR from 1997 - 2007.

Tuberculosis

The visual appraisal of the time series of the raw (unsmoothed) IR of tuberculosis over all three study regions showed summer and winter peaks consistently throughout the 11-year period (Figure 4.14b). The loess smoothed time series produced a flat line at an IR of zero, indicating that there was possibly no seasonal pattern exhibited in the time series of the IR. There appeared to be no trend in the average IR from 1997 - 2004, however there appears to be an increasing trend from 2004 - 2007.

Verotoxin producing *Escherichia coli* (VTEC)

The time series of the raw (unsmoothed) IR of VTEC over all three study regions indicated sporadic cases with no visible cyclicity or trend throughout the period 1997 - 2007 inclusive (Figure 4.15).

Yersiniosis

The time series of the average IR of yersiniosis over all three study regions showed seasonal peaks in IR throughout the summer period of each year (Figure 4.16a). There appeared to be a slight decrease in trend in the average IR from 1997 - 2005, followed by a slight increase in trend from 2005 - 2007. The IR of yersiniosis stratified by region was higher in the Canterbury study region, followed by the Waikato region then Manawatu (Figure 4.16b). The Canterbury and Waikato study regions showed seasonality as seen in the average IR. The Manawatu study region appeared to have no deviation above zero IR.

Assessing Seasonality

The cumulative periodograms for each disease exhibiting seasonality, over all three locations, are shown in Figure 4.17. An annual cycle was present in campylobacteriosis, legionellosis, meningococcal disease and yersiniosis. The diseases found to show two cycles per year were cryptosporidiosis, influenza (hospitalisations), salmonellosis and shigellosis. The diseases that did not have any significant cyclicity were found to be giardiasis, leptospirosis, listeriosis, rheumatic fever, tuberculosis and VTEC. The evidence of one or two cycles per year in the IR of disease indicates the possible presence of a seasonal or climatic effect on the IR of disease throughout the year. The diseases that do not exhibit a significant cyclicity will not be taken through into the modelling chapters.

Selection of Diseases for Further Investigation

The diseases that failed to meet both requirements outlined in section 4.2.1 (i.e. 1. the IR of disease is less than 5 per 100,000 over the 11 years in all three locations, and 2. the disease showed no seasonal aspect in IR throughout the year) were legionellosis, leptospirosis, listeriosis, rheumatic fever, shigellosis, tuberculosis and VTEC. The remaining seven diseases identified as potential indicator diseases were campylobacteriosis, cryptosporidiosis, giardiasis, influenza hospitalisations, meningococcal disease, salmonellosis and yersiniosis.

4.4 Summary

Our results indicate that the potential indicator diseases for the effect of climate change on infectious diseases in humans could be any of the seven diseases identified during the disease selection process. Four of the seven possible indicator diseases concur with the indicator diseases endorsed by the Advisory Panel. Hence the following diseases were chosen as indicator diseases for the HAIFA project and to be used in subsequent chapters: campylobacteriosis, cryptosporidiosis, influenza hospitalisations and meningococcal disease.

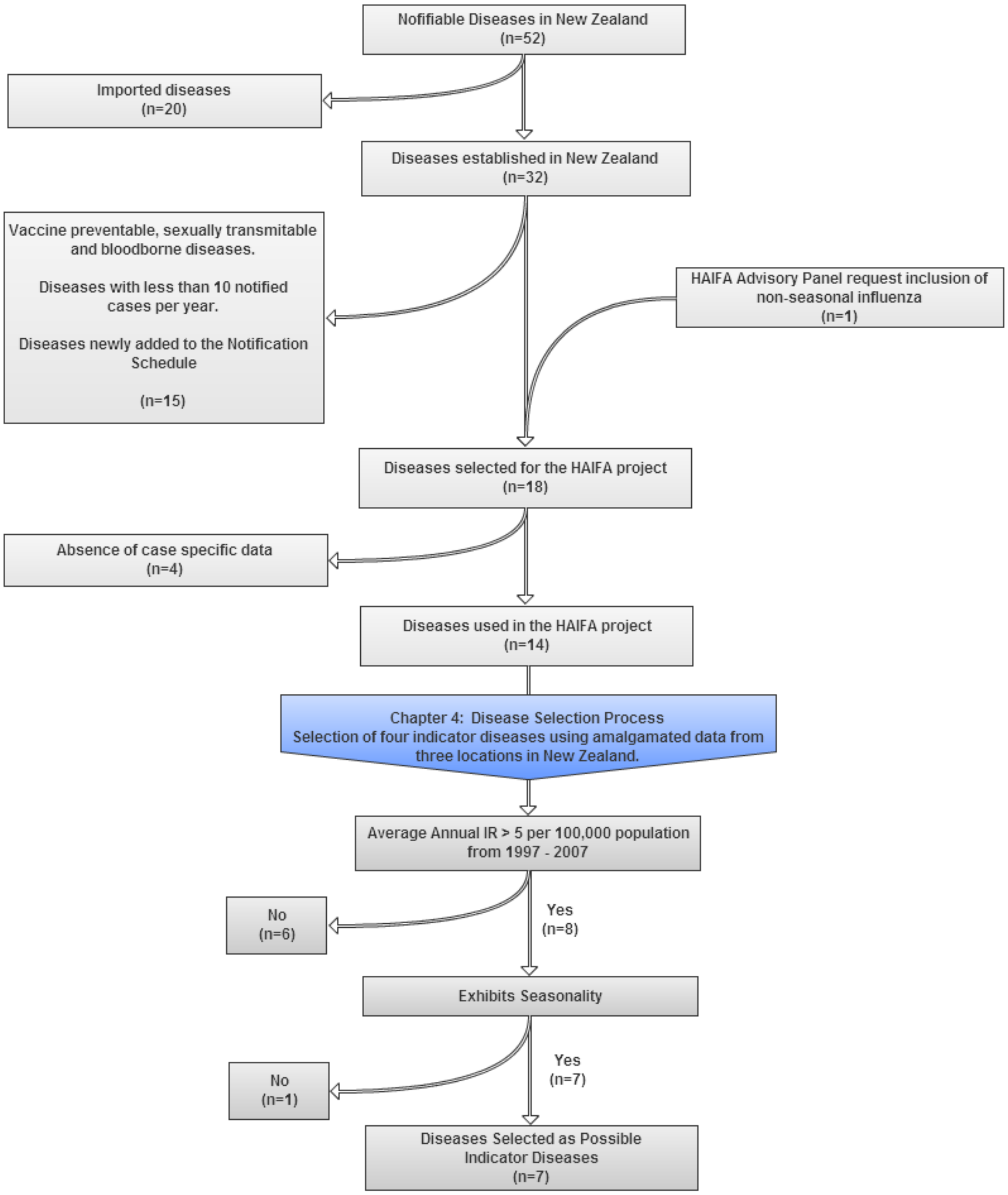


Figure 4.1: An overview of the disease selection process.

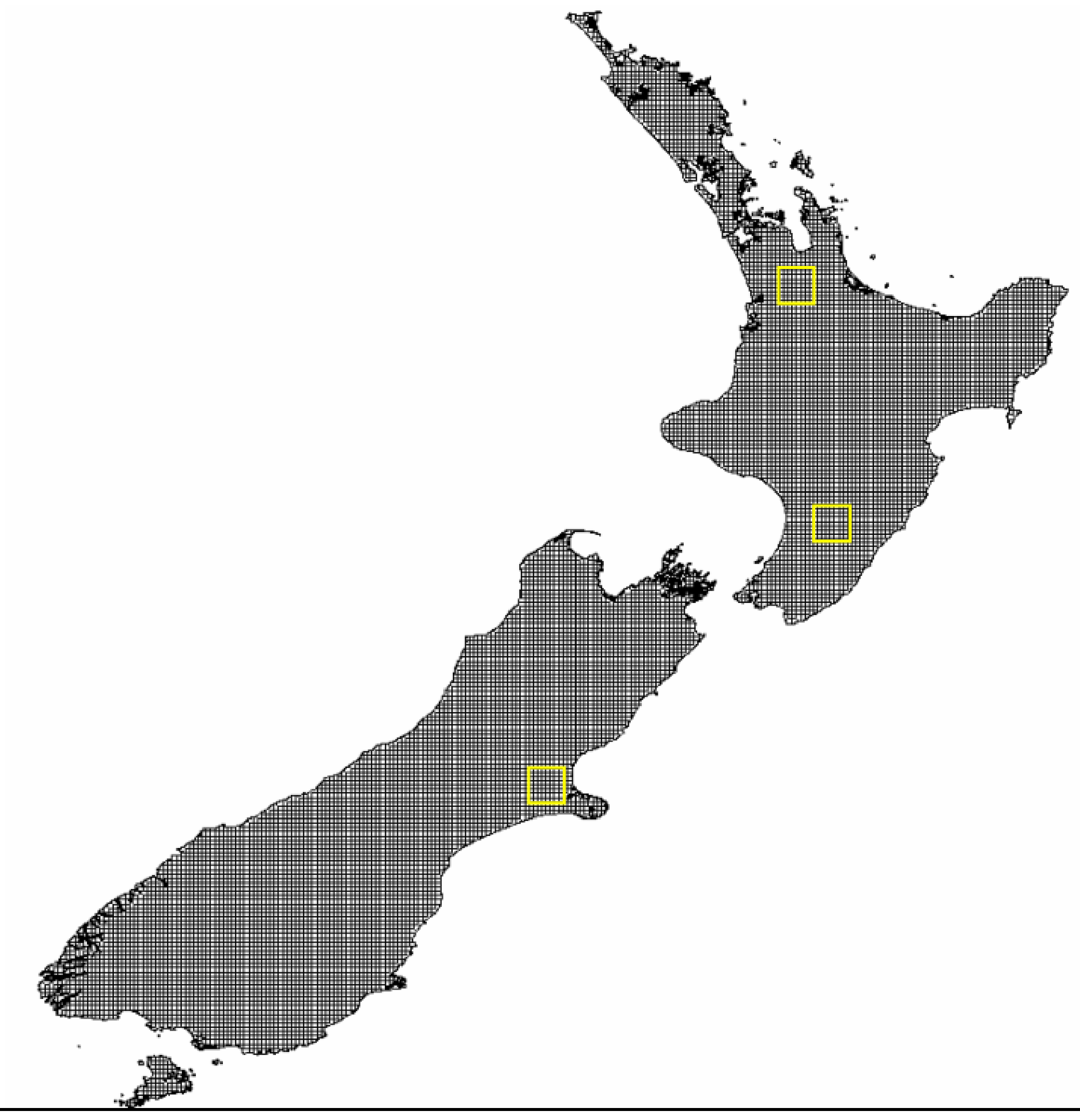


Figure 4.2: Map of the three study regions in New Zealand used during the indicator disease selection process. From top to bottom: Waikato, Manawatu and Canterbury.

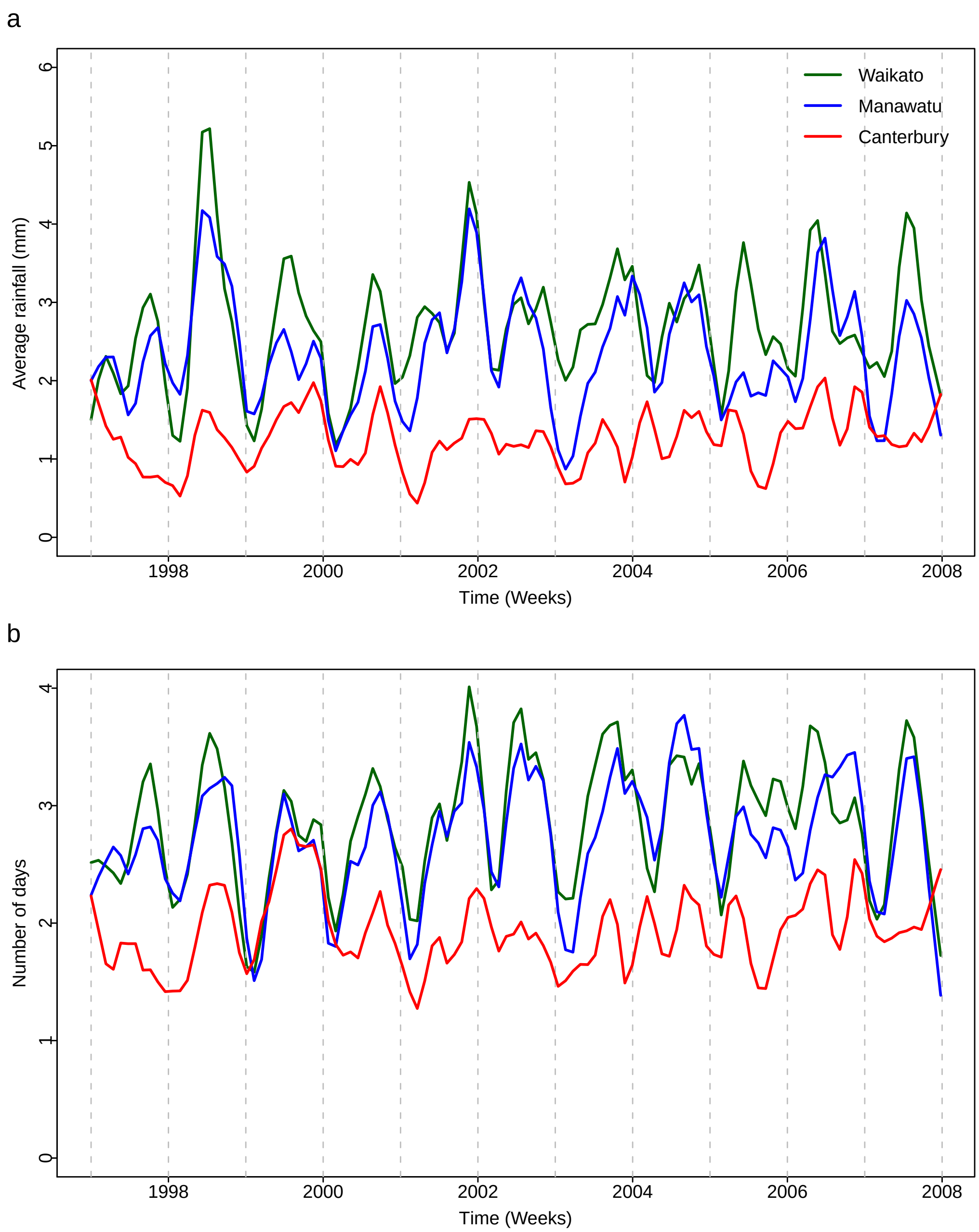


Figure 4.3: Loess smoothed time-series of a) the average weekly rainfall and b) average number of rainy days per week in the Canterbury, Manawatu and Waikato study regions, from 1997 - 2007.

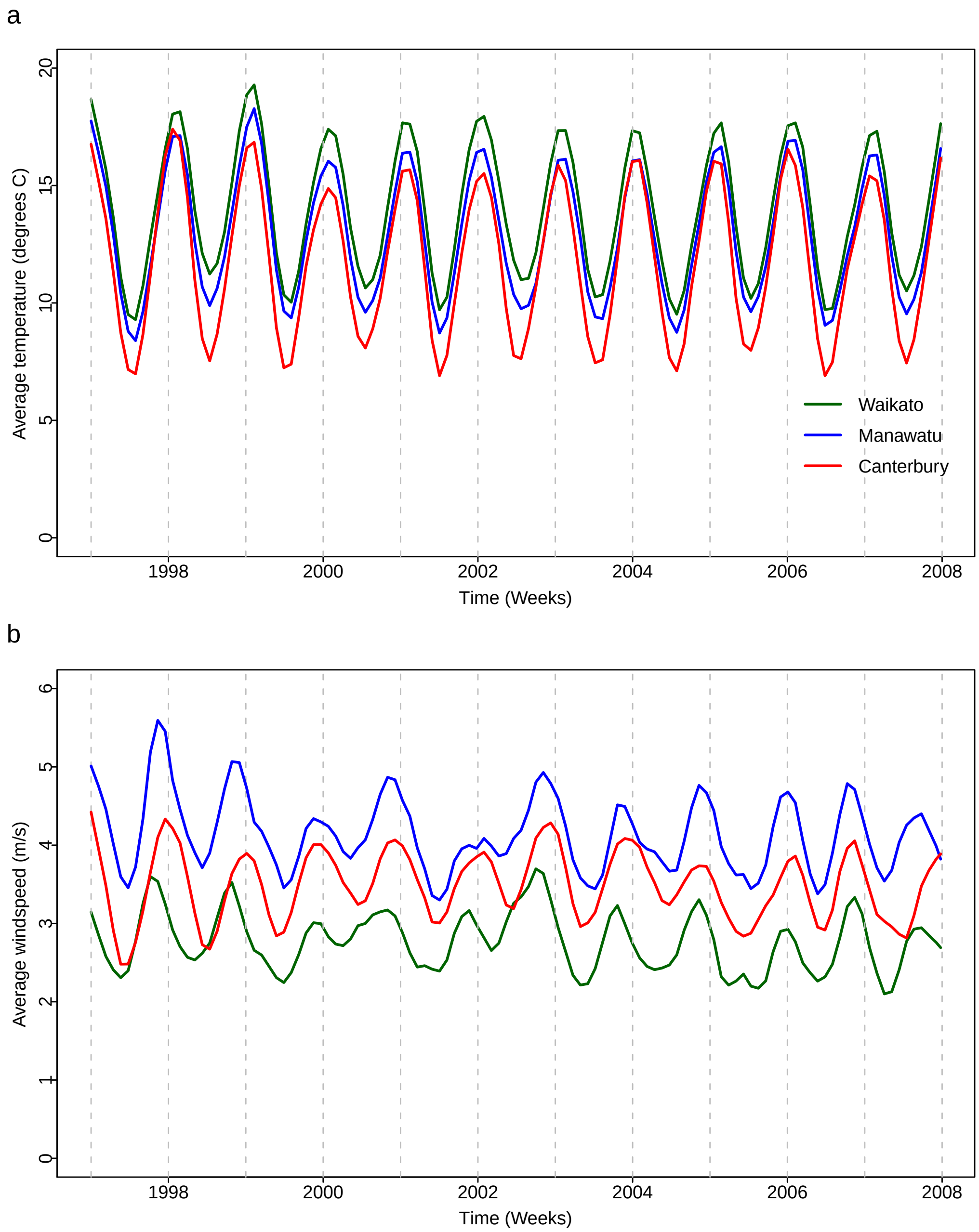


Figure 4.4: Loess smoothed time-series of a) the average weekly temperature and b) wind speed in the Canterbury, Manawatu and Waikato study regions, from 1997 - 2007.

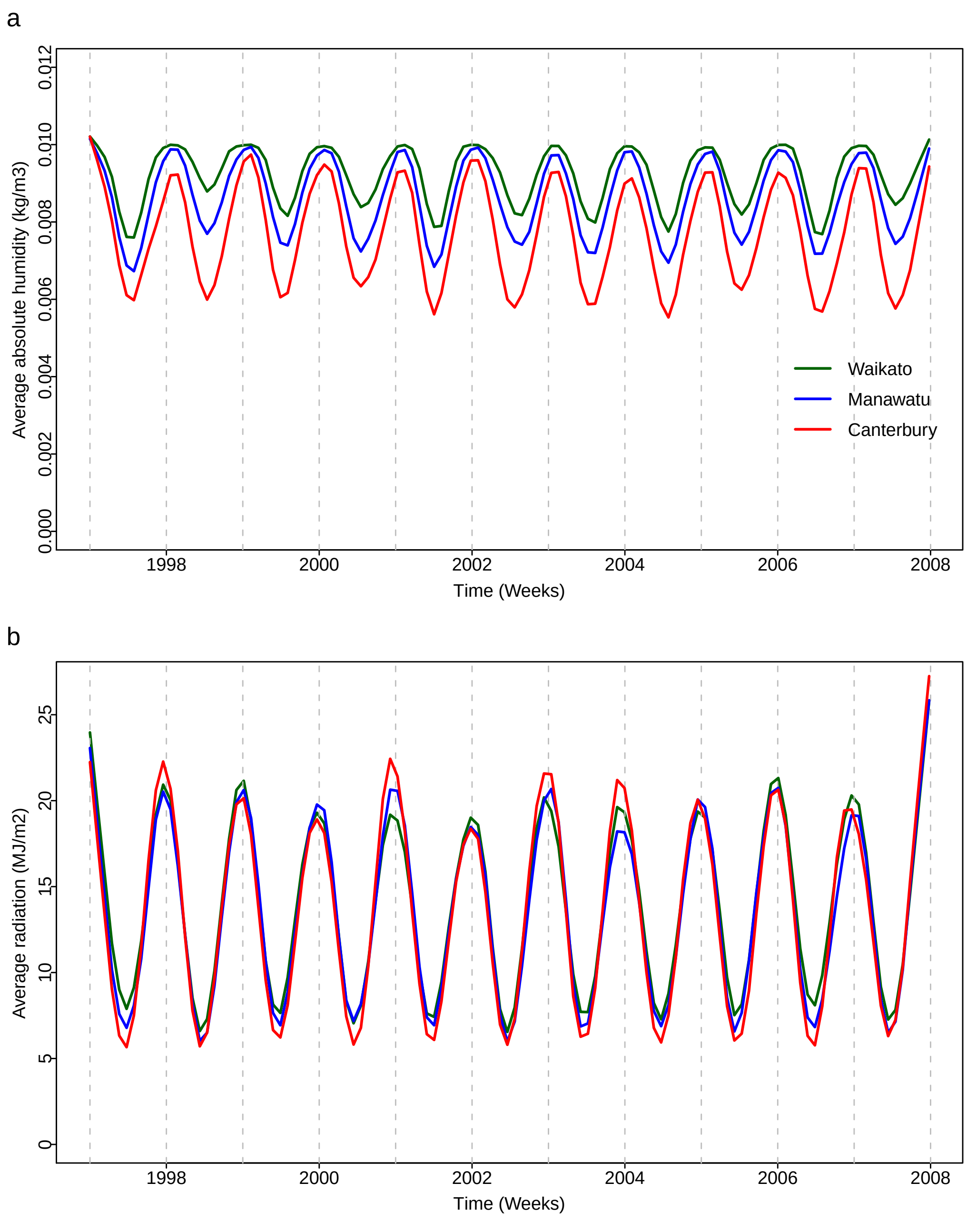


Figure 4.5: Time-series of a) the average absolute humidity and b) the average weekly radiation levels in the Canterbury, Manawatu and Waikato study regions, from 1997 - 2007.

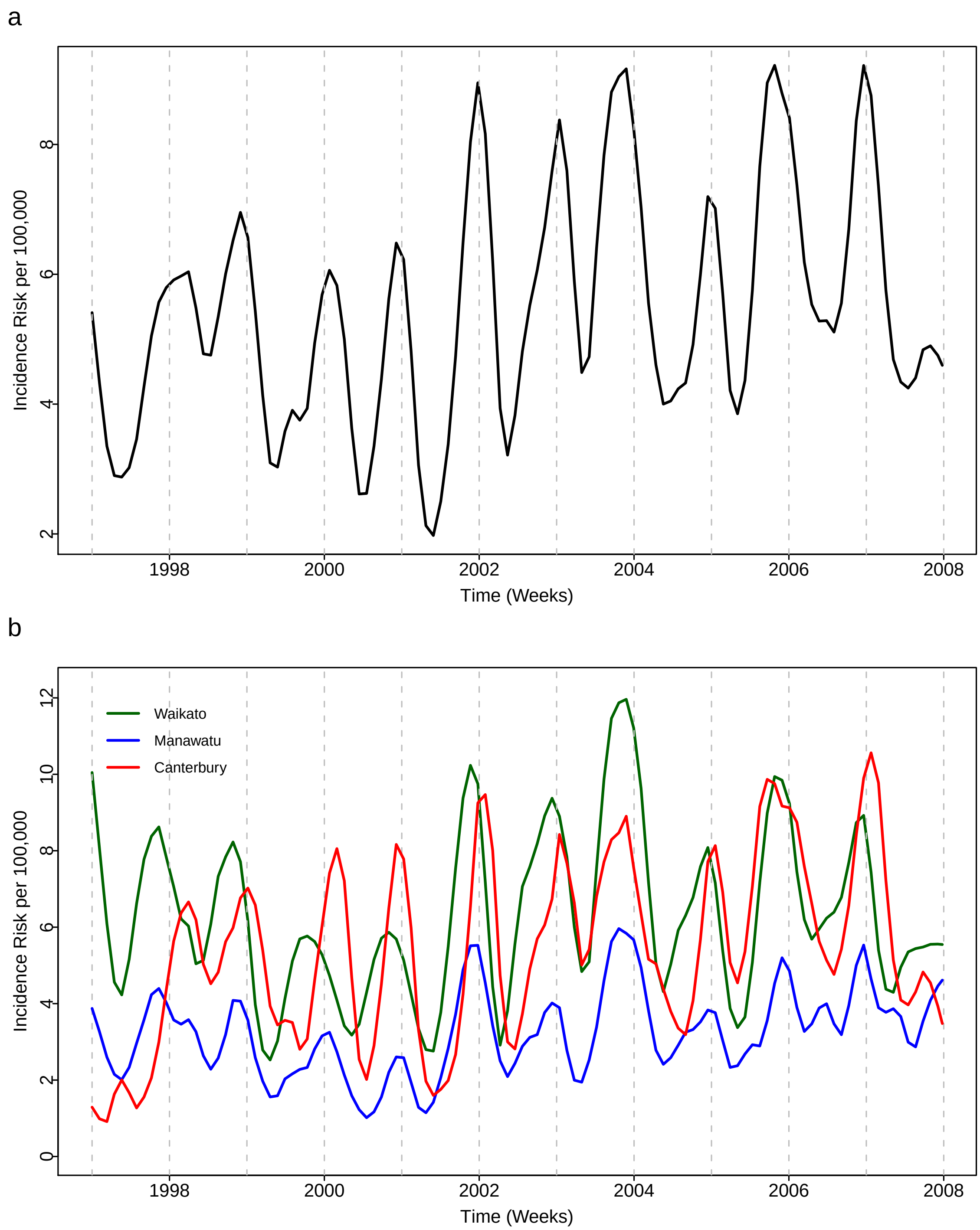


Figure 4.6: Loess smoothed time series of a) the average IR of campylobacteriosis across the three study regions, and b) the IR of campylobacteriosis stratified by study region, from 1997 - 2007.

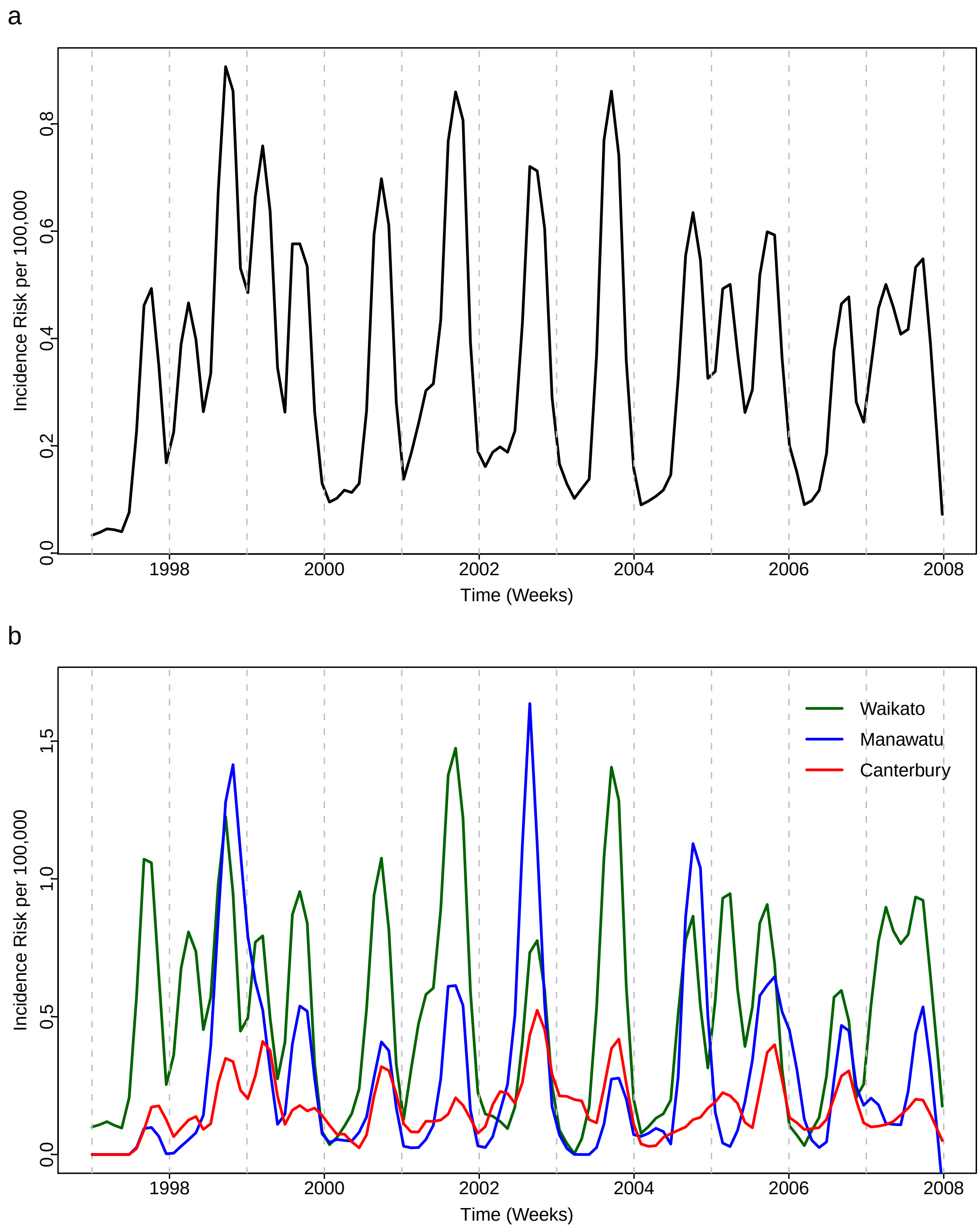


Figure 4.7: Loess smoothed time series of a) the average IR of cryptosporidiosis across the three study regions, and b) the IR of cryptosporidiosis stratified by study region, from 1997 - 2007.

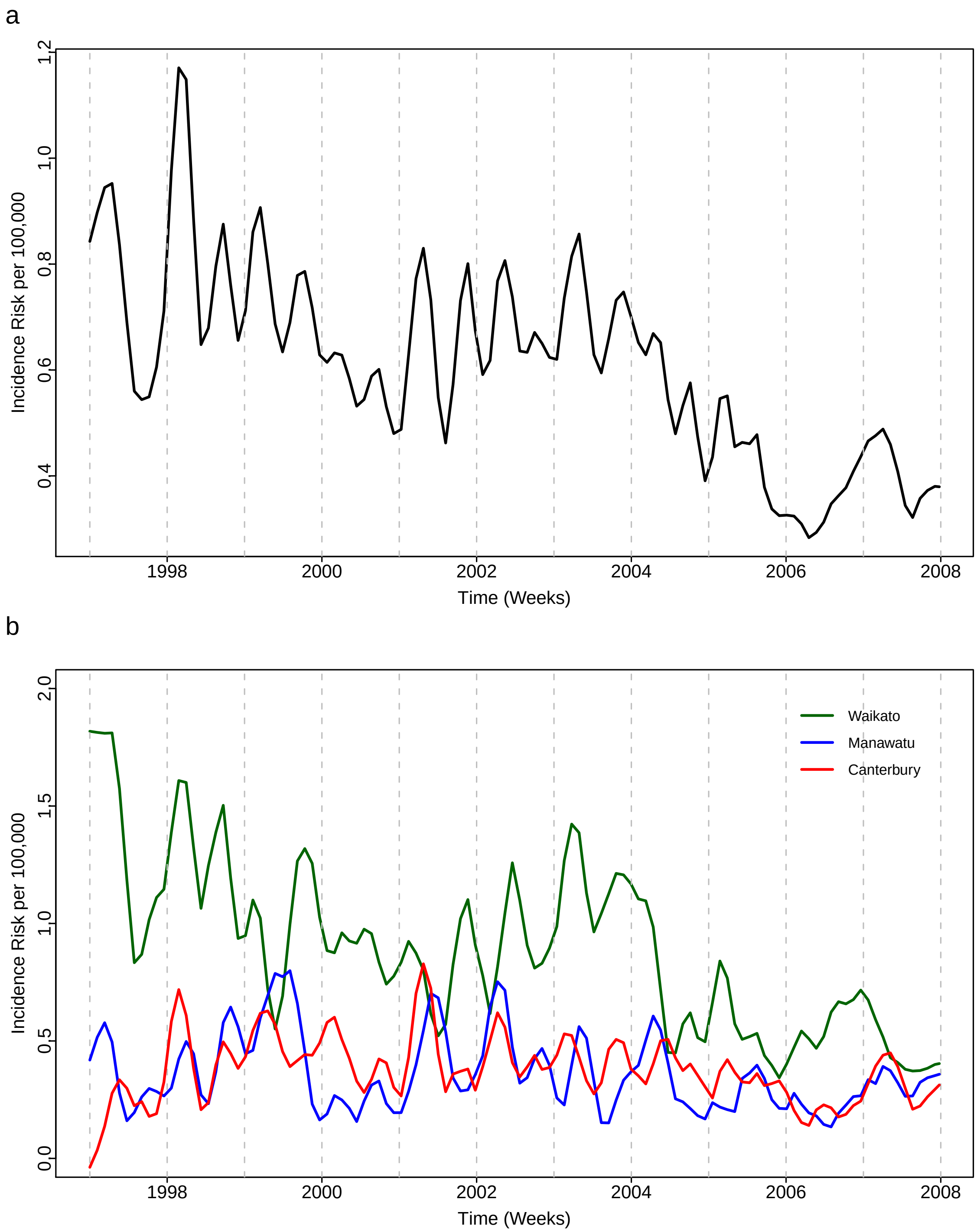


Figure 4.8: Loess smoothed time series of a) the average IR of giardiasis across the three study regions, and b) the IR of giardiasis stratified by study region, from 1997 - 2007.

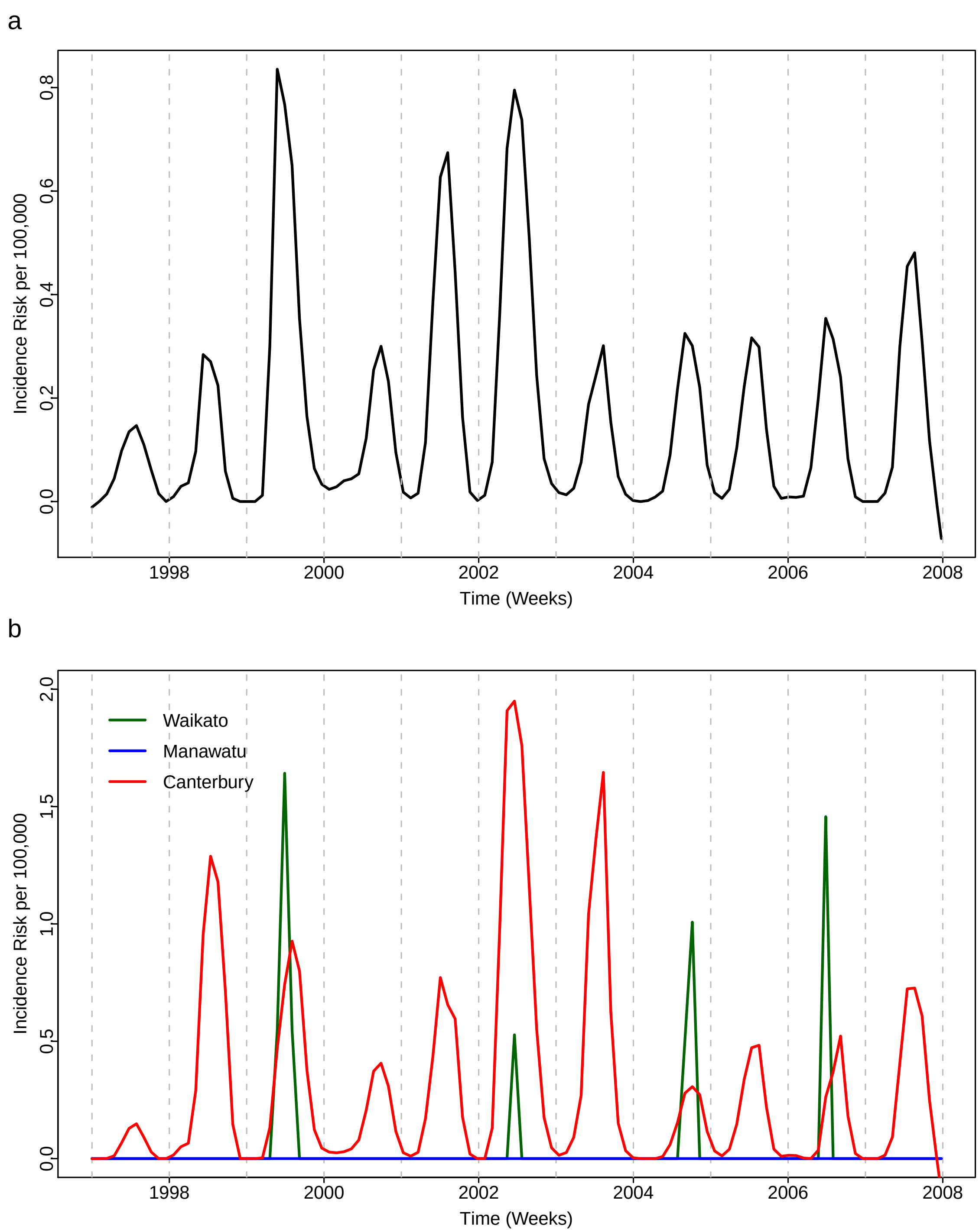


Figure 4.9: Loess smoothed time series of a) the average IR of influenza hospitalisations across the three study regions, and b) the IR of influenza hospitalisations stratified by study region, from 1997 - 2007.

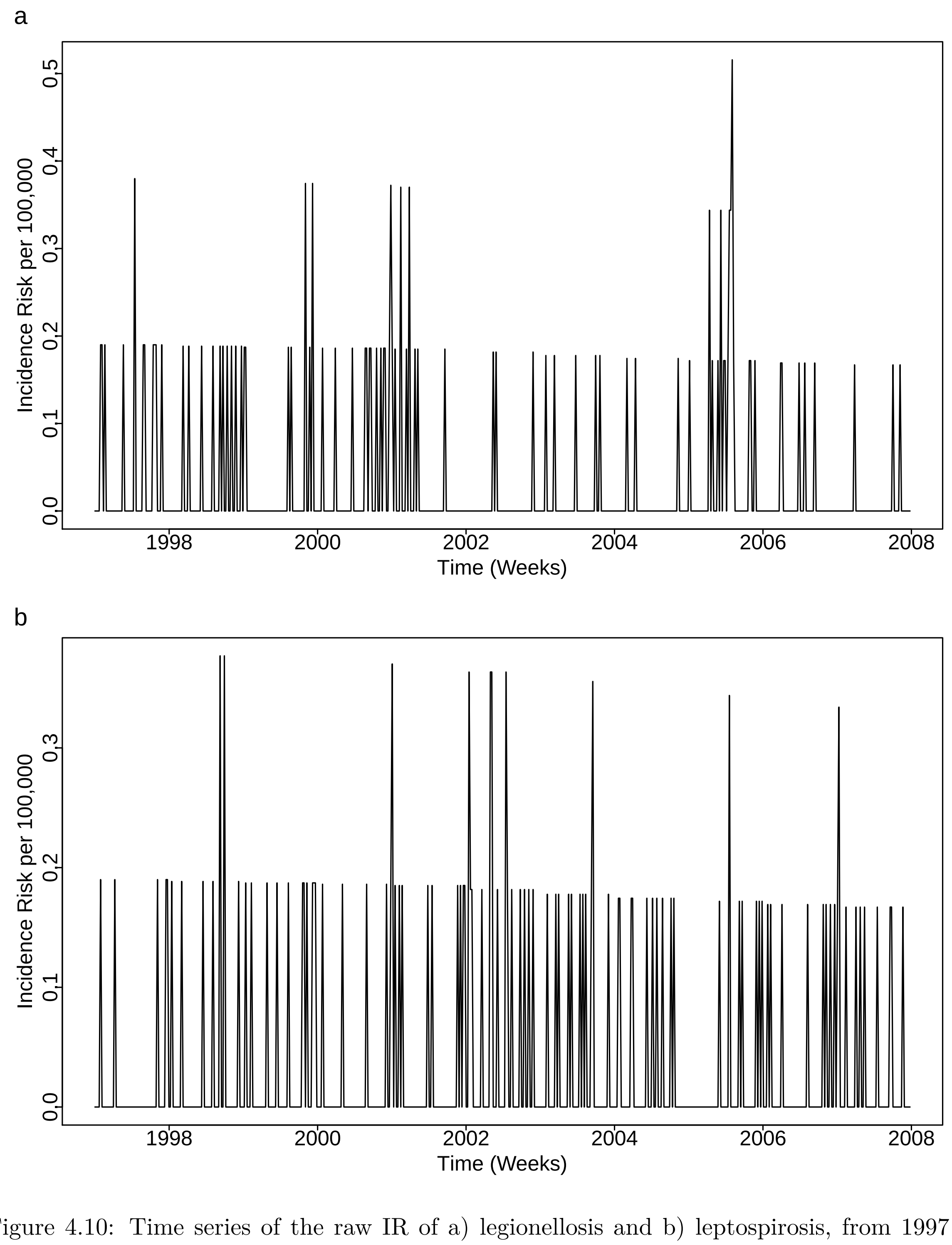
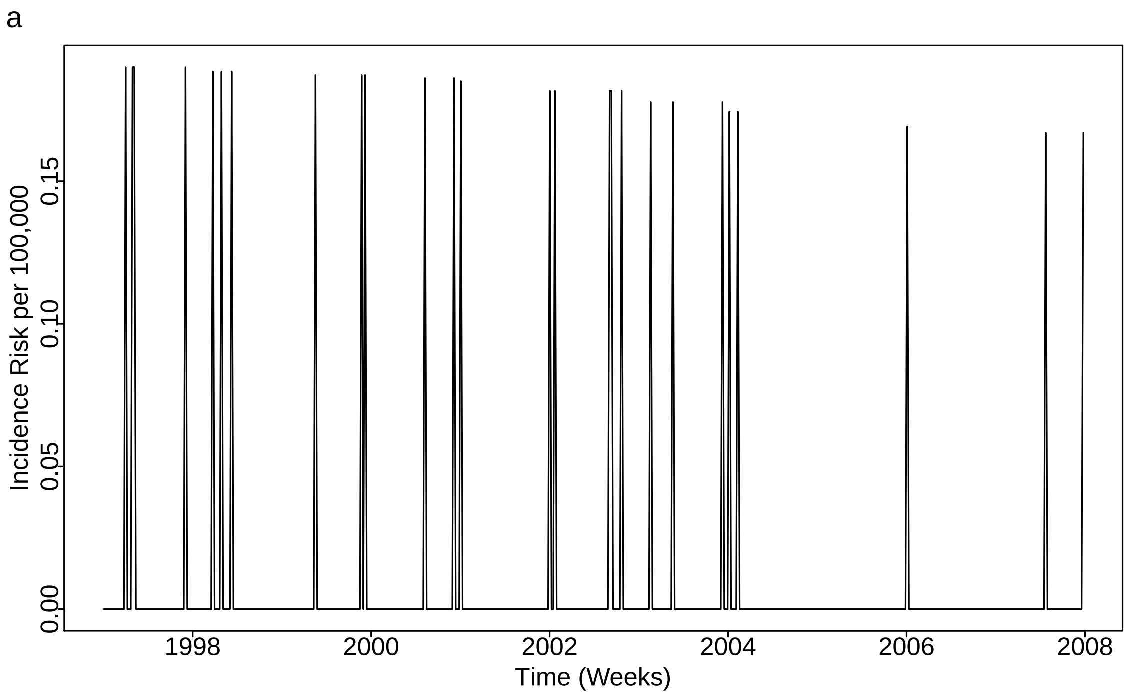


Figure 4.10: Time series of the raw IR of a) legionellosis and b) leptospirosis, from 1997 - 2007.



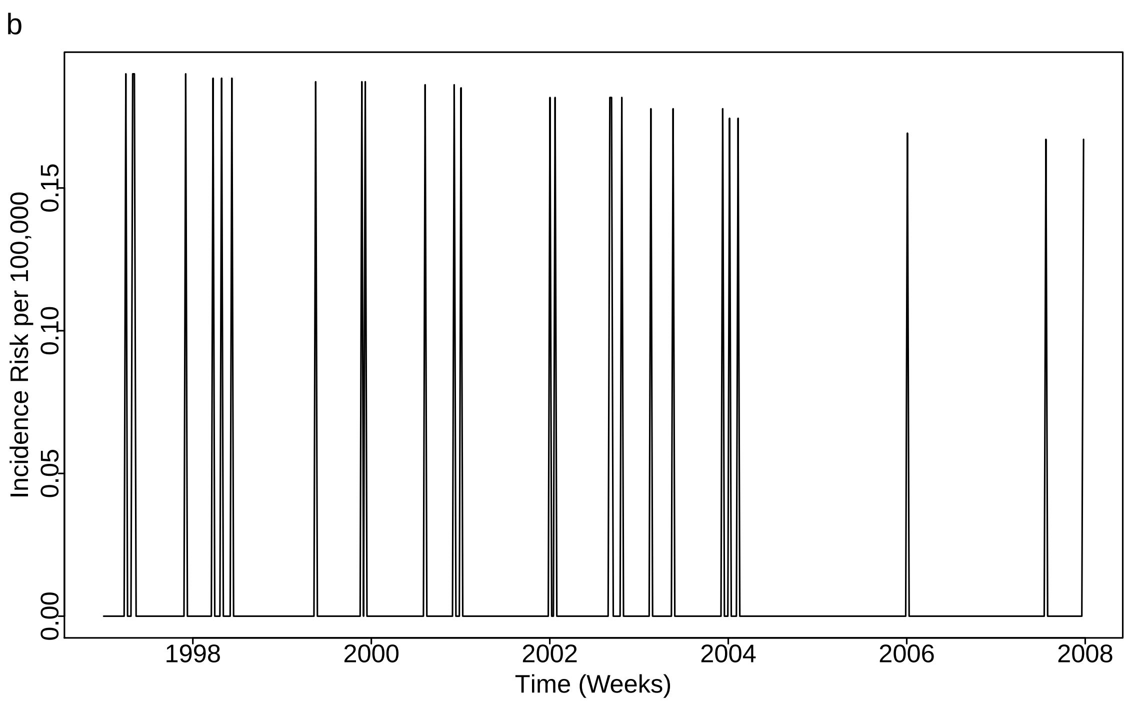


Figure 4.11: Time series of the raw IR of a) listeriosis and b) rheumatic fever, from 1997 - 2007.

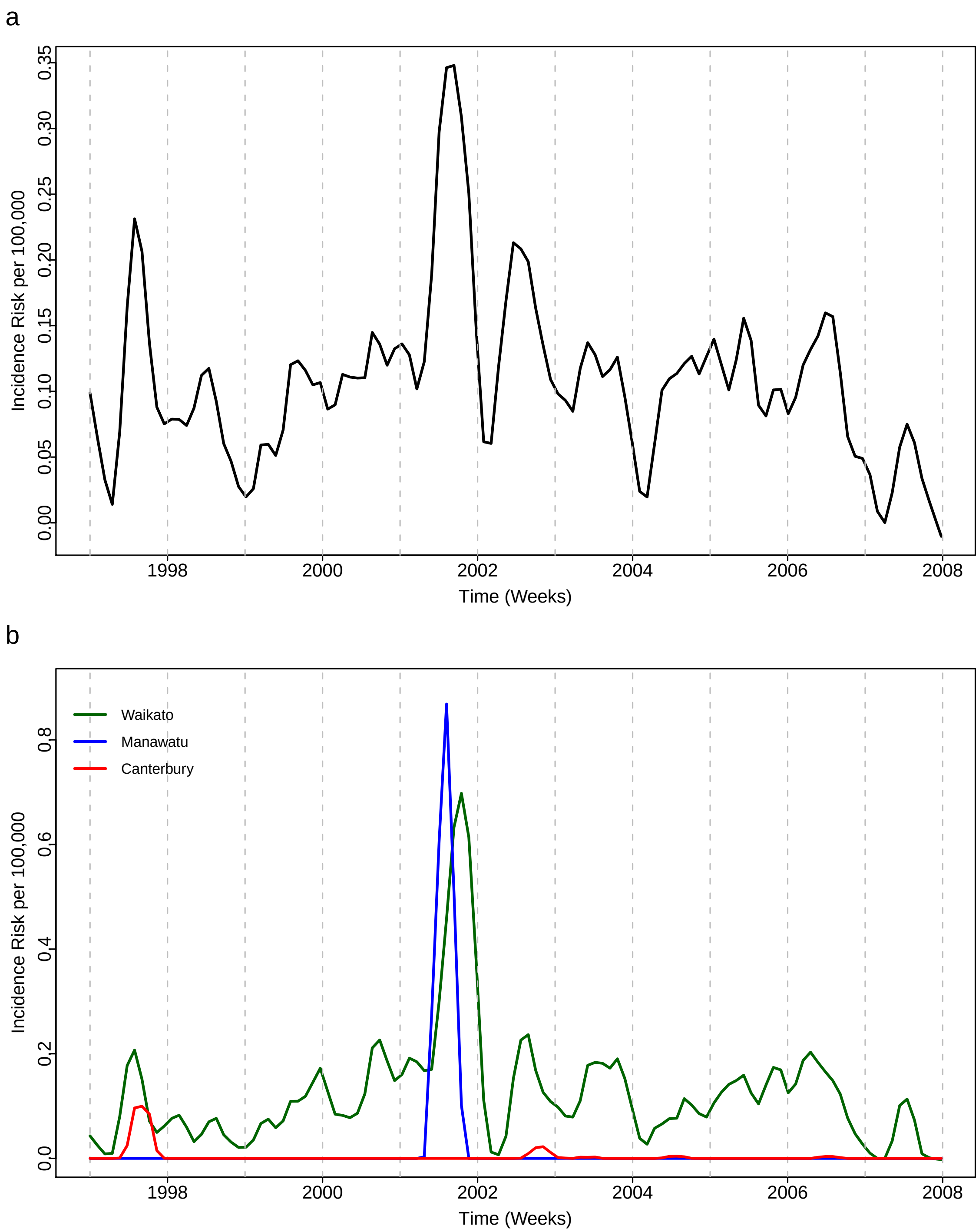


Figure 4.12: Loess smoothed time series of a) the average IR of meningococcal disease across the three study regions, and b) the IR of meningococcal disease stratified by study region, from 1997 - 2007.

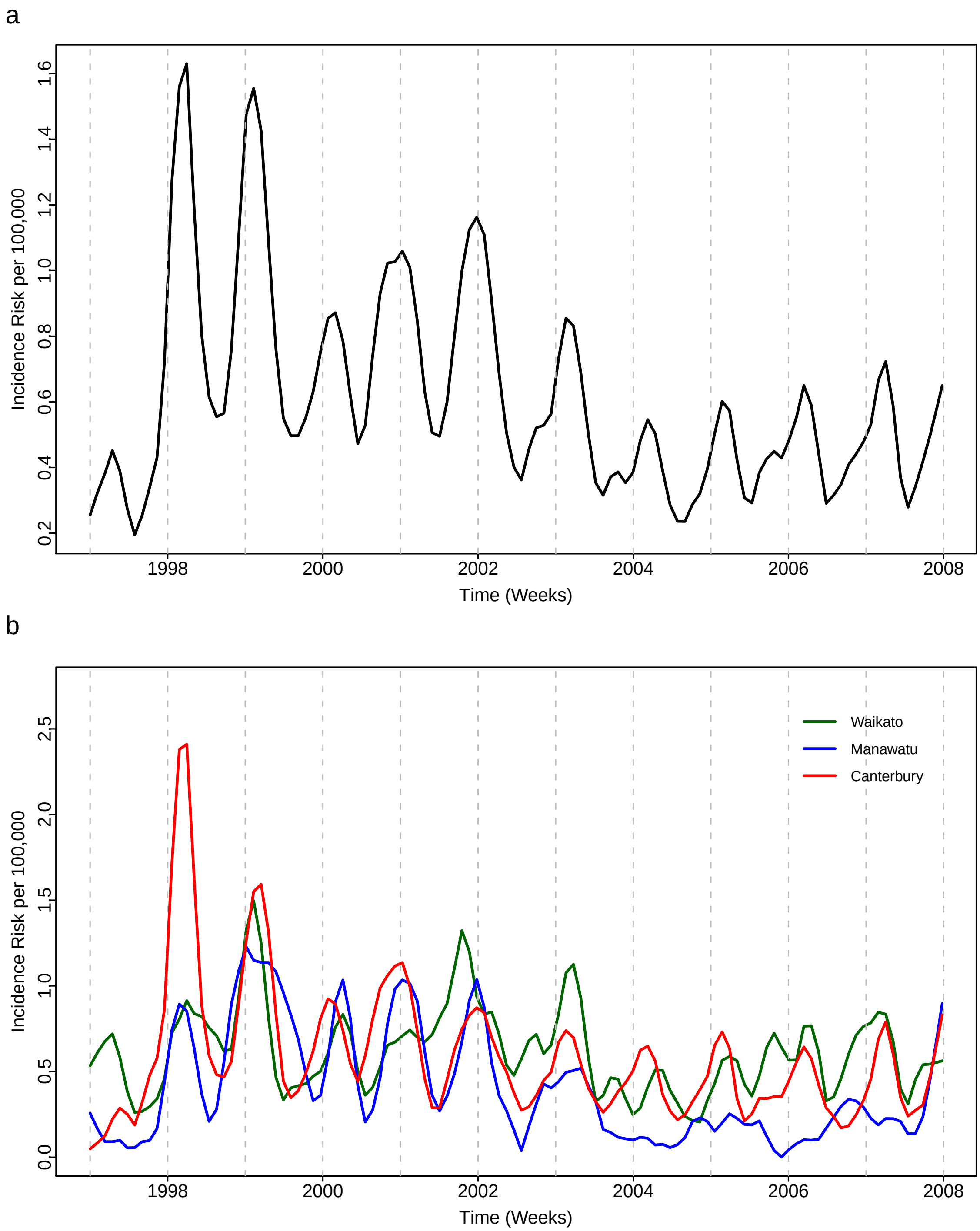


Figure 4.13: Loess smoothed time series of a) the average IR of salmonellosis across the three study regions, and b) the IR of salmonellosis stratified by study region, from 1997 - 2007.

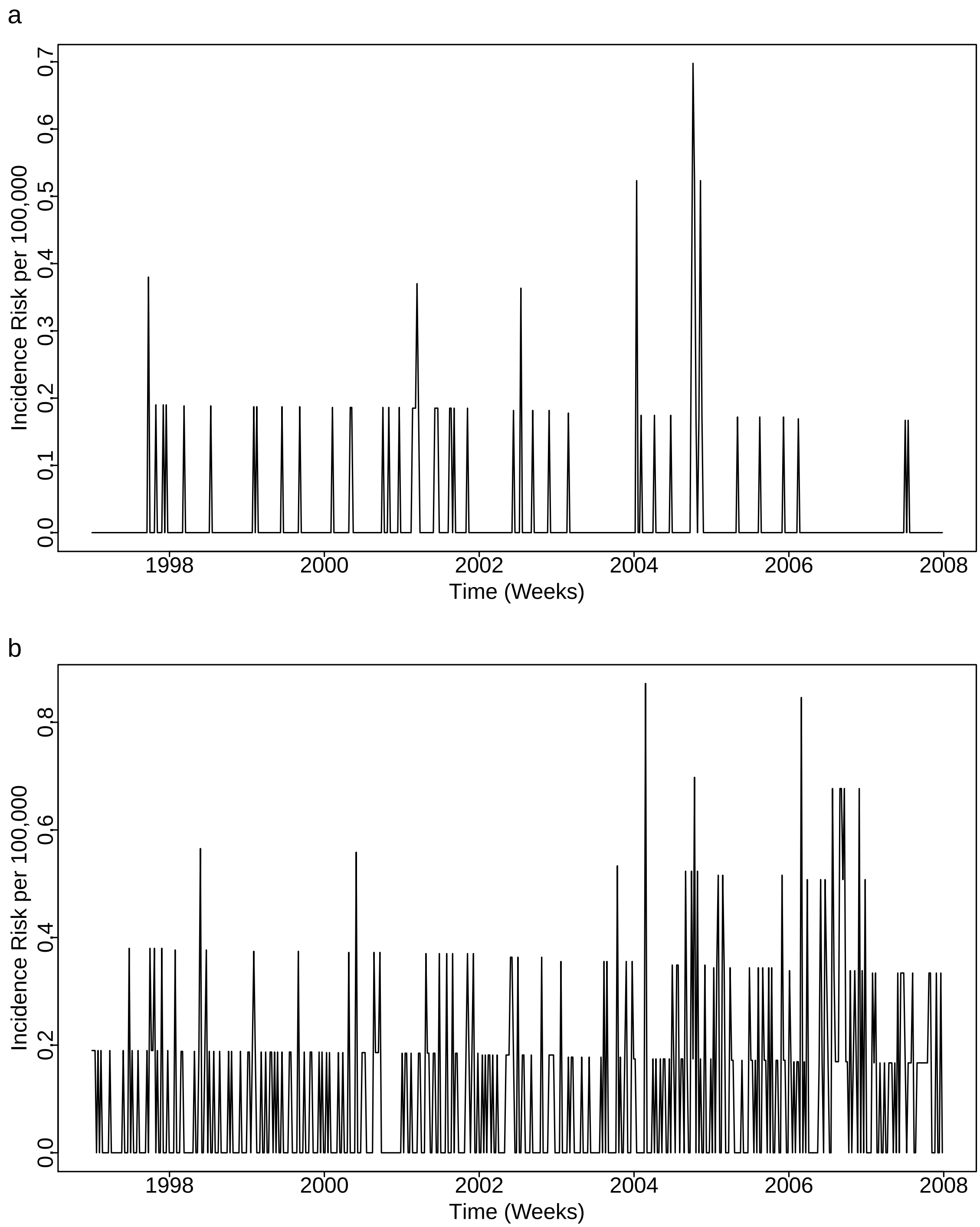


Figure 4.14: Time series of the raw IR of a) shigellosis and b) tuberculosis, from 1997 - 2007.

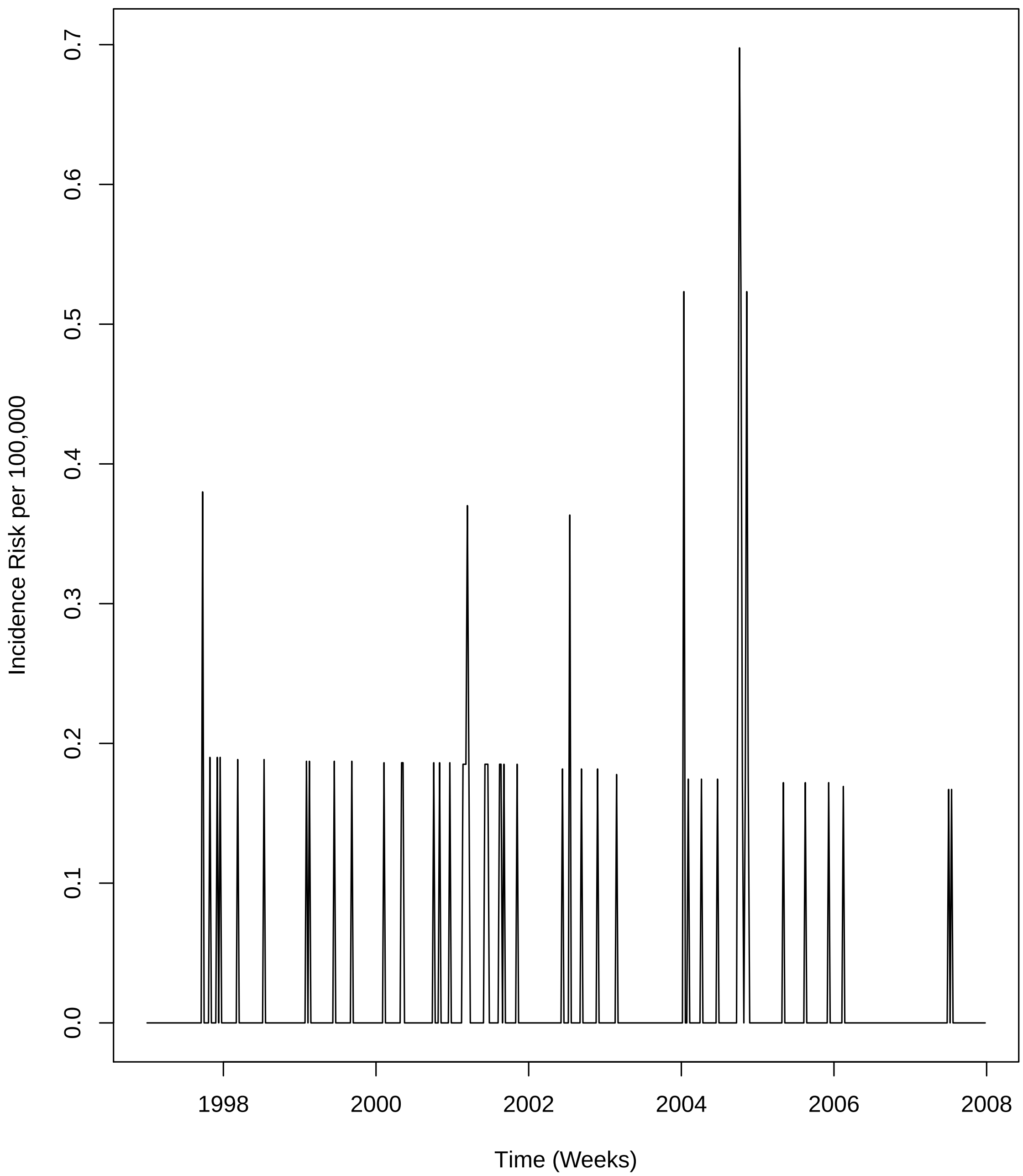


Figure 4.15: Time series of the raw IR of VTEC from 1997 - 2007.

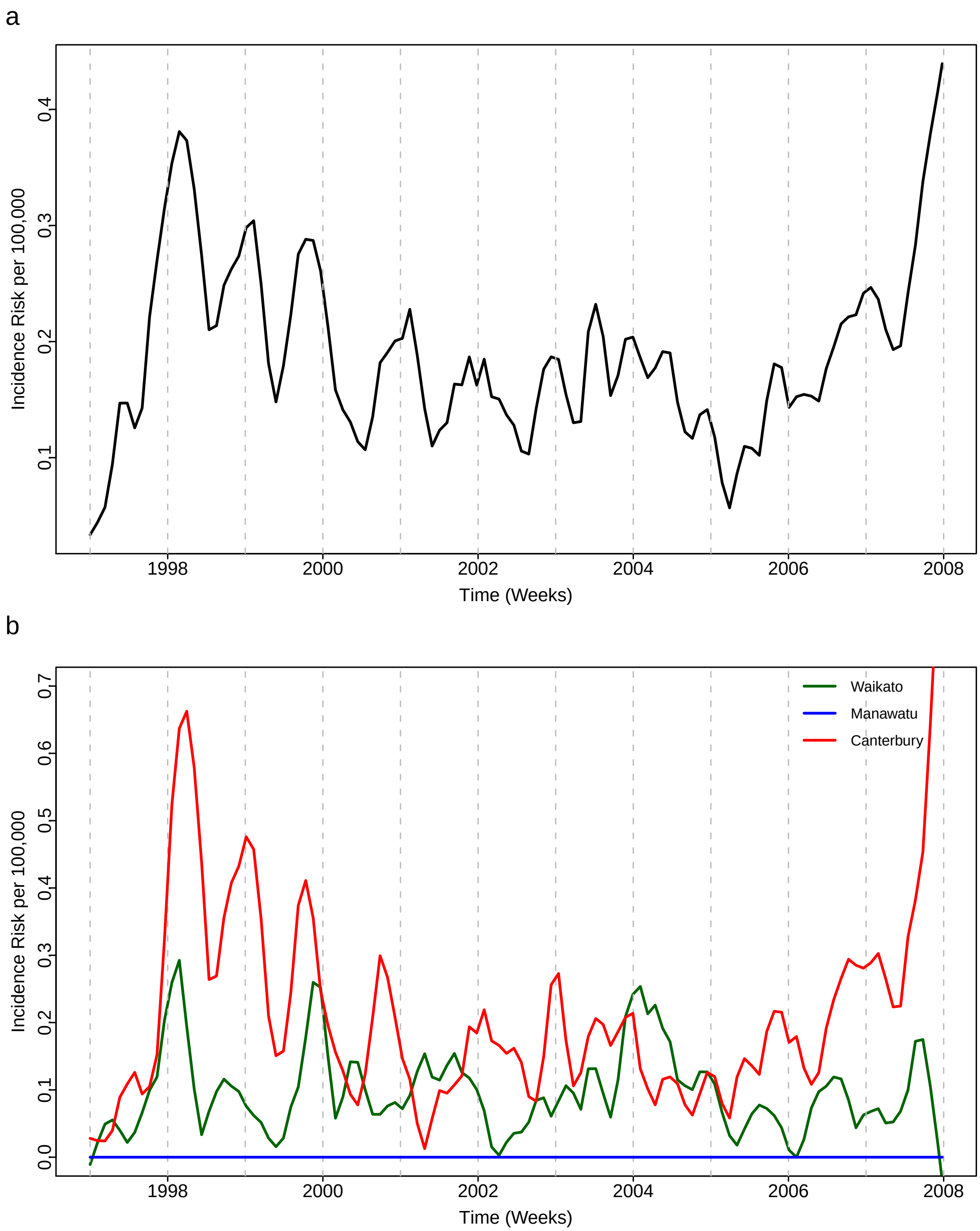


Figure 4.16: Loess smoothed time series of a) the average IR of yersiniosis across the three study regions, and b) the IR of yersinosis stratified by study region, from 1997 - 2007.

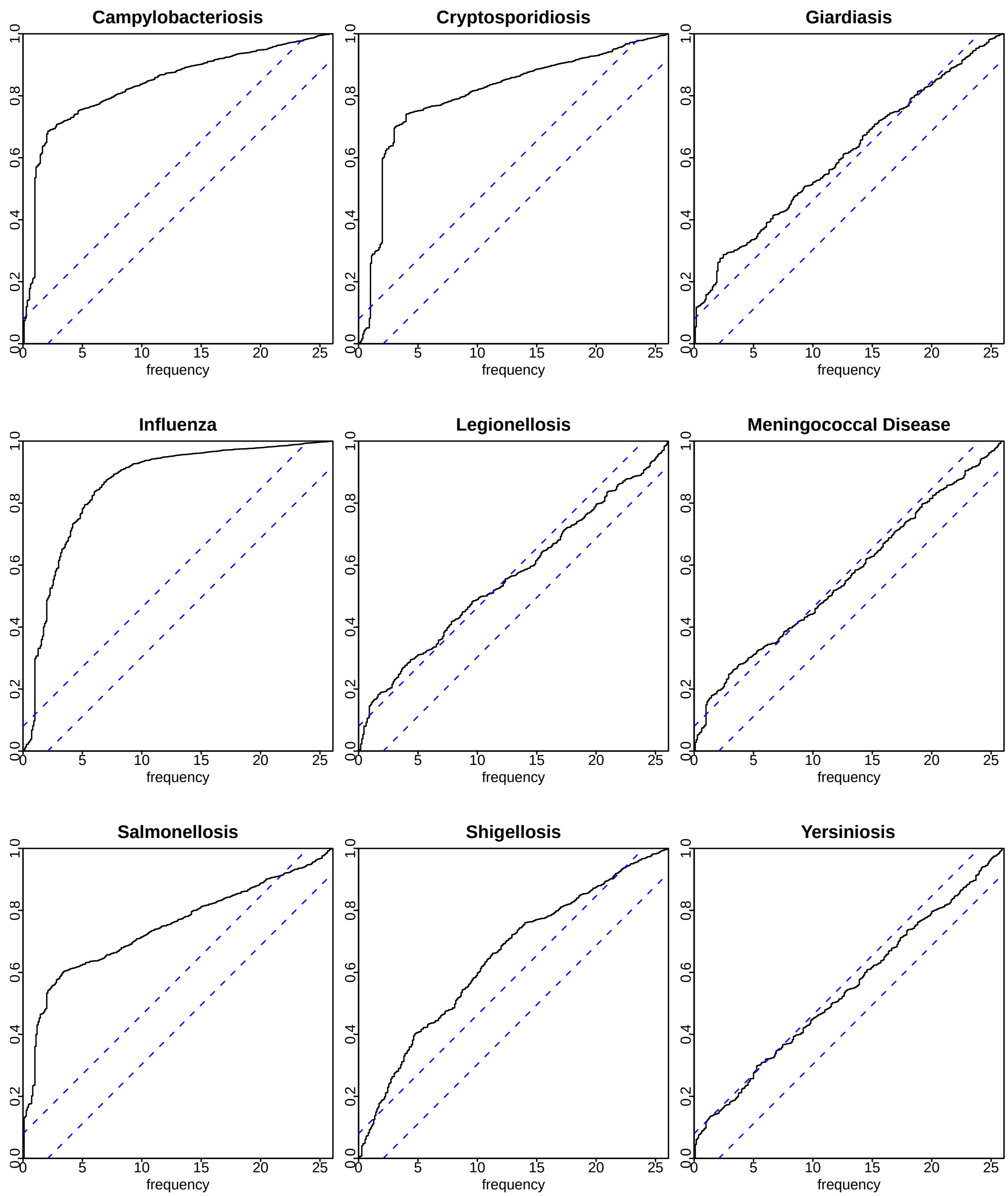


Figure 4.17: The cumulative spectral analysis of each disease to check for seasonality in the average IR throughout the three study locations, from 1997 - 2007.

Chapter 5

Variable Screening

5.1 Introduction

The aim of this chapter was to explore the associations of weather variables and other possible confounders when regressed against the IR of each indicator disease. The findings will be used to inform variable selection for the past association model (Chapters 6) and are further discussed in Chapter 8.

5.2 Method

5.2.1 Data

We now return to the weekly measurements for each weather, disease and confounding variables spanning 1997 - 2007, for 11,491 VCS grids as defined by NIWA [74].

Weather Data

The following weather variables were taken forward into the modelling process: the weekly average absolute humidity, weekly average temperature and weekly average rainfall. These weather variables were selected due to the availability of climate projection data for 2015, 2040 and 2090 from NIWA, as required in Chapter 7.

The weather data was standardized using Equation 5.1. Standardizing continuous variables produces estimated coefficients from the model output that once exponentiated, represent the change in incidence risk from an increase of one standard deviation in that variable.

$$Z_{kit} = \frac{W_{kit} - \bar{W_k}}{s.e.(W_k)} \tag{5.1}$$

where Z_{kit} is the standardized version of the weather variable W_k, in the grid i during week t; $\bar{W_k}$ is the mean and $s.e.(W_k)$ is the standard error for the weather variable W_k over the 11 year study period.

Disease Data

During the disease selection stage (Chapter 4), the following diseases were chosen to be taken into the modelling phase: campylobacteriosis, cryptosporidiosis, influenza hospitalisations and meningococcal disease. During screening, the offset was set to the population at risk (PAR). The addition of an offset is useful when modelling the counts of disease as it models the incidence risk (IR) of disease instead of the raw count, thus taking into account differing population sizes. An offset is included by the addition of a $\log(PAR_t)$ term to the right-hand side of the equation (see Equation 5.3).

Demographic Data

Grids with no population at risk (PAR = 0) were excluded from the screening process (an average of 23 out of 11,491 grids per year), as the offset $log(PAR)$ is undefined when PAR equals zero. The excluded VCS grids typically represented mountains, lakes or river areas.

For variables with multiple categories (DHB and SDI), the choice of baseline is selected by taking a group that is "typical". A baseline choice of a group that is rare will cause the coefficient estimates of the other groups to be poorly estimated with unrealistic standard errors. In the case of DHB, Counties Manukau was set as the baseline as it is a DHB that is representative of a mid-range PAR sized DHB in New Zealand (see Table 3.2). The baseline for social deprivation index (SDI) was set to a SDI level of 5. (Refer to Table 3.3.)

Similarly, with proportion variables (age, gender and ethnicity) a baseline category was assigned. For age, the proportion between 4 and 65 years of age in each grid was set as the baseline category. For gender, the proportion of males in each grid was set as the baseline. For ethnicity, the proportion of European or Other Ethnicity (including New Zealander) was set as the baseline. The proportion variables were then standardized using Equation 5.1. For example, in the case of gender this resulted in a variable measuring the expected change in IR from an increase of one standard deviation in the proportion of females in each VCS grid.

Environmental Data

The count of beef, dairy and sheep animals per grid was recalculted as the animal density per km^2 was calculated using the area of each VCS grid (approximately 25 km^2). Each animal density was then standardized (Equation 5.1) to produce the standardized animal density per km^2.

The percentage of each category of land use was standardized to produce the standardized percentage of each category of land use in each grid (Equation 5.1). The baseline was set as Scrub, as this was the land use type with the median overall percentage of land coverage in New Zealand (9.5%).

Drinking water quality was entered into the model as a factor, with the baseline set to compliant, good quality water.

Rurality was also entered as a factor, with urban areas set as the baseline. (Refer to Table 3.5.)

Other Variables

The influenza vaccination coverage per 100,000 was standardized using Equation 5.1, to produce the standardized rate of influenza vaccination per 100,000 population.

School holidays was entered into the model as a factor, where non-holiday weeks were set as the baseline.

5.2.2 Size and Complexity of Dataset

Running a statistical model on a dataset of this size could not be achieved with a Microsoft 32-bit operation system due to the memory restrictions in the statistical package R. To resolve this issue, screening was carried out on a 64-bit linux operating system with four CPU's (Xeon X5482 CPU with 16 Gb RAM and 4 Tb hard drive). The 64-bit Linux operating system allowed an increase in the R memory allocation.

All data manipulations and statistical analyses in this chapter were performed using R version 2.12.1 [62].

5.2.3 Screening

To adjust for seasonality in both the outcome (disease) and the explanatory variables, sine and cosine variables were added within the Poisson generalized linear model (GLM). The inclusion of the sine and cosine terms as explanatory variables were used to account for the annual seasonal cycle observed in the IR of disease and the weather variables, otherwise a spurious association might appear purely because both variables are seasonal. The sine and cosine variables were created with a period of 52.18 (365.25 days per year / 7 days per week) and t equal to the week number (1 to 574) spanning from January 1997 to December 2007, as shown below:

$$X_1 = \sin\left(\frac{2\pi t}{52.18}\right)$$
$$X_2 = \cos\left(\frac{2\pi t}{52.18}\right) \tag{5.2}$$

A Poisson generalized linear model (GLM) was used in the screening process to identify any significant associations of weather variables and other possible confounders when regressed against the IR of each indicator disease. In the Poisson GLM, spatial and temporal dependencies of the data are unable to be specifically accounted for. The offset was set to PAR in order to take into account the differing population sizes in each grid.

The significance of association required to be considered for selection into the past association and projection calculation stages (Chapter 6 and 7, respectively) was a p-value $\leq$ 0.05. In epidemiology, a more common significance level during screening is p $\leq$ 0.20, however as discussed earlier, the Poisson GLM model does not adjust for spatial or temporal correlation over and above that contained in the covariates, hence we must be conservative in our interpretation of p-values during screening. The required significance level was set to p-value $\leq$ 0.05 in the hope it would reduce the chances of taking variables through to the past association modelling stages (Chapter 6) which are in fact only significant due to the large number of "repeated measures" (6,448,316 measures of "one location" derived from the 574 weekly measurements per grid with PAR $>$ 0).

The residual deviance was compared to the number of degrees of freedom to check for overdispersion. When evident, a quasi-Poission GLM was used in both screening phases for each disease. The quasi-Poisson GLM has the potential to additionally account for the effects of the spatial and temporal structure of the data.

The screening was carried out in two phases. Phase One involved screening the weather and seasonal terms. Significant associations found in Phase One were then used as a base model (B) to which possible confounders were added one at a time, to screen for additional significant associations in Phase Two.

Phase One: Screening of Weather and Seasonal Terms.

A quasi-Poisson generalised linear model (GLM) was used to screen the three standardized weather variables (average absolute humidity, average temperature and average rainfall), along with the seasonal terms (sine and cosine). The expected value of the count of disease (Y) is Poisson distributed with a meant of μ, as described in Equation 5.3.

$$\log(\mu) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z_1 + \beta_4 Z_2 + \beta_5 Z_3 + \log(PAR) \tag{5.3}$$

where $\log(\mu)$ denotes the expected count of disease (Y), α denotes the intercept, $\beta_1 - \beta_5$ denote the estimated coefficients; X_1 and X_2 denote the seasonal terms sine and cosine, Z_1 denotes the standardized weekly average absolute humidity, Z_2 denotes the standardized weekly average temperature, Z_3 denotes standardized weekly average rainfall and $\log(PAR)$ denotes the offset of the population at risk in each grid.

As discussed above, the level required to be considered significantly associated with the IR of disease was set at $p \leq 0.05$. These significant associations were then used to form a base model, denoted as B.

Phase Two: Screening of Possible Confounders.

Possible confounders were added, one at a time, to the base model B to explore the over and above association with the IR of disease, having adjusted for the significant weather and seasonal variables (Equation 5.4). With some confounders, more than one covariate was added to the base model at once to account for each level within the confounder. An example of the levels within a confounder variable include the standardized proportion of each ethnicity in a grid, which contains four levels: Maori, Pacific Peoples, Asian, Middle Eastern/Latin American/African, with the baseline set to European or Other Ethnicity (including New Zealander); or the drinking water quality in each grid, with three levels: intermediate drinking water quality, poor drinking water quality and unknown drinking water quality, and the baseline set to good drinking water quality.

$$\log(\mu) \sim \alpha + B + \beta_6 Z_4 + ... + \beta_k Z_k + \log(PAR) \tag{5.4}$$

where $\log(\mu)$ denotes the expected count of disease (Y), α denotes the intercept; B denotes the baseline model from Phase One containing the significant associations with weather and seasonal variables; β_6 to β_k denote the estimated coefficients, Z_4 to Z_k denote the categories within each possible confounder screened one at a time and $\log(PAR)$ is the offset, taking into account the different PAR in each grid.

5.3 Results

During Phase One, overdispersion was found to be present in all four disease models when Poisson GLM models were initially used, hence quasi-Poisson GLMs were used in both the Phase One and Two screening.

5.3.1 Campylobacteriosis

In Phase One, the weather variables found to be significantly associated with the incidence risk (IR) of campylobacteriosis in New Zealand were the average absolute humidity (β_3=-0.0227, p-value=0.000), average rainfall (β_4=0.0505, p-value=0.000) and average temperature (β_5=0.0284, p-value=0.000). The sine (β_1=-0.2110, p-value=0.000) term was found to be significantly associated with the IR of campylobacteriosis in New Zealand. These results are displayed in Table 5.1.

An example of the interpretation of the Phase One screening results for campylobacteriosis can be illustrated using the weekly average absolute humidity (β_3=-0.0227, p-value=0.000). For an increase of one standard deviation in the weekly average absolute humidity in grid i, the quasi-Poisson GLM suggests that the expected IR of campylobacteriosis in that grid changes by a factor of 0.978 ($= e^{-0.0227}$), when the effects of the other weather and seasonal covariates are held constant.

The demographic confounders indicating a significant association with the IR of campylobacteriosis during Phase Two, when added to the base model one confounder at a time, were age, DHB, ethnicity, gender and SDI (see Table 5.2). The significant environmental confounders were dairy, beef and sheep density, drinking water quality and land use (see Table 5.3). School holiday weeks were also found to be significantly associated with the incidence risk of campylobacteriosis, having adjusted for the effect of the base model (see Table 5.4).

To illustrate the interpretation of the Phase Two screening results, consider the demographic confounder *gender*. For each one standard deviation increase in the proportion of females in VCS grid i (β_5=0.7335, p-value=0.001), the quasi-Poisson GLM suggests that the IR of campylobacteriosis in that grid changes by a factor of 2.082 ($= e^{0.7335}$), assuming all other weather and seasonal covariates in the base model are held constant.

The explanatory variables not associated with the IR of campylobacteriosis during Phase One and Two screening were found to be the cosine seasonality term and rurality.

Variable	β_k	s.e.(β_k)	p-value
Sine	-0.2110	0.0043	0.000
Cosine	0.0047	0.0055	0.392
Average Absolute Humidity	-0.0227	0.0059	0.000
Average Rainfall	0.0505	0.0075	0.000
Average Temperature	0.0284	0.0071	0.000

Table 5.1: Campylobacteriosis Phase One screening results for each weather and seasonal explanatory variable, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

5.3.2 Cryptosporidiosis

In Phase One, the weather variables found to be significantly associated with the IR of cryptosporidiosis in New Zealand were the average weekly rainfall (β_4=-0.1331, p-value=0.000) and average temperature (β_5=-0.3695, p-value=0.000). Both the sine (β_1=-0.6852, p-value=0.000) and cosine (β_2=-0.1730, p-value=0.000) terms were included in the base model, as they were found to be significantly associated with the IR of cryptosporidiosis in New Zealand. These results are displayed in Table 5.5.

The significant demographic confounders during Phase Two were found to be age, DHB, ethnicity, gender and SDI, having adjusted for the effect of the base model (see Table 5.6). The significant environmental confounders were found to be dairy and sheep density, intermediate and unknown drinking water quality, land use and rurality (see Table 5.7).

The explanatory variables found to be insignificantly associated with the IR of cryptosporidiosis during Phase One and Two screening were found to be the weekly average absolute humidity, beef density and school holidays (see Table 5.8).

5.3.3 Influenza Hospitalisations

In Phase One, the weather variables found to be significantly associated with the IR of influenza hospitalisations in New Zealand were the average absolute humidity (β_3=-0.4873, p-value=0.000) and average temperature (β_5=-0.0868, p-value=0.040). Both the sine (β_1=0.3928, p-value=0.000) and cosine (β_1=-1.1434, p-value=0.000) terms were found to be significantly associated with the IR of influenza hospitalisations in New Zealand. These results are displayed in Table 5.9.

The significant demographic confounders during Phase Two were found to be age, DHB, ethnicity and gender, having adjusted for the effect of the base model (see Table 5.10). The significant environmental confounders were found to be beef, dairy and sheep density, drinking water quality, land use and rurality (see Table 5.11). Influenza vaccination coverage and school holiday weeks were also found to be significantly associated with the incidence risk of influenza hospitalisations, having adjusted for the effect of the base model (see Table 5.12).

The explanatory variables found to be insignificantly associated with the IR of influenza hospitalisations during Phase One and Two screening were found to be the weekly average rainfall and SDI.

5.3.4 Meningococcal Disease

In Phase One, the weather variables found to be significantly associated with the IR of meningococcal disease in New Zealand were the average absolute humidity (β_3=0.2324, p-value=0.000) and average temperature (β_5=-0.1279, p-value=0.001). The cosine term (β_2=-0.4863, p-value=0.000) was found to be significantly associated with the IR of meningococcal disease in New Zealand. These results are displayed in Table 5.13.

The significant demographic confounders during Phase Two were found to be age, DHB, ethnicity, gender and SDI, having adjusted for the effect of the base model (see Table 5.14). The significant environmental confounders were found to be beef and sheep density, drinking water quality, land use and rurality (see Table 5.15). School holiday weeks were also found to be significantly associated with the incidence risk of meningococcal disease, having adjusted for the effect of the base model (see Table 5.16).

The explanatory variables found to be insignificantly associated with the IR of meningococcal disease during Phase One and Two screening were found to be the average weekly rainfall, the sine seasonality term and dairy density.

Demographic Confounder	**Categories**	β_k	**s.e.(β_k)**	**p-value**
Age Group (baseline those aged between 4 and 65 years of age)	Less than 4 Years	-13.9228	0.2259	0.000
	Over 65 Years	-2.6674	0.0902	0.000
DHB (baseline Counties Manukau)	Northland	0.6461	0.0176	0.000
	Waitemata	0.5931	0.0183	0.000
	Auckland	0.4075	0.0180	0.000
	Waikato	0.6015	0.0180	0.000
	Lakes	0.3000	0.0213	0.000
	Bay of Plenty	0.2119	0.0191	0.000
	Tairawhiti	-0.0344	0.0321	0.284
	Taranaki	0.5092	0.0226	0.000
	Hawke's Bay	0.1837	0.0236	0.000
	Whanganui	0.3138	0.0266	0.000
	Mid Central	0.0675	0.0226	0.003
	Hutt Valley	0.7586	0.0203	0.000
	Capital & Coast	0.9735	0.0175	0.000
	Wairarapa	0.0535	0.0300	0.075
	Nelson-Melborough	0.0834	0.0224	0.000
	West Coast	0.3557	0.0304	0.000
	Canterbury	0.6895	0.0185	0.000
	South Canterbury	1.0251	0.0218	0.000
	Otago	0.7211	0.0191	0.000
	Southland	0.6343	0.0220	0.000
Ethnicity (baseline European or Other (Including New Zealander))	Maori	-1.8364	0.0455	0.000
	Pacific Islander	-1.6052	0.0521	0.000
	Asian	0.4960	0.0543	0.000
	Middle Eastern/Latin American/African	5.1159	0.3401	0.000
Gender (baseline Male)	Female	0.7335	0.2271	0.001
SDI (baseline 5)	SDI Score of 0	1.1690	0.1349	0.000
	SDI Score of 1	1.0244	0.1350	0.000
	SDI Score of 2	1.0767	0.1348	0.000
	SDI Score of 3	1.1191	0.1348	0.000
	SDI Score of 4	1.0311	0.1348	0.000
	SDI Score of 6	0.8453	0.1347	0.000
	SDI Score of 7	0.8958	0.1346	0.000
	SDI Score of 8	0.8518	0.1346	0.000
	SDI Score of 9	0.7716	0.1346	0.000
	SDI Score of 10	0.3998	0.1347	0.003

Table 5.2: Campylobacteriosis Phase Two screening results for each demographic confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Environmental Confounder	Categories	β_k	s.e.(β_k)	p-value
Animal Density	Beef Density	-0.0953	0.0046	0.000
	Dairy Density	0.0383	0.0030	0.000
	Sheep Density	0.0901	0.0060	0.000
Drinking Water Quality	Intermediate	-0.1291	0.0066	0.000
(baseline Good)	Poor	-0.1324	0.0056	0.000
	Quality Unknown	-0.2880	0.0053	0.000
Land Use (baseline Scrub)	Artificial Surfaces	-0.6574	0.0365	0.000
	Bare Surfaces	-0.5489	0.4256	0.197
	Water Bodies	-0.7907	0.0920	0.000
	Cropland	-1.1552	0.0537	0.000
	Grassland	-0.7401	0.0363	0.000
	Sedge/Saltmarsh	-8.9560	0.5545	0.000
	Forest	-0.9637	0.0556	0.000
Rurality (baseline Urban)	Rural	-0.0090	0.0062	0.149

Table 5.3: Campylobacteriosis Phase Two screening results for each environmental confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Other Confounder	Category	β_k	s.e.(β_k)	p-value
School Holidays	Holiday Week	0.0549	0.0052	0.000

Table 5.4: Campylobacteriosis Phase Two screening results for other confounders, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Coefficients	β_k	s.e.(β_k)	p-value
Sine	-0.6852	0.0231	0.000
Cosine	-0.1730	0.0293	0.000
Average Absolute Humidity	0.0214	0.0359	0.550
Average Rainfall	-0.1331	0.0331	0.000
Average Temperature	-0.3695	0.0426	0.000

Table 5.5: Cryptosporidiosis Phase One screening results for each weather and seasonal explanatory variable, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Demographic Confounder	**Categories**	β_k)	**s.e.(**β_k	**p-value**
Age Group (baseline those aged between 4 and 65 years of age)	Less than 4 Years	2.2596	0.8578	0.008
	Over 65 Years	2.2081	0.3273	0.000
DHB (baseline Counties Manukau)	Northland	-0.6196	0.0936	0.000
	Waitemata	-0.7221	0.1007	0.000
	Auckland	-0.5347	0.0936	0.000
	Waikato	1.0923	0.0786	0.000
	Lakes	0.5758	0.0946	0.000
	Bay of Plenty	-0.1604	0.0959	0.094
	Tairawhiti	0.2275	0.1333	0.088
	Taranaki	-0.0030	0.1150	0.979
	Hawke's Bay	0.5540	0.0935	0.000
	Whanganui	0.3981	0.1152	0.001
	Mid Central	0.7236	0.0886	0.000
	Hutt Valley	0.8294	0.0899	0.000
	Capital & Coast	0.5701	0.0831	0.000
	Wairarapa	0.5087	0.1214	0.000
	Nelson-Melborough	-0.1866	0.1092	0.088
	West Coast	1.0431	0.1154	0.000
	Canterbury	0.3668	0.0840	0.000
	South Canterbury	1.3752	0.0935	0.000
	Otago	0.4429	0.0906	0.000
	Southland	0.7489	0.0962	0.000
Ethnicity (baseline European or Other (Including New Zealander))	Maori	-0.9174	0.1363	0.000
	Pacific Islander	-2.6720	0.2672	0.000
	Asian	-5.4939	0.4109	0.000
	Middle Eastern/Latin American/African	-13.6698	4.6665	0.003
Gender (baseline Male)	Female	-13.3064	0.4170	0.000
SDI (baseline 5)	SDI Score of 0	3.2242	1.2733	0.011
	SDI Score of 1	3.0318	1.2735	0.017
	SDI Score of 2	2.8370	1.2733	0.026
	SDI Score of 3	2.8780	1.2733	0.024
	SDI Score of 4	2.9073	1.2732	0.022
	SDI Score of 6	2.7711	1.2729	0.030
	SDI Score of 7	2.6715	1.2727	0.036
	SDI Score of 8	2.6121	1.2726	0.040
	SDI Score of 9	2.5322	1.2726	0.047
	SDI Score of 10	2.3487	1.2730	0.065

Table 5.6: Cryptosporidiosis Phase Two screening results for each demographic confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Environmental Confounder	**Categories**	β_k	**s.e.**(β_k)	**p-value**
Animal Density	Beef Density	0.0258	0.0138	0.061
	Dairy Density	0.2721	0.0083	0.000
	Sheep Density	0.3706	0.0141	0.000
Drinking Water Quality	Intermediate	0.1558	0.0354	0.000
(baseline Good)	Poor	0.0595	0.0311	0.056
	Quality Unknown	0.1325	0.0277	0.000
Land Use (baseline Scrub)	Artificial Surfaces	-1.6104	0.1006	0.000
	Bare Surfaces	-1.5102	0.5638	0.007
	Water Bodies	0.3084	0.1964	0.116
	Cropland	-0.5747	0.1271	0.000
	Grassland	0.0751	0.0940	0.424
	Sedge/Saltmarsh	-8.6878	1.2764	0.000
	Forest	-1.0822	0.1338	0.000
Rurality (baseline Urban)	Rural	0.4880	0.0227	0.000

Table 5.7: Cryptosporidiosis Phase Two screening results for each environmental confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Other Confounder	**Category**	β_k	**s.e.**(β_k)	**p-value**
School Holidays	Holiday Week	-0.0292	0.0279	0.295

Table 5.8: Cryptosporidiosis Phase Two screening results for other confounders, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Coefficients	β_k	**s.e.**(β_k)	**p-value**
Sine	0.3928	0.0239	0.000
Cosine	-1.1434	0.0330	0.000
Average Absolute Humidity	-0.4873	0.0346	0.000
Average Rainfall	-0.0358	0.0292	0.221
Average Temperature	-0.0868	0.0423	0.040

Table 5.9: Influenza Hospitalisations Phase One screening results for each weather and seasonal explanatory variable, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Demographic Confounder	**Categories**	β_k	**s.e.(β_k)**	**p-value**
Age Group (baseline those aged between 4 and 65 years of age)	Less than 4 Years	5.6546	0.9262	0.000
	Over 65 Years	0.1548	0.3464	0.655
DHB (baseline Counties Manukau)	Northland	-0.9398	0.0626	0.000
	Waitemata	-0.4360	0.0603	0.000
	Auckland	-0.5416	0.0571	0.000
	Waikato	-0.5702	0.0575	0.000
	Lakes	-1.8049	0.0946	0.000
	Bay of Plenty	-0.2472	0.0556	0.000
	Tairawhiti	-0.0038	0.0795	0.962
	Taranaki	-1.4879	0.1084	0.000
	Hawke's Bay	-1.0255	0.0838	0.000
	Whanganui	-1.2093	0.1085	0.000
	Mid Central	-1.9968	0.1077	0.000
	Hutt Valley	-1.7330	0.1047	0.000
	Capital & Coast	-1.0577	0.0629	0.000
	Wairarapa	-1.5422	0.1265	0.000
	Nelson-Melborough	-0.6173	0.0662	0.000
	West Coast	-0.0169	0.0795	0.831
	Canterbury	0.1088	0.0560	0.052
	South Canterbury	-0.8492	0.0818	0.000
	Otago	-1.1878	0.0678	0.000
	Southland	-1.9846	0.1047	0.000
Ethnicity (baseline European or Other (Including New Zealander))	Maori	0.8886	0.1262	0.000
	Pacific Islander	1.2998	0.1600	0.000
	Asian	1.3906	0.3025	0.000
	Middle Eastern/Latin American/African	-0.4437	3.6835	0.904
Gender (baseline Male)	Female	5.3620	0.9871	0.000
SDI (baseline 5)	SDI Score of 0	10.2560	49.5517	0.836
	SDI Score of 1	10.5751	49.5517	0.831
	SDI Score of 2	10.6008	49.5517	0.831
	SDI Score of 3	10.8169	49.5517	0.827
	SDI Score of 4	10.6342	49.5517	0.830
	SDI Score of 6	10.9660	49.5517	0.825
	SDI Score of 7	11.1001	49.5517	0.823
	SDI Score of 8	10.8353	49.5517	0.827
	SDI Score of 9	11.1118	49.5517	0.823
	SDI Score of 10	11.3088	49.5517	0.819

Table 5.10: Influenza hospitalisations Phase Two screening results for each demographic confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Environmental Confounder	Categories	β_k	s.e.(β_k)	p-value
Animal Density	Beef Density	0.0839	0.0163	0.000
	Dairy Density	-0.1089	0.0156	0.000
	Sheep Density	-0.4109	0.0313	0.000
Drinking Water Quality	Intermediate	0.1679	0.0259	0.000
(baseline Good)	Poor	0.2511	0.0226	0.000
	Quality Unknown	-0.1463	0.0239	0.000
Land Use (baseline Scrub)	Artificial Surfaces	1.8144	0.1244	0.000
	Bare Surfaces	1.4535	0.5373	0.007
	Water Bodies	1.2776	0.2336	0.000
	Cropland	1.6095	0.1464	0.000
	Grassland	1.2666	0.1228	0.000
	Sedge/Saltmarsh	4.3744	0.8698	0.000
	Forest	1.4459	0.1601	0.000
Rurality (baseline Urban)	Rural	-0.3207	0.0283	0.000

Table 5.11: Influenza hospitalisations Phase Two screening results for each environmental confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Other Confounder	Category	β_k	s.e.(β_k)	p-value
Influenza Vaccination Coverage		0.0805	0.0070	0.000
School Holidays	Holiday Week	0.1345	0.0176	0.000

Table 5.12: Influenza hospitalisations Phase Two screening results for other confounders, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Coefficient	β_k	s.e.(β_k)	p-value
Sine	0.0360	0.0207	0.075
Cosine	-0.4863	0.0202	0.000
Average Absolute Humidity	0.2324	0.0340	0.000
Average Rainfall	0.0283	0.0261	0.278
Average Temperature	-0.1279	0.0375	0.001

Table 5.13: Meningococcal disease Phase One screening results for each weather and seasonal explanatory variable, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Demographic Confounder	Categories	β_k	s.e.(β_k)	p-value
Age Group (baseline those aged between 4 and 65 years of age)	Less than 4 Years	28.4962	1.1923	0.000
	Over 65 Years	-1.4441	0.6059	0.017
DHB (baseline Counties Manukau)	Northland	-0.8508	0.0664	0.000
	Waitemata	-0.3583	0.0630	0.000
	Auckland	0.2117	0.0563	0.000
	Waikato	-0.5225	0.0652	0.000
	Lakes	-0.0719	0.0738	0.330
	Bay of Plenty	-0.4480	0.0691	0.000
	Tairawhiti	-0.3386	0.1113	0.002
	Taranaki	-1.0275	0.1083	0.000
	Hawke's Bay	-0.4787	0.0831	0.000
	Whanganui	-0.7559	0.1160	0.000
	Mid Central	-1.1001	0.0951	0.000
	Hutt Valley	-0.9517	0.0961	0.000
	Capital & Coast	-1.3308	0.0828	0.000
	Wairarapa	-0.5928	0.1209	0.000
	Nelson-Melborough	-1.8284	0.1313	0.000
	West Coast	-1.0278	0.1631	0.000
	Canterbury	-1.3235	0.0800	0.000
	South Canterbury	-1.1926	0.1377	0.000
	Otago	-0.3982	0.0757	0.000
	Southland	-0.9249	0.1025	0.000
Ethnicity (baseline European or Other (Including New Zealander))	Maori	2.3727	0.1217	0.000
	Pacific Islander	3.8812	0.1173	0.000
	Asian	-2.2212	0.2958	0.000
	Middle Eastern/Latin American/African	4.7076	2.2160	0.034
Gender (baseline Male)	Female	7.6915	1.5075	0.000
SDI (baseline 5)	SDI Score of 0	-0.0223	0.5298	0.966
	SDI Score of 1	-0.1279	0.5306	0.810
	SDI Score of 2	-0.1461	0.5270	0.782
	SDI Score of 3	-0.2293	0.5281	0.664
	SDI Score of 4	-0.0466	0.5262	0.929
	SDI Score of 6	0.0477	0.5237	0.927
	SDI Score of 7	0.0442	0.5224	0.933
	SDI Score of 8	0.6270	0.5213	0.229
	SDI Score of 9	0.6047	0.5212	0.246
	SDI Score of 10	1.2708	0.5212	0.015

Table 5.14: Meningococcal disease Phase Two screening results for each demographic confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Environmental Confounder	**Categories**	β_k**)**	**s.e.(**β_k	**p-value**
Animal Density	Beef Density	0.0417	0.0177	0.018
	Dairy Density	-0.0121	0.0134	0.366
	Sheep Density	-0.1943	0.0313	0.000
Drinking Water Quality	Intermediate	0.3659	0.0308	0.000
(baseline Good)	Poor	0.1276	0.0280	0.000
	Quality Unknown	0.0901	0.0256	0.000
Land Use (baseline Scrub)	Artificial Surfaces	1.2035	0.1634	0.000
	Bare Surfaces	-29.2027	3.0710	0.000
	Water Bodies	2.3989	0.2698	0.000
	Cropland	1.0052	0.2057	0.000
	Grassland	0.9422	0.1638	0.000
	Sedge/Saltmarsh	4.5611	1.1488	0.000
	Forest	0.4263	0.2204	0.053
Rurality (baseline Urban)	Rural	-0.0539	0.0253	0.033

Table 5.15: Meningococcal disease Phase Two screening results for each environmental confounder, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Other Confounder	**Category**	β_k	**s.e.(**β_k**)**	**p-value**
School Holidays	Holiday Week	-0.0700	0.0262	0.008

Table 5.16: Meningococcal disease Phase Two screening results for other confounders, the estimated coefficient (β_k), standard error (s.e.(β_k)) and level of significance (p-value).

Chapter 6

Past Association Models

6.1 Introduction

The aim of this chapter is to describe the method used to model the effect of past climate variation on the four indicator diseases in New Zealand from 1997 - 2007, at a fine spatial and temporal level (574 weekly measurements for each of the 11,491 VCS grids) having adjusted for the effect of confounders. Expert opinion on epidemiological plausibility, alongside the screening results from Chapter 5 were used to select the confounding covariates for each model.

6.1.1 Size and Complexity of Dataset

The requirement for the past association model to adequately model a large, complex dataset resulted in a complicated, yet rewarding challenge. The data for each disease, weather and level within each confounding variable consisted of 11,491 VCS grids containing 574 weekly measurements per grid. Additionally, to further increase the complexity of the model, the spatial and temporal structure of each disease and explanatory covariate needed to be adjusted for. This ensured the assumption for Poisson distributed data, where the occurrence of a notified case of disease is independent, was upheld.

The length of time to completion for each MCMC iteration increased linearly with the number of variables entered into the model. With the number of iterations required for convergence in the order of 10,000, this proved to be a timely exercise, hence traditional methods for model building were not used.

All data manipulations and statistical analyses in this chapter were performed using R version 2.12.1 [62].

6.2 Data

The format for the variables used in the past association models were the same as that defined during the screening process (see Chapter 5.2.1). The only exception to this was the demographic variable social deprivation index (SDI). During screening, SDI was entered into the quasi-Poisson GLM models for each disease as a factor with 10-levels and the baseline a grid score of 5. For the past association model, SDI was entered as a linear predictor. During an examination of the screening results, it was noted that SDI appeared to have a linear effect on the association of disease and therefore decided that SDI would be entered as a linear variable in the past association models. The advantages of this decision were that it decreased the number of degrees of freedom used in the model, and ten levels within the variable were reduced to one, which was favoured in the confounding variable selection process.

6.3 Method

6.3.1 Explanatory Variable Selection

The variable selection process for each model investigating the association between past climate variation and the four indicator diseases in New Zealand occurred in two steps. In the first step, the weather and other possible confounding variables were screened, as described in Chapter 5. The computing limitations of such a complex model resulted in a restriction of the maximum number of weather and possible confounding variables to be included in each model set to eight. The decision was made to include all significant weather explanatory variables. The second step involved ranking the confounding variables found to be significantly associated with the IR of disease during screening, using expert opinion as to the epidemiological plausibility of an association [8, 27].

Furthermore, in situations where both rurality and animal density were found to be significantly associated with the IR of disease during screening, rurality would be used as a proxy for the animal density in each VCS grid. The inclusion of rurality (1 variable) over animal density (3 variables) also allowed for the inclusion of other significantly associated confounders in the past association model, and prevented collinearity occurring. The limit on the number of explanatory variables resulted in the decision that the demographic variable District Health Board (DHB) would not be included in any past association model, as it is a factor consisting of 20 levels. It was hoped that any effect of DHB on under- or over-reporting would be seen in a map of the posterior spatial effects (U_i) produced within the output of each model. Such spatial effects are incorporated implicitly in our projections, on the assumption that they remain constant over time.

6.3.2 Modifying the Knorr-Held Richardson Model

The past association model utilised a modified version of the Bayesian hierarchical model developed by Leonhard Knorr-Held and Sylvia Richardson to analyse space-time surveillance data on meningococcal disease incidence in France [45]. As described in Chapter 2.2.3, the Knorr-Held Richardson model contains latent parameters that capture temporal, seasonal and spatial trends in the meningococcal disease incidence for endemic periods, by decomposing the log-relative-risks of disease ($\log(\lambda_{it})$) into three components, as seen in Equation 6.1.

$$\log(\lambda_{it}) = R_t + S_t + U_i \tag{6.1}$$

where λ_{it} is assumed to be Poisson distributed with mean $e_{it}\lambda_{it}$, where e_{it} is the number of expected cases; R_t denotes the temporal component; S_t denotes the seasonal component; and U_i denotes the spatial component.

In order to use the Knorr-Held Richardson model in the past association modelling, Dr. Jonathan Marshall of Massey University and Dr. Simon Spencer of Warwick University carried out modifications to incorporate linear explanatory covariates, whilst removing the seasonal component, S_t (Equation 6.2).

In the French meningococcal disease data, daily notifications were modelled. The seasonal term (S_t) took into account the "day of week" effect, where notifications only occur from Monday to Friday. For the HAIFA project, ESR converted the daily number of disease notifications, obtained from EpiSurv, into weekly measurements thus avoiding the need to adjust for this effect. Additionally, any long term seasonality was assumed to be absorbed by the flexible temporal component, R_t.

The expected number of notified cases for each disease (Y_{it}) in each grid i during week t, was assumed to follow a Poisson distribution with rate $PAR_{it}\lambda_{it}$. Here the outcome variable λ_{it} represents the rate of notified cases in grid i during week t, and PAR_{it} represents the population at risk.

The offset was set to the population at risk (PAR_{it}) in each grid i during week t, to account for differing population sizes. For grids with no population (i.e. $PAR_{it} = 0$), the model ensured these grids did not contribute to the likelihood.

$$log(\lambda_{it}) = \alpha + R_t + U_i + \hat{\beta}_1 Z_{1it} + ... + \hat{\beta}_k Z_{kit} + log(PAR_{it}) \tag{6.2}$$

where λ_{it} represents the rate of notified cases in grid i during week t; α is the intercept (or fixed effect) representing the average per-person risk over New Zealand (assuming covariates at baseline); R_t represents the temporal component; U_i represents the spatial component; $\hat{\beta}_1, ..., \hat{\beta}_k$ are the posterior coefficient estimates from the posterior distributions $\beta_1, ..., \beta_k$; $Z_{1it}, ..., Z_{kit}$ are explanatory covariates, such as weather, demographic and environmental variables; and $\log(PAR_{it})$ is the offset of the population at risk in each grid i during week t.

The non-informative prior distributions for the beta coefficient of each weather and confounding variable (Z_{kit}) was set to a default value of $\beta_{Z_k} \sim N(0, 1/0.5)$.

As with the original Knorr-Held Richardson model [45], the temporal component (R_t) assumed a Gaussian second-order random walk prior (Equation 6.3), that is, the change in log-relative-risk from week t to week $t+1$ will be similar to the change in risk from week $t-1$ to week t [52, 73].

$$R_{t+1}|R_t, R_{t-1}, \kappa_R \sim R_t + (R_t - R_{t-1}) + N(0, \frac{1}{\kappa_R}) \tag{6.3}$$

where flat priors were assumed for R_{t-1} and R_t; and the hyperparameter (κ_R) was assumed to have a non-informative conjugate Gamma prior [52].

For the spatial component (U_i), an intrinsic Gaussian Markov random field prior was used which ensured grid i took the average value of its surrounding grids [52]. A queen neighbourhood adjacency matrix was created, providing information of the immediate neighbours for each grid i, which aids in the adjustment for the spatial correlation between grid i and its immediate neighbours. A conjugate Gamma prior was used for κ_U. (See Equation 6.4).

$$U_i \mid \{U_j : j \in n(i)\}, \kappa_U \sim \frac{1}{|n(i)|} \sum_{j \in n(i)} U_j + N\left(0, \frac{1}{\kappa_U |n(i)|}\right) \tag{6.4}$$

6.3.3 Past Association Models

In equation (6.2), the fixed effect (α) is the baseline or intercept, representing the log IR when all covariates are at their default values. The starting values for the fixed effect were estimated as the average IR of each disease per week. These differed between diseases and were calculated using the equation below:

$$\alpha = \log\left(\frac{N}{52 \times PAR}\right)$$

where α denotes the fixed effect; N denotes the average total number of cases of each disease per year over the study period; and PAR denotes the total population at risk in New Zealand.

The resulting starting values were:

- Campylobacteriosis α = -11.
- Cryptosporidiosis α = -13.
- Influenza hospitalisations α = -12.
- Meningococcal disease α = -14.

To achieve convergence, each model was run for 10,000 iterations sampling every 10th iteration, after an initial burn-in of 500. The posterior coefficient estimate ($\hat{\beta}_k$) is equal to the median of the posterior distribution of β_k produced from the MCMC iterations and were taken through into the projection modelling process (Chapter 7).

The significance of association required with the past IR of disease was a p-value ≤ 0.05, having adjusted for the effects of other explanatory variables. The 95% credible interval (C.I.) for each posterior coefficient estimate ($\hat{\beta}_k$) corresponding to each explanatory variable (Z_{kit}) was calculated as the 2.5th and 97.5th percentiles of the β_k posterior distribution.

The multiplicative effect of a one standard deviation change in each weather or continuous confounding variable on the incidence risk (IR) of each disease in New Zealand, having adjusted for the effects of other covariates and confounders in the model, is given by Equation 6.5.

$$\Delta\lambda_{it} = e^{\hat{\beta}_k} \tag{6.5}$$

where $\Delta\lambda_{it}$ denotes the change in rate of a notified case of disease in grid i during week t; and $\hat{\beta}_k$ denotes the posterior coefficient estimate of each covariate Z_{kit}.

The posterior spatial effects (U_i) represent the multiplicative difference in the IR of disease for grid i compared to the New Zealand average (i.e. e^{U_i}). A choropleth map of the posterior spatial effects (U_i) for each grid i was created and grouped into quartiles. They were visually appraised to check if there was evidence of spatial correlation or patterns that were not accounted for in the weather and confounding variables. For example, this could be related to missing covariates such as the effect of differing DHB reporting practices.

6.4 Results

6.4.1 Explanatory Variable Selection

The use of expert opinion alongside the screening results from Chapter 5, suggested the inclusion of the following weekly weather and confounding variables in the past association models for each disease:

- Campylobacteriosis: average absolute humidity, average rainfall, average temperature, age, animal density, drinking water quality, ethnicity and SDI.
- Cryptosporidiosis: average rainfall, average temperature, age, drinking water quality, ethnicity, rurality and SDI.
- Influenza Hospitalisations: average absolute humidity, average temperature, age, ethnicity, influenza vaccination coverage, rurality and SDI.
- Meningococcal Disease: average absolute humidity, average temperature, age, ethnicity, rurality and SDI.

6.4.2 Past Association Models

Campylobacteriosis

The weekly average absolute humidity was the only weather variable found to be significantly associated with the rate of notified campylobacteriosis cases in grid i during week t in New Zealand from 1997 - 2007, having adjusted for the effects of other covariates in the model ($\hat{\beta}_1$=0.143, p-value=0.000). It follows that for an increase of one standard deviation in the weekly average absolute humidity in grid i, the expected rate of campylobacteriosis notification in that grid changes by a factor of 1.154 ($\Delta\lambda_{it} = e^{0.143}$), when all other variables are held constant.

The confounders found to be significantly associated with the rate of notified cases in grid i during week t were the standardized proportion aged less than four years ($\hat{\beta}_4$=-0.067, p-value=0.008), the standardized proportion aged over sixty-five years ($\hat{\beta}_5$=0.183, p-value=0.000), the standardized beef density ($\hat{\beta}_6$=0.031, p-value=0.011), the standardized dairy density ($\hat{\beta}_7$=0.042, p-value=0.000), intermediate drinking water quality ($\hat{\beta}_9$=0.056, p-value=0.000), poor drinking water quality ($\hat{\beta}_{10}$=0.089, p-value=0.000), the standardized proportion who identify with Maori ethnicity ($\hat{\beta}_{11}$=-0.404, p-value=0.000), the standardized proportion who identify with Pacific Islander ethnicity ($\hat{\beta}_{13}$=0.063, p-value=0.002) and the standardized proportion who identify with Asian ethnicity ($\hat{\beta}_{14}$=0.024, p-value=0.002).

An example of the interpretation of the effect of significant confounders on the rate of notified cases in grid i during week t can be illustrated with the proportion of those ages less than four years in grid i. The posterior coefficient estimate, $\hat{\beta}_4$=-0.067, defines the magnitude of the multiplicative change in the rate of notified cases in grid i, given by $\Delta\lambda_{it} = e^{-0.067} = 0.935$. This indicates that for one standard deviation increase in the proportion of those aged less than four years (i.e. an increase of 2.1%) at the expense of the proportion aged between four and 65 years, the rate of notified cases of campylobacteriosis changes by a factor of 0.935, assuming all other covariates remain constant (p-value=0.008). The 95% credible interval states that we are 95% sure the true change in rate of notified cases lies between 0.908 and 0.972, assuming all other variables remain constant. This infers that an increase in the proportion of those aged less than four years in grid i decreases the rate of notified cases of campylobacteriosis in that grid.

An example of how this can be applied to the weekly IR of campylobacteriosis in New Zealand can be illustrated with the use of VCS grid 30341 in Manawatu. The proportion aged less than four years old in 2007 was approximately 6.8%, the average weekly IR of notified campylobacteriosis was 0.52 cases and the PAR residing in the grid was approximately 13,109 people. An increase of one standard deviation in the proportion of those aged less than four years in 2007 to 8.9% at the expense of the proportion aged between four and 65 years, suggests that the weekly campylobacteriosis IR will change by a factor of 0.935, resulting in an average of 0.49 notified cases per week. The 95% credible interval estimates the true number of weekly notified cases would lie within the range of 0.47 and 0.51 notified cases.

For more detailed results for the past association modelling of campylobacteriosis see Table 6.1.

Cryptosporidiosis

The weekly average rainfall and weekly average temperature were found to be significantly associated with the rate of notified cryptosporidiosis cases in grid i in New Zealand, having adjusted for the effects of other covariates in the model ($\hat{\beta}_1$=-0.094, p-value=0.002; and $\hat{\beta}_2$=0.124, p-value=0.004, respectively). This indicates that an increase of one standard deviation in the average weekly temperature in grid i results in a change in the rate of cryptosporidiosis notifications by a factor of 0.910 (i.e. $\Delta\lambda_{it} = e^{-0.094}$), when all other covariates are held constant. Likewise, for an increase of one standard deviation in weekly average rainfall in grid i, the

Explanatory Variable, Z_i	$\hat{\beta}_k$	$e^{\hat{\beta}_k}$ (95% C.I.)	p-value
Average Absolute Humidity	0.143	1.154 (1.135,1.173)	0.000
Average Rainfall	0.007	1.007 (0.991, 1.022)	0.387
Average Temperature	0.012	1.012 (0.978,1.064)	0.505
Aged < 4 Years	-0.067	0.935 (0.908,0.972)	0.008
Aged > 65 Years	0.183	1.201 (1.100,1.326)	0.000
Beef Density	0.031	1.032 (1.012,1.051)	0.011
Dairy Density	0.042	1.043 (1.017,1.073)	0.000
Sheep Density	0.006	1.006 (0.968,1.036)	0.678
Intermediate Drinking Water Quality	0.056	1.057 (1.028,1.094)	0.000
Poor Drinking Water Quality	0.089	1.093 (1.062,1.135)	0.000
Unknown Drinking Water Quality	0.048	1.051 (0.994,1.190)	0.080
Maori	-0.404	0.667 (0.611,0.887)	0.000
Pacific Islander	0.063	1.065 (1.007,1.106)	0.002
Asian	0.024	1.024 (1.005,1.069)	0.002
Middle Eastern/Latin American/African	0.027	1.027 (0.993,1.066)	0.131

Table 6.1: The past association modelling results for campylobacteriosis for each explanatory variable (Z_i), its associated posterior coefficient estimate ($\hat{\beta}_k$), the expected change in the rate of notified case ($e^{\hat{\beta}_k}$) with the corresponding 95% credible interval, and level of significance (p-value). The baseline confounding variables are the proportion in each grid who are aged between 4 and 65 years of age, grids with good drinking water quality and the proportion in each grid who identify with European and other ethnicity (including New Zealand European).

rate of cryptosporidiosis notifications changes by a factor of 1.132 ($\Delta\lambda_{it} = e^{0.124}$), once again when all other covariates are held constant.

The confounders found to be significantly associated with the rate of notified cases of cryptosporidiosis in grid i during week t, were poor drinking water quality ($\hat{\beta}_6$=0.102, p-value=0.048), unknown drinking water quality ($\hat{\beta}_7$=-0.350, p-value=0.002), the standardized proportion who identify with Maori ethnicity ($\hat{\beta}_8$=-0.340, p-value=0.000) and those living in rural areas ($\hat{\beta}_{12}$=0.451, p-value=0.000).

For more detailed results for the past association modelling of cryptosporidiosis, see Table 6.2.

Influenza Hospitalisations

There were no weather variables found to be significantly associated with influenza hospitalisations in New Zealand, having adjusted for the effects of other covariates in the model.

The proportion who identify with Maori ethnicity in grid i during week t was the only confounder found to be significantly associated with the rate of hospitalisations where influenza was the primary diagnosis in New Zealand ($\hat{\beta}_5$=0.197, p-value=0.008).

For more detailed results for the past association modelling of influenza hospitalisations,

Explanatory Variable, Z_i	$\hat{\beta}_k$	$e^{\hat{\beta}_k}$ (95% C.I.)	p-value
Average Rainfall	-0.094	0.910 (0.864,0.958)	0.002
Average Temperature	0.124	1.132 (1.047,1.214)	0.004
Aged < 4 Years	0.162	1.175 (1.095,1.258)	0.000
Aged > 65 Years	-0.029	0.971 (0.899,1.032)	0.362
Intermediate Drinking Water Quality	0.069	1.072 (0.980,1.178)	0.149
Poor Drinking Water Quality	0.102	1.108 (1.003,1.219)	0.048
Unknown Drinking Water Quality	-0.350	0.705 (0.619,0.811)	0.002
Maori	-0.340	0.711 (0.636,0.831)	0.000
Pacific Islander	0.003	1.003 (0.955,1.065)	0.987
Asian	0.015	1.015 (0.987,1.043)	0.392
Middle Eastern/Latin American/African	-0.029	0.971 (0.856,1.068)	0.731
Rural	0.451	1.570 (1.280,1.868)	0.000
SDI (baseline 0)	0.020	1.020 (0.945,1.104)	0.644

Table 6.2: The past association modelling results for cryptosporidiosis for each explanatory variable (Z_i), its associated posterior coefficient estimate ($\hat{\beta}_k$), the expected change in rate of notified cases ($e^{\hat{\beta}_k}$) with the corresponding 95% credible interval, and level of significance (p-value). The baseline confounding variables are the proportion in each grid who are aged between 4 and 65 years of age, grids with good drinking water quality, the proportion in each grid who identify with European and other ethnicity (including New Zealand European), grids in urban areas and a SDI grid score of 0.

see Table 6.3.

Meningococcal Disease

There were no weather variables found to be significantly associated with meningococcal disease in New Zealand, having adjusted for the effects of other covariates in the model.

The confounders found to be significantly associated with the rate of notified cases of meningococcal disease in grid i during week t, were the standardised proportion of those aged less than four years ($\hat{\beta}_3$=0.133, p-value=0.017), the standardized proportion who identify with Maori ethnicity ($\hat{\beta}_5$=0.169, p-value=0.000), the standardized proportion who identify with Asian ethnicity ($\hat{\beta}_7$=-0.079, p-value=0.044) and the social deprivation index (SDI) score of the grid ($\hat{\beta}_{10}$=0.226, p-value=0.000).

For more detailed results for the past association modelling of meningococcal disease, see Table 6.4.

6.4.3 Posterior Spatial Effects

The posterior spatial effects were mapped to show any multiplicative difference in the IR of disease for grid i compared to the New Zealand average (i.e. e^{U_i}). The boundaries on each map represent the twenty-one DHBs in New Zealand.

Explanatory Variable, Z_i	$\hat{\beta}_k$	$e^{\hat{\beta}_k}$ (95% C.I.)	p-value
Average Absolute Humidity	0.046	1.047 (0.757,1.123)	0.918
Average Temperature	-0.127	0.88 (0.776,1.118)	0.541
Aged < 4 Years	0.140	1.151 (0.915,1.234)	0.297
Aged > 65 Years	0.008	1.008 (0.926,1.311)	0.528
Maori	0.197	1.218 (1.076,1.602)	0.008
Pacific Islander	-0.007	0.993 (0.809,1.073)	0.756
Asian	-0.081	0.922 (0.891,1.063)	0.815
Middle Eastern/Latin American/African	0.025	1.026 (0.806,1.074)	0.960
Influenza Vaccination Coverage	0.025	1.026 (0.860,1.082)	0.918
Rural	0.077	1.079 (0.318,1.284)	0.743
SDI (baseline 0)	0.215	1.24 (0.993,1.375)	0.067

Table 6.3: The past association modelling results for influenza hospitalisations for each explanatory variable (Z_i), its associated posterior coefficient estimate ($\hat{\beta}_k$), the expected change in rate of hospitalisation due to influenza ($e^{\hat{\beta}_k}$) with the corresponding 95% credible interval, and level of significance (p-value). The baseline confounding variables are the proportion in each grid who are aged between 4 and 65 years of age; the proportion in each grid who identify with European and other ethnicity (including New Zealand European); grids in urban areas; and a SDI grid score of 0.

Campylobacteriosis

The posterior spatial effects (U_i) corresponding to the past association model for the IR of notified campylobacteriosis from 1997 - 2007 were mapped and shown in Figure 6.1a. The map suggests that the IR of notified campylobacteriosis is higher than the national average in the North Island and east of the South Island. The IR is lower than the national average in the west of the South Island.

Cryptosporidiosis

The posterior spatial effects (U_i) corresponding to the past association model for the IR of notified cryptosporidiosis from 1997 - 2007 were mapped and shown in Figure 6.1b. The map suggests that the IR of notified cryptosporidiosis is predominantly higher than the national average in the North Island and east of the South Island. The IR is lower than the national average in Northland, Bay of Plenty and the west of the South Island.

Influenza Hospitalisations

The posterior spatial effects (U_i) corresponding to the past association model for the IR of influenza hospitalisations from 1997 - 2007 were mapped and is shown in Figure 6.1c. The map suggests that the IR of influenza hospitalisations in the top half of the North Island

Explanatory Variable, Z_i	$\hat{\beta}_k$	$e^{\hat{\beta}_k}$ (95% C.I.)	p-value
Average Absolute Humidity	0.048	1.049 (0.970,1.153)	0.257
Average Temperature	-0.117	0.889 (0.778,1.070)	0.166
Aged $<$ 4 Years	0.133	1.142 (1.030,1.256)	0.017
Aged $>$ 65 Years	-0.004	0.996 (0.930,1.072)	0.983
Maori	0.169	1.185 (1.056,1.336)	0.000
Pacific Islander	0.000	1.000 (0.986,1.016)	0.989
Asian	-0.079	0.924 (0.901,0.996)	0.044
Middle Eastern/Latin American/African	0.024	1.024 (0.931,1.080)	0.472
Rural	0.075	1.077 (0.892,1.316)	0.450
SDI (baseline 0)	0.226	1.254 (1.149,1.370)	0.000

Table 6.4: The past association modelling results for meningococcal disease for each explanatory variable (Z_i), its associated posterior coefficient estimate ($\hat{\beta}_k$), the expected change in rate of meningococcal disease notifications ($e^{\hat{\beta}_k}$) with the corresponding 95% credible interval, and level of significance (p-value). The baseline confounding variables are the proportion in each grid who are aged between 4 and 65 years of age; the proportion in each grid who identify with European and other ethnicity (including New Zealand European), grids in urban areas and a SDI grid score of 0.

is predominantly higher than the national average and approximately equal to the national average in the bottom half, with the Tairawhiti DHB estimated to be lower than the national average. The map suggests a decreased risk in the top half of the South Island and an increased risk at the bottom of the South Island, with a band across the bottom of the South Canterbury and West Coast DHBs and the top of the Otago and Southland DHBs suggesting no difference in IR of influenza hospitalisations compared to the national average.

Meningococcal Disease

The posterior spatial effects (U_i) corresponding to the past association model for the IR of notified meningococcal disease from 1997 - 2007 were mapped and shown in Figure 6.1d. The map suggests that the IR of notified meningococcal disease in the North Island is predominantly higher than the national average, with the Tairawhiti DHB estimated to be lower than the national average. The map suggests a decreased risk in the top half of the South Island and an increased risk at the bottom of the South Island, with a band accross the bottom of the South Canterbury and West Coast DHBs and the top of the Otago and Southland DHBs suggesting no difference in IR of meningococcal disease compared to the national average.

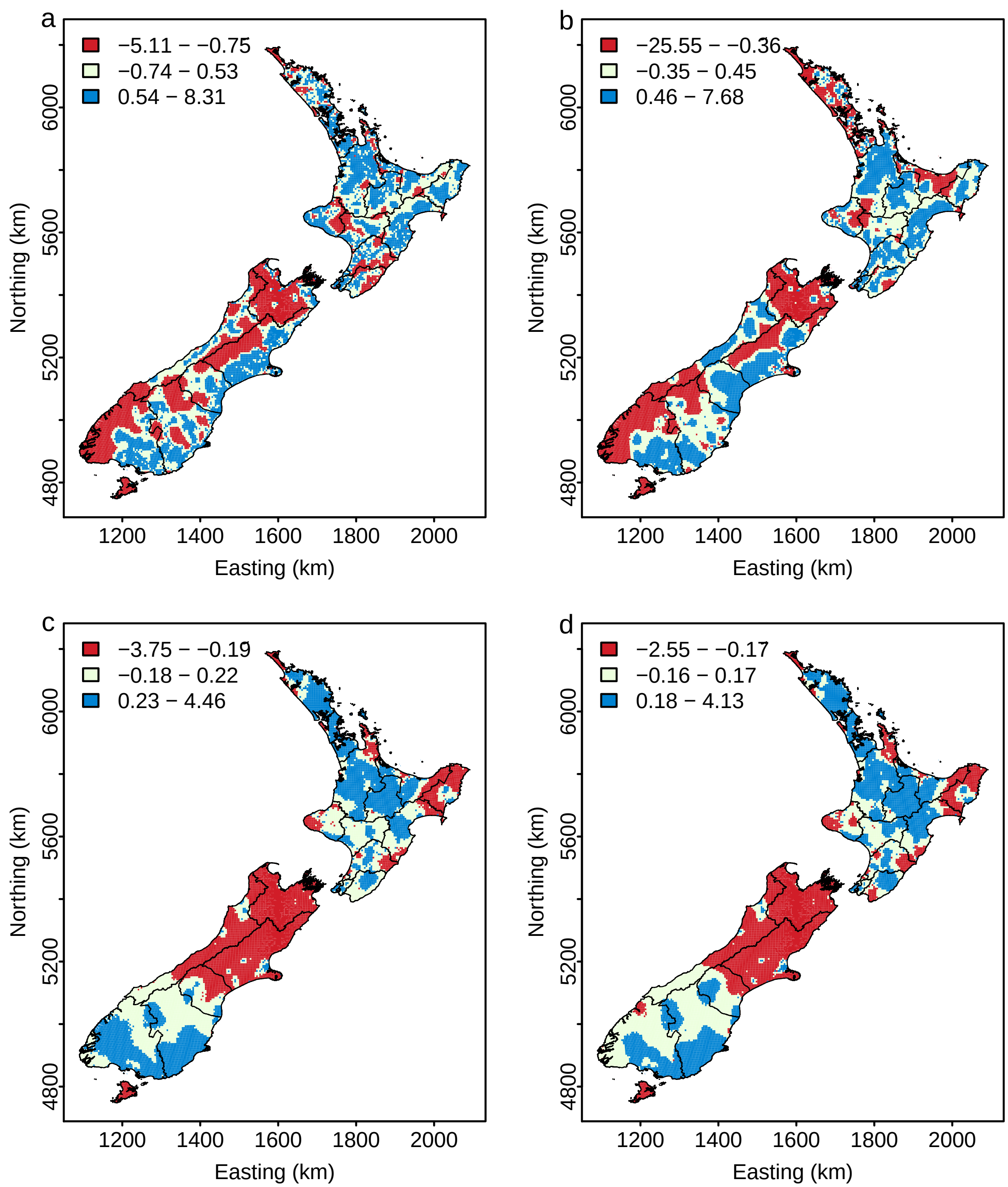


Figure 6.1: Posterior spatial effects (U_i) for the incidence risk of a) campylobacteriosis, b) cryptosporidiosis, c) influenza hospitalisations, and d) meningococcal disease in New Zealand from 1997 - 2007.

Chapter 7

Projection Calculations

7.1 Introduction

The aim of this chapter is to describe the method used to project the percentage change in incidence risk (IR) of each indicator disease in New Zealand for the years 2015, 2040 and 2090. This was carried out using a combination of 3 IPCC emission scenarios (A2, B2 and A1B) and 12 downscaled global climate models.

7.2 Data

The climate change projection data for the weather variables were provided by Dr. Andrew Tait of NIWA [74]. The projection data were prepared in two steps. Firstly the methods used by Dr. Andrew Tait are described, followed by a description of the necessary data manipulations carried out on each weather variable to be in the correct format for projection purposes.

The projected average rainfall data were provided by NIWA in the format of the total change in rainfall (mm) per season for each emission scenario. Producing this data was a step-wise process carried out by Dr. Andrew Tait, as outlined below:

1. The monthly percentage change in rainfall for 2030 - 2049 (with midpoint reference year 2040) and 2080 - 2099 (midpoint reference year 2090) was calculated by Dr. Brett Mullan of NIWA for the New Zealand Climate Change Centre (NZCCC) Adaptation Conference in 2009 from 12 global climate models. NIWA statistically downscaled these projections to the 11,491 VCS grids spanning New Zealand used for the HAIFA project.

2. The mean total change in rainfall for the years 2040 and 2090 were calculated from the projected percentage monthly change and the 1980 - 1999 mean monthly total change.

3. The seasonal total change in rainfall was then calculated for 2040 and 2090.

4. The climate change projections were then set with respect to the 11 year study period, spanning 1997 - 2007, and the total change in mean seasonal rainfall for this base period was calculated (midpoint reference year 2002).

5. The seasonal total change in rainfall from 2002 to 2040 and to 2090 was then calculated for each emission scenario (A2, B1 and A1B).

6. The total change in rainfall for 2015 was calculated by linear interpolation, where total changes in rainfall between 1997 - 2007 (midpoint reference year 2002) and 2030 - 2049 (midpoint reference year 2040) were multiplied by 13/38.

For projection purposes, the projected seasonal change in rainfall needed to be converted into a weekly measurement. This was achieved by dividing the projected seasonal change by 52/4, as there are approximately 52/4 weeks in each season.

The projected average temperature data were provided in the format of the total change in temperature (°C). The step-wise process carried out by Dr. Andrew Tait is outlined below:

1. Starting with the monthly temperature projections calculated by Dr. Brett Mullan for the NZCCC Adaption Conference for 2030 - 2049, the mean seasonal temperature changes for the 11 year study period (1997 - 2007) were calculated.

2. The seasonal differences between the original baseline for the climate change projection data from 1980 - 1999 and the required baseline for the study period 1997 - 2007 were calculated. The difference between these was then subtracted from the projected increases in temperature for 2040 and 2090.

3. To derive the temperature changes in 2015, the projected temperature changes between 1997 - 2007 (midpoint reference year 2002) and 2030 - 2049 (midpoint reference year 2040) were multiplied by 13/38, i.e. a linear interpolation.

The projected change in temperature data provided from Dr. Andrew Tait were in the desired format for the projection models, hence no data manipulations were required.

The average absolute humidity projections were provided as the seasonal percentage change in absolute humidity (kg/m^3) per degree change in temperature (°C). The step-wise process carried out by Dr. Andrew Tait and Dr. Abha Sood of NIWA is outlined below:

1. The projection data are based on the A2 emission scenario, where the change in absolute humidity and temperature from 1971 - 2000 to 2071 - 2100 were derived from NIWA's Regional Climate Model.

2. The percentage change in absolute humidity per degree warming (temperature increase) was calculated for each season. The data for the years 2015, 2040 and 2090 under each emission scenario still required calculation.

For projection purposes, the percentage change in absolute humidity per season for each year (2015, 2040 and 2090) under each emission scenario was required. The projected percentage change in absolute humidity was linked to each temperature change projection for 2015, 2040 and 2090 under each emission scenario. This was achieved using the following equation:

$$\Delta \hat{AH}_{cit} = \frac{\Delta AH_{it}}{100\%} \times \Delta \hat{T}_{cit} \times \bar{AH}$$

where $\Delta \hat{AH}_{cit}$ denotes the projected percentage change in absolute humidity for grid i and season t of the three future years under emission scenario c; ΔAH_{it} denotes the projected change in absolute humidity per degree warming provided by Dr. Andrew Tait; $\Delta \hat{T}_{cit}$ denotes the projected seasonal change in temperature for 2015, 2040 and 2090 under each emission scenario; and $\bar{AH}$ denotes the average absolute humidity for each season from 1997 - 2007.

The average absolute humidity for each season spanning 1997 - 2007 were as follows:

- Summer: $\bar{AH} = 0.00936\ kg/m^3$.
- Autumn: $\bar{AH} = 0.00824\ kg/m^3$.
- Winter: $\bar{AH} = 0.00633\ kg/m^3$.
- Spring: $\bar{AH} = 0.00775\ kg/m^3$.

Projection data for confounders in the past association models were assumed to stay constant between the study period and the projection years. The reasons for this were two fold. Firstly, projection data was not available for the confounding variables, and secondly, creating these projections would introduce additional uncertainty about our disease projections. There were limited projections for the population but these were not of a sufficient spatial scale to be useful.

All data manipulations and statistical analyses in this chapter were performed using R version 2.12.1 [62].

7.3 Method

The projection calculations were used to estimate the percentage change in IR for the years 2015, 2040 and 2090. The past association models outlined in Chapter 6 were taken to be the model of best choice for each disease, modelling the association between infectious disease incidence and past climate in New Zealand from 1997 - 2007. The projection calculation utilises the posterior coefficient estimates ($\hat{\beta}_k$) for each weather and confounding variable produced from the modified Knorr-Held Richardson models, outlined in Chapter 6.3.3. The weather and confounding variable selection utilised in the past association models was not modified for the projection model, therefore variables found to be insignificantly associated with the IR of disease (p-value > 0.05) having adjusted for the other covariates, were not excluded from the projection calculation. The calculations were run for each of the 3 climate scenarios provided by NIWA, which were a combination of 3 IPCC emission scenarios: the high A2, medium A1B and low B1; and an average of 12 NIWA downscaled global climate models.

The second aim of this project was to use the identified effect of climate variation on four indicator diseases throughout New Zealand from 1997 - 2007, at a fine spatial and temporal level, to calculate the percentage change in incidence risk of disease in the years 2015, 2040 and 2090. As the emphasis of this project is on the effect of climate variation, it was assumed the demographic, environmental and other confounders would remain constant with respect to time.

7.3.1 Projection Calculation

The posterior coefficient estimates ($\hat{\beta}_k$) for the standardized covariates (Z_k) identified in the past association models, were used to calculate the percentage change in IR for each disease in 2015, 2040 and 2090 for each grid i, under each emission scenario (Equation 7.2). The rational for the projection equation is defined below:

During the past association modelling, the count of disease in each grid-week, Y_{it}, was assumed to follow a Poisson distribution with mean $PAR_{it}\lambda_{it}$, where PAR_{it} represents the population at risk in grid i, week t, and the outcome variable λ_{it} represents the rate of notified cases of disease in grid i during week t. The extended Knorr-Held Richardson from Chapter 6 was of the form:

$$\log\left(\frac{\lambda_{it}}{PAR_{it}}\right) = \alpha + R_t + U_i + \sum_{i=1}^{n} \beta_k Z_{kit}$$

where λ_{it} denotes the rate of notified cases in grid i during week t; PAR_{it} is the population at risk offset; α denotes the intercept; R_t denotes the temporal component; U_i denotes the spatial component; $\hat{\beta}_k$ denotes the posterior coefficient estimate; and Z_{kit} denotes the kth of n standardized covariates (weather variables and other confounders).

As described in Equation 5.1 (Chapter 5.2.1), standardising the covariates (W_k) were carried out using the following relationship:

$$Z_{kit} = \frac{W_{kit} - \bar{W}_k}{s.e.(W_k)}$$

where Z_{kit} is the standardized version of covariate W_k, in the grid i, week t; $\bar{W}_k$ is the mean and $s.e.(W_k)$ is the standard error for the covariate W_k over the 11 year study period.

This, along with the law of logarithms, implies that

$$IR_{it} \equiv \frac{\lambda_{it}}{PAR_{it}} = \exp\left(\alpha_i + R_t + U_i + \sum_{i=1}^{n} \hat{\beta}_k \frac{W_{kit} - \bar{W}_k}{s.e.(W_k)}\right)$$

To calculate the percentage change in IR ($\Delta \tilde{IR}_{it}$), this involves taking the ratio of the projected incidence risk (IR^*_{it}) and the IR over the study period (IR_{it}), for each grid i. Let $IR_{it} = \dfrac{\lambda_{it}}{PAR_{it}}$ denote the IR from 1997 - 2007 and $\hat{IR}_{it*} = \dfrac{\lambda_{it*}}{PAR_{it*}}$ denote the projected incidence risk for grid i at future time $t*$. Therefore, assuming no change in PAR:

$$\frac{\hat{IR}_{it*}}{IR_{it}} = \exp\left(\left[\sum_{k=1}^{n}(\hat{\beta}_k \frac{(W_{kit} + \hat{w}_{kit}) - \bar{W}_k}{s.e.(W_k)}) + U_{i*} + R_{t*}\right] - \left[\sum_{k=1}^{n} \hat{\beta}_k \frac{W_{kit} - \bar{W}_k}{s.e.(W_k)} + U_i + R_t\right]\right)$$

where $\frac{IR_{it*}}{IR_{it}}$ denotes the change in incidence risk; $\hat{\beta}_k$ denotes the posterior coefficient estimate for the weather variables; $\hat{w_{kit}}$ denotes the projected change in weather variable k; U_i and R^*_t denote the projected spatial and temporal components for grid i in future time $t*$. U_i and R_t denote the spatial and temporal components spanning the 11 year study period; and $\bar{W}_k$ denotes the mean and $s.e.(W_k)$ denotes the standard error for the covariates W_k over the 11 year study period.

As indicated earlier, non-weather covariates (confounders) were assumed to remain constant in relation to future projections (i.e. the change in weather, $\hat{w}_{kit} = 0$), thus eliminating them from the equation. For the purposes of projection, the spatial and temporal effects (U_i and R_t, respectively) were incorporated implicitly, as they are assumed to remain constant over time. Hence, it was assumed that $\exp(U^*_i - U_i)$ equals 1, as with $\exp(R^*_t - R_t)$. This resulted in the projected change in IR_{it} given by

$$\frac{IR_{it*}}{IR_{it}} = \exp\left(\sum_{k=1}^{m} \frac{\hat{\beta}_k}{s.e.(W_k)} \hat{w_{kit}}\right) \tag{7.1}$$

where $\frac{IR_{it*}}{IR_{it}}$ denotes the projected change in IR; $\hat{\beta}_k$ denotes the posterior coefficient estimate for m weather variables; $s.e.(W_k)$ denotes the standard error of the weather variable W_k over the 11 year study period; and $\hat{w}_{kit}$ denotes the projected change in weather variable W_k.

The projected seasonal percentage change in IR_{it} for grid i, year t under each emission scenario, is given by Equation 7.2:

$$\Delta \tilde{IR}_{it} = \exp\left(\sum_{k=1}^{m} (\tilde{\beta}_k \hat{w}_{kit}) - 1\right) \times 100\% \tag{7.2}$$

where $\Delta \tilde{IR}_{it}$ denotes the projected percentage change in IR; $\hat{\beta}_k$ denotes the projection coefficient $\left(\text{i.e. } \tilde{\beta}_k = \frac{\hat{\beta}_k}{s.e.(W_k)}\right)$; and $\hat{w}_{kit}$ denotes the projected change in weather variable W_k.

The results were displayed on choropleth maps representing the projected percentage change in disease incidence risk in each grid (i) spanning New Zealand. For each projection year and climate scenario combination, there were 4 maps produced representing the four seasons (summer, autumn, winter and spring). The percentage change in IR of each disease for each projection year is divided into quartiles resulting in the three emission scenarios (A2, A1B and B2) having the same breaks for each year. This aids in comparing the magnitude of change projected between the emission scenarios.

7.4 Results

7.4.1 Campylobacteriosis

The estimated coefficients for the weather variables in the projection calculation are shown in Table 7.1. The weekly weather variable, average absolute humidity was found to be significantly associated with the IR campylobacteriosis in New Zealand from 1997 - 2007. The weather covariates, average weekly rainfall and temperature, were found to be insignificantly associated with the IR of campylobacteriosis however are included in the projection calculation as they represent an unbiased estimate for the future incidence.

Campylobacteriosis Projections for 2015

A2 Climate Scenario

The seasonal projected percentage changes in campylobacteriosis IR throughout New Zealand

Weather Covariate (W_k)	$\hat{\beta}_k$	s.e.(W_k) (1997 - 2007)	$\tilde{\beta}_k$	p-value
Average Absolute Humidity	0.1429	0.0017	85.0448	0.000
Average Rainfall	0.0069	7.6419	0.0009	0.387
Average Temperature	0.0118	4.5039	0.0026	0.505

Table 7.1: The campylobacteriosis projection calculation estimate for each weather variable (W_k), the posterior coefficient estimates ($\hat{\beta}_k$), the standard error of the weather variables throughout the study period (s.e.(W_k)), the coefficient estimate for the projection calculation ($\tilde{\beta}_k$), and the level of significance (p-value).

in 2015 under the A2 climate scenario are mapped in Figure 7.1. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR throughout New Zealand predominantly lies between 1.19% to 4.92%.

- In autumn, the change in IR ranges between 0.48% and 1.18% throughout the most part of New Zealand, with the Nelson-Marlborough, Canterbury and Southland DHBs projected to change between -1.40% and 0.47%.

- In winter, IR changes between -1.40% and 0.75% are seen throughout New Zealand, with the Hawke's Bay, Tairawhiti, West Coast and Otago DHBs IR projected to change between 0.76% and 1.18%.

- In spring, the projected change in IR ranges from -1.40 to 0.75% throughout the majority of New Zealand, with the projected change in the Hawke's Bay, Tairawhiti, West Coast and Otago DHBs ranging from 0.76% to 1.18%.

A1B Climate Scenario

The seasonal projected percentage changes in campylobacteriosis IR throughout New Zealand in 2015 under the A1B climate scenario are mapped in Figure 7.2. The projected percentage changes for each season are outlined below:

- In summer, the change in IR throughout the North Island and the West Coast, Canterbury, South Canterbury and Otago DHBs range from 1.19% to 4.92%. Parts of the Nelson Marlborough and Southland DHBs IR is projected to change between -1.40% and 0.47%.

- In autumn, the projected change in IR in the North Island ranges from 0.48% to 1.18%. The projected change in IR in the South Island ranges between -1.40% and 0.75%, and from 1.19% to 4.92% in the South Canterbury DHB.

- In winter, a projected change between 0.48% and 1.18% is suggested throughout New Zealand, with a change between -1.40% and 0.47% in the Hawke's Bay, Tairawhiti, Nelson-Marlborough, Canterbury and Southland DHBs, and from 1.19% to 4.92% in the West Coast and South Canterbury DHBs.

- In spring, the projected change in IR ranges between 0.76% and 1.18% in the North Island, and -1.40% to 0.75% in the South Island with 1.19% to 4.92% in the West Coast DHB.

B1 Climate Scenario
The seasonal projected percentage changes in campylobacteriosis IR throughout New Zealand in 2015 under the B1 climate scenario are mapped in Figure 7.3. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR in the North Island ranges from 1.19% to 4.92%, with 0.76% to 1.18% in the Mid Central DHB. The South Island projected change ranges between 0.76% and 4.92%, with -1.40% to 0.75% projected in parts of the Nelson-Marlborough, Canterbury and Southland DHBs.

- In autumn, an IR change of -1.40% to 0.47% is projected throughout New Zealand, with 0.48% to 0.47% projected in the Tairawhiti, Hawke's Bay, Wairarapa, West Coast and South Canterbury DHBs.

- In winter, the projected change in IR in the North Island ranges from 0.48% to 0.75%. In the Tairawhiti, Nelson-Marlborough, Canterbury and Southland DHBs the projected change ranges from -1.40% to 0.47%, and in the West Coast, Otago and Southland DHBs ranges from 0.76% to 4.92%.

- In spring, the suggested change in IR ranges between 0.48% and 1.18% in the North Island and between -1.40% and 0.75% in the South Island, with 0.76% to 1.18% projected in the West Coast DHB.

Campylobacteriosis Projections for 2040

A2 Climate Scenario
The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2040 under the A2 climate scenario are mapped in Figure 7.4. The projected percentage changes for each season are outlined below:

- In summer, the expected change in IR throughout New Zealand predominantly ranges from 3.64% to 15.06%.

- In autumn, the projected changes in IR suggest the North Island and the South Canterbury and West Coast DHBs range from 1.43% to 3.63%. The changes in the Nelson-Marlborough, Canterbury and Southland DHBs range from -4.10% to 1.42%, and from 3.64% to 15.06% in the South Canterbury DHB.

- In winter and spring, the projected changes in the North Island range from 1.43% to 3.63%, with the changes projected in the Tairawhiti and Hawke's Bay DHBs ranging between -4.10% and 1.42%. Projected IR changes between -4.10% and 2.27% are suggested in the Nelson-Marlborough, Canterbury, Otago and Southland DHBs, and from 2.28% to 15.06% in the South Canterbury and West Coast DHBs.

A1B Climate Scenario

The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2040 under the A1B climate scenario are mapped in Figure 7.5. The projected percentage changes for each season are outlined below:

- In summer, an IR change ranging from 3.64% to 15.06% is suggested in the North Island, and from 2.28% to 15.06% in the South Island.

- In autumn, the projected change in the North Island ranges from 2.28% to 3.63%. In the Nelson-Marlborough, Canterbury and Southland DHBs the change ranges between -4.10% and 1.42% and between 2.28% and 15.06% in the West Coast, South Canterbury and Otago DHBs.

- In winter, the change in IR is suggested to range between 1.43% and 2.27% throughout New Zealand, with a change between -4.10% and 1.42% in the Northland, Tairawhiti, Auckland, Hawke's Bay, Mid Central, Nelson-Marlborough and Canterbury DHBs, and between 2.28% and 3.63% in the West Coast and South Canterbury DHBs.

- In spring, the suggested change in campylobacteriosis IR in the North Island predominantly ranges from 2.28% to 3.63%. The change in the South Island ranges from -4.10% to 2.27%, and 3.64% to 15.06% in West Coast DHB.

B1 Climate Scenario

The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2040 under the B1 climate scenario are mapped in Figure 7.6. The projected percentage changes for each season are outlined below:

- In summer, the suggested changes in IR throughout New Zealand range predominantly between 2.28% and 15.06%, with parts of the Nelson-Marlborough and Southland DHBs projected change between -4.10% and 1.42%.
- In autumn, the change in New Zealand predominantly ranges from -4.10% to 1.42%, with a change in IR ranging between 1.43% and 3.63% in the Tairawhiti, Hawke's Bay, Whanganui, Wairarapa, Capital and Coast, South Canterbury and the West Coast DHBs.
- In winter, the suggested change in IR in the North Island predominantly ranges from 1.43% to 2.27%. A projected change ranging from -4.10% to 1.42% is suggested in the Tairawhiti, Hawke's Bay, Nelson-Marlborough, Canterbury and Southland DHBs, and the West Coast, South Canterbury and Otago DHBs are projected to change between 1.43% and 3.63%.
- In spring, the projected change in IR in the North Island ranges from 1.43% to 3.63%. The range in the South Island predominantly lies between -4.10% and 2.27%, with 2.28% to 3.63% projected in the West Coast DHB.

Campylobacteriosis Projections for 2090

A2 Climate Scenario

The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2090 under the A2 climate scenario are mapped in Figure 7.7. The projected percentage changes for each season are outlined below:

- In summer, the projected change in campylobacteriosis IR in New Zealand ranges predominantly between 6.92% and 30.10%.
- In autumn, the suggested change in IR in the North Island ranges predominantly between 9.47% and 30.10%, and from 6.92% to 30.10% in the South Island.
- In winter, the suggested projected change in the North Island ranges between 6.92% and 9.46%, with a range between 4.72% and 6.91% projected in the Northland, Tairawhiti, Bay of Plenty and Hawke's Bay DHBs. The projected change in IR in the South Island ranges from 4.72% to 9.46%, with 9.47% to 30.10% projected in the West Coast DHB.

- In spring, the projected change in IR suggested in the North Island ranges from 6.92% to 9.46%, and from 4.72% to 6.91% in the South Island, with 6.92% to 30.10% in the West Coast DHB.

A1B Climate Scenario

The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2090 under the A1B climate scenario are mapped in Figure 7.8. The projected percentage changes for each season are outlined below:

- In summer, the change in projected IR in the North Island lies between 9.47% and 30.10%, and between 6.92% and 30.10% in the South Island.

- In autumn, the suggested change in New Zealand ranges predominantly between 4.72% and 9.46%.

- In winter, parts of the Tairawhiti, Hawke's Bay and Canterbury DHBs are projected to change between -1.37% and 4.71%. In parts of the West Coast DHB the projected change in IR range from 9.47% to 30.10%. The projected change in IR throughout the rest of New Zealand ranges between 4.72% and 6.91%.

- In spring, the change in IR in the North Island ranges from 4.72% and 9.46%. The projected in the South Island ranges from -1.37% to 6.91% and from 6.92% to 9.46% in the West Coast DHB.

B1 Climate Scenario

The seasonal projected percentage change in campylobacteriosis IR throughout New Zealand in 2090 under the B1 climate scenario are mapped in Figure 7.9. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR in the North Island predominantly lies between 6.92% and 30.10%. The projections in the South Island range between -1.37% and 4.71%, with 4.72% to 6.91% in the Mid Central, South Canterbury and West Coast DHBs.

- In autumn, the changes in IR projected throughout New Zealand predominantly range between -1.37% and 4.71%, with a change between 4.72% and 6.91% projected in the Tairawhiti, Hawke's Bay, Wairarapa, Capital and Coast, South Canterbury and West Coast DHBs.

Weather Covariate (W_k)	$\hat{\beta}_k$	s.e.(W_k) (1997 - 2007)	$\tilde{\beta}_k$	p-value
Average Rainfall	-0.0943	7.6419	-0.0123	0.002
Average Temperature	0.1244	4.5039	0.0276	0.004

Table 7.2: The cryptosporidiosis projection calculation estimate for each weather variable (W_k), the posterior coefficient estimates ($\hat{\beta}_k$), the standard error of the weather variables throughout the study period (s.e.(W_k)), the coefficient estimate for the projection calculation ($\tilde{\beta}_k$), and the level of significance (p-value).

- In winter, the suggested change in campylobacteriosis IR in New Zealand predominantly ranges from -1.37% to 4.71%, with the change in the West Coast DHB between 4.72% and 9.46%.
- In spring, the projected change in IR ranges between -1.37% and 4.71% throughout New Zealand.

7.4.2 Cryptosporidiosis

The estimated coefficients for the weather variables in the projection calculation are shown in Table 7.2. The weather covariates, average weekly rainfall and temperature, were found to be significantly associated with the IR of campylobacteriosis in New Zealand from 1997 - 2007.

Cryptosporidiosis Projections for 2015

A2 Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2015 under the A2 climate scenario are mapped in Figure 7.10. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR in the North Island predominantly lies between 0.76% and 10.03%. The projected change in the west of the South Island lies between -19.07% and -0.27%, and between -0.26% and 10.03% in the east.
- In autumn, the projected change in cryptosporidiosis IR in New Zealand ranges between -1.35% and -0.27%, with a range of 0.76% to 10.03% in the Northland and Bay of Plenty DHBs.
- In winter, the projected change in IR in the Counties Manukau, Taranaki DHBs and the South Island ranges from -19.07% to -0.27%. The projected change in the Northland, Tairawhiti, Hawke's Bay, Mid Central and Wairarapa DHBs ranges from 0.76% to 10.03%, and from -0.26% and 0.75% in the Bay of Plenty, Lakes and Whanganui DHBs.

- In spring, the projected change in IR predominantly ranges between -0.26% and 0.75% throughout New Zealand, with the exception of the West Coast DHB (-19.07% to -1.36%), and the Waitemata DHB and bottom half of the North Island (0.76% to 10.03%).

A1B Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2015 under the A1B climate scenario are mapped in Figure 7.11. The projected percentage changes for each season are outlined below:

- In summer and spring, the projected change in IR predominantly ranges between 0.76% and 10.03% in the North Island, with a range of -0.26% to 0.75% in the Bay of Plenty, Lakes and Hawke's Bay DHBs. The projected change in IR in the west of the South Island ranges between -19.07% and -1.36%, and between -1.35% and 0.75% in the east of the South Island.

- In autumn, the change in IR predominantly ranges from -19.07% to -0.27% throughout New Zealand, with 0.76% to 10.03% projected in the Northland, Bay of Plenty and in parts of the Southland DHBs.

- In winter, the projected change in IR predominantly ranges from -0.26 to 10.03% in the North Island and from -19.07% to -0.27% in the South Island.

B1 Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2015 under the B1 climate scenario are mapped in Figure 7.12. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR in the North Island is projected to range between 0.76% and 10.03%, with a range from -0.26% to 0.75% in parts of the Waikato, Bay of Plenty, Lakes and Hawke's Bay DHBs. The projected change in the west of the South Island ranges from -19.07% to -1.36%, and predominantly ranges from -1.35% to 0.75% in the east of the South Island.

- In autumn, the projected change in IR ranges from -19.07% to -0.27% throughout New Zealand, with a projected change from 0.76% to 10.03% in parts of the Northland, Bay of Plenty and Southland DHBs.

- In winter, the suggested change in IR ranges from -19.07% to -1.36% in the Counties Manukau, Waikato, Nelson Marlborough and West Coast DHBs. A range of -1.35%

to 0.75% is suggested for the Bay of Plenty, Lakes, Whanganui, Taranaki, Canterbury, South Canterbury, Otago and Southland DHBs. In the Northland, Tairawhiti, Hawke's Bay, Mid Central, Wairarapa, Hutt and Capital and Coast DHBs a range from 0.76% to 10.03% is projected.

- In spring, the projected change in IR ranges from 0.76% to 10.03% in the bottom half of the North Island. A range in the projected IR change from -1.35% to 0.75% is suggested in the upper half of the North Island and east of the South Island. The change in the west of the South Island is suggested to range between -19.07% and -1.36%.

Cryptosporidiosis Projections for 2040

A2 Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2040 under the A2 climate scenario are mapped in Figure 7.13. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR throughout most of the North Island and parts of the Canterbury and South Canterbury DHBs were projected to range from 2.23% - 32.24%. A range from -46.14% to -3.87% is projected in the west of the South Island and from -3.86% to 2.22% elsewhere.

- In autumn, the projected change in IR is suggested to range from -3.86% to 2.22% throughout New Zealand, with a range of 2.23% to 32.24% in the Northland, Bay of Plenty and parts of Southland DHBs.

- In winter, the projected change in IR ranges from -3.86% to 2.22% throughout New Zealand, with 2.23% to 32.24% projected in the Northland DHB and east coast of the North Island, and from -46.14% to -3.87% on the west of the South Island.

- In spring, the projected change in IR throughout New Zealand predominantly ranges from -0.70% to 2.22%, with 2.23% to 32.24% projected in the bottom half of the North Island, and -46.14% to -3.87% projected in the West Coast DHB.

A1B and B1 Climate Scenarios

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2040 under the A1B and B1 climate scenarios are mapped in Figures 7.14 and 7.15, respectively. The spatial patterns in the projected percentage change of cryptosporidiosis IR for each season were found to be the same in both scenarios, and are outlined below:

- In summer, the projected change in IR in the North Island are suggested to range between 2.23% and 32.24%. The projected change in the Bay of Plenty and Lakes DHBs and the east of the South Island ranges from -3.86% to 2.22%, with the range in the west of the South Island between -46.14% and -3.87%.
- In autumn, the projected change in IR predominantly lies between -3.86% and -0.71% throughout New Zealand, with the projected change in the Northland, Bay of Plenty and parts of Southland DHBs ranging from 2.23% to 32.24%.
- In winter, the projected change throughout New Zealand predominantly ranges between -3.86% and 2.22%, with a range from 2.23% to 32.24% projected in the Northland DHB and east coast of the North Island, and from -46.14% to -3.87% in the Waikato DHB and west coast of South Island.
- In spring, the suggested change in IR ranges between -3.86% and 2.22% throughout New Zealand, with the bottom half of the North Island expecting changes from 2.23% to 32.24%, and between -46.14% and -3.87% on the west coast of South Island.

Cryptosporidiosis Projections for 2090

A2 Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2090 under the A2 climate scenario are mapped in Figure 7.16. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR ranges from 5.68% to 39.12% in the North Island and in parts of the east of the South Island, with projections ranging between -50.71% and -1.90% on the west of the South Island.
- In autumn, the projected change in IR throughout New Zealand ranges from -1.89% to 5.67%, except in the Northland, Bay of Plenty and parts of Southland DHBs (5.68% to 39.12%) and on the east coast of the North Island (-50.71% to -1.90%).
- In winter, the projected change in the IR of cryptosporidiosis ranges from 5.68% to 39.12% in the North Island and parts of the Canterbury DHB, and from -50.71% to -1.90% in parts of the Waikato and Taranaki DHBs, and along the east coast of the South Island.

- In spring, the change in IR is projected to range between 5.68% and 39.12% in the North Island. On the east of the South Island, the projected change in IR ranges from -1.89% to 5.67%, and from -50.71% to -1.90% on the west.

A1B Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2090 under the A1B climate scenario are mapped in Figure 7.17. The projected percentage changes for each season are outlined below:

- In summer, the projected change in cryptosporidiosis IR in the North Island ranges from 5.68% to 39.12%. Throughout the majority of the South Island the projected change ranges from -1.89% to 5.85%, with -50.71% to -1.90% in parts of the West Coast and Southland DHBs.

- In autumn, the projected change in IR throughout New Zealand predominantly ranges from -1.89% to 2.13%, with 5.68% to 39.12% in the Northland, Bay of Plenty and parts of Southland DHBs, and -50.71% to -1.90% in parts of the east coast of North Island.

- In winter, the projected change in IR throughout New Zealand ranges between -1.89% to 5.67%, with a range from 5.68% - 39.12% in the east coast of the North Island and Northland DHB and from -50.71% to -1.90% in parts of the east coast of the South Island and the Waikato and Taranaki DHBs.

- In spring, the projected change in IR in the North Island ranges between 2.14% and 39.12%, between -1.89% and 5.67% in the east of the South Island, and between -50.71% and -1.90% in the west of the South Island.

B1 Climate Scenario

The seasonal projected percentage change in cryptosporidiosis IR throughout New Zealand in 2090 under the B1 climate scenario are mapped in Figure 7.18. The projected percentage changes for each season are outlined below:

- In summer, the projected change in New Zealand ranges between -1.89% and 5.67%, with a change of 5.68% to 39.12% projected in parts of the bottom half of North Island, and between -50.71% and -1.90% on west coast of the South Island.

- In autumn, the change in IR is projected between -1.89% and 2.13% throughout New Zealand, with 5.68% to 39.12% in parts of the Northland, Waitemata, Auckland, Bay

Weather Covariate (W_k)	$\hat{\beta}_k$	s.e.(W_k) (1997 - 2007)	$\tilde{\beta}_k$	p-value
Average Absolute Humidity	0.0462	0.0017	27.4708	0.918
Average Temperature	-0.1274	4.5039	-0.0283	0.541

Table 7.3: The influenza hospitalisation projection calculation estimate for each weather variable (W_k), the posterior coefficient estimates ($\hat{\beta}_k$), the standard error of the weather variables throughout the study period (s.e.(W_k)), the coefficient estimate for the projection calculation ($\tilde{\beta}_k$), and the level of significance (p-value).

of Plenty and Southland DHBs, and -50.71% to -1.90% in parts of the east coast of the North Island and parts of the West Coast, Canterbury, South Canterbury and Otago DHBs.

- In winter, the projected change in cryptosporidiosis IR predominantly ranges from -1.89% to 2.13% throughout New Zealand, with -50.71% to -1.90% projected in the Waikato, Taranaki, Nelson-Marlborough, West Coast and Southland DHBs, and 5.68% to 39.12% in the Northland, Waitemata, Tairawhiti, Hawke's Bay DHBs.
- In spring, the change in IR throughout New Zealand predominantly ranges from -1.89% to 2.13%, with the projected change on the west coast of the South Island ranging from -50.71% to -1.90%, and from 5.68% to 39.12% in parts of the Tairawhiti and Taranaki DHBs.

7.4.3 Influenza Hospitalisations

The estimated coefficients for the weather variables in the projection calculation are shown in Table 7.3. It must be noted that the weather variables were found to be insignificantly associated with the IR of influenza hospitalisations in New Zealand from 1997 - 2007, during the past association modelling in Chapter 6. Therefore, it is statistically plausible that there may be in fact no change in the IR of influenza hospitalisations in 2015, 2040 and 2090 in comparison to that seen in the study period (1997 - 2007, inclusive). However, our projections still represent our "best guess" based on the data and the fitted model.

Influenza Hospitalisation Projections for 2015

A2 and A1B Climate Scenarios

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2015 under the A2 and A1B climate scenarios are mapped in Figures 7.19 and 7.20, respectively. The spatial patterns in the projected percentage change of influenza hospitalisations IR for each season were found to be the same in both scenarios, and are outlined below:

- In summer, the projected change in IR of influenza hospitalisations throughout New Zealand lies predominantly between -1.12% and -0.31%.
- In autumn, the projected change in IR ranges from -1.12% to -0.31% throughout New Zealand, with a range from -0.18% to 0.55% in parts of the Waitemata, Auckland, Waikato, Nelson-Marlborough, Canterbury and Southland DHBs.
- In winter and spring, the projected change in IR ranges from -1.12% to -0.31% throughout New Zealand, with -0.18% to 0.55% in parts of Nelson-Marlborough, Canterbury and Southland DHBs.

B1 Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2015 under the B1 climate scenario are mapped in Figure 7.21. The projected percentage changes for each season are outlined below:

- In summer, the change in IR throughout New Zealand is projected to predominately range between -0.41% and -0.19%.
- In autumn, the projected change in IR ranges predominantly from -0.30% to 0.55% throughout New Zealand.
- In winter and spring, the projected change in influenza hospitalisation IR ranges between -1.12% and -0.42% in the West Coast, Canterbury and Otago DHBs, between -0.41% and -0.19% in the North Island, and between -0.18% and 0.55% in parts of the Nelson-Marlborough, Canterbury and Southland DHBs.

Influenza Hospitalisation Projections for 2040

A2 Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2040 under the A2 climate scenario are mapped in Figure 7.22. The projected percentage changes for each season are outlined below:

- In summer, the projected change in influenza hospitalisation IR predominantly lies between -3.26% and -0.90% throughout New Zealand.

- In autumn, the projected change in IR ranges between -0.89% and 1.16% throughout New Zealand.

- In winter and spring, the projected change in IR predominantly lies between -3.26% and -0.90% throughout New Zealand, with parts of Nelson-Marlborough, Canterbury and Southland DHBs projected to range from -0.55% to 1.16%.

A1B Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2040 under the A1B climate scenario are mapped in Figure 7.23. The projected percentage changes for each season are outlined below:

- In summer, the projected change in influenza hospitalisation IR throughout New Zealand lies predominantly within the range of -3.26% to -0.90%.

- In autumn, the projected change throughout New Zealand ranges between -0.89% and 1.16%, with a range of -3.26% to -1.21% in the South Canterbury and Otago DHBs.

- In winter, the projected change throughout New Zealand ranges from -3.26% to -0.56%, with the range in parts of the Nelson-Marlborough, Canterbury and Southland DHBs between -0.55% and 1.16%.

- In spring, the projected change in IR throughout New Zealand ranges from -3.26% to -0.90%, with the range in parts of the Nelson-Marlborough, Canterbury and Southland DHBs from -0.55% to 1.16%.

B1 Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2040 under the B1 climate scenario are mapped in Figure 7.24. The projected percentage changes for each season are outlined below:

- In summer, the projected change in influenza hospitalisations IR predominately range between -1.20% and -0.56% throughout New Zealand.

- In autumn, the suggested change in IR ranges from -0.89% to 1.16% throughout New Zealand.

- In winter and spring, the change in IR is projected to range between -1.20% and -0.56% throughout New Zealand, with a change from -0.55% to 1.16% projected in parts of the Nelson-Marlborough, Canterbury and Southland DHBs.

Influenza Hospitalisation Projections for 2090

A2 Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2090 under the A2 climate scenario are mapped in Figure 7.25. The projected percentage changes for each season are outlined below:

- In summer and autumn, the projected change in influenza hospitalisations IR throughout New Zealand ranges between -6.46% and -2.68%.
- In winter, the projected change in IR ranges from -6.46 to -3.39% throughout New Zealand.
- In spring, the change in IR predominantly ranges between -6.46% and -3.39% throughout New Zealand, with the change in the west coast of the South Island projected to range between -3.38% and -1.87%.

A1B Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2090 under the A1B climate scenario are mapped in Figure 7.26. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR ranges predominantly between -6.46% and -2.68% throughout New Zealand, with the change in the Northland, Waitemata, Auckland, Counties Manakau, Waikato and Hawke's Bay DHBs ranging between -2.67% and -1.87%.
- In autumn, the projected change in IR ranges between -2.67% and -1.87% throughout New Zealand.
- In winter and spring, the change in IR ranges predominantly between -6.46% and -2.68% throughout New Zealand.

B1 Climate Scenario

The seasonal projected percentage change in influenza hospitalisations IR throughout New Zealand in 2090 under the B1 climate scenario are mapped in Figure 7.27. The projected percentage changes for each season are outlined below:

- In summer, winter and spring, the projected change in influenza hospitalisation IR ranges between -2.67% and 0.82% throughout New Zealand.

Weather Covariate (W_k)	$\hat{\beta}_k$	s.e.(W_k) (1997 - 2007)	$\tilde{\beta}_k$	p-value
Average Absolute Humidity	0.0483	0.0017	28.7192	0.257
Average Temperature	-0.1173	4.5039	-0.0260	0.166

Table 7.4: The meningococcal disease projection calculation estimate for each weather variable (W_k), the posterior coefficient estimates ($\hat{\beta}_k$), the standard error of the weather variables throughout the study period (s.e.(W_k)), the coefficient estimate for the projection calculation ($\tilde{\beta}_k$), and the level of significance (p-value).

- In autumn: the projected change in influenza hospitalisation IR throughout New Zealand ranges between -1.86% and 0.82%.

7.4.4 Meningococcal Disease

The estimated coefficients for the weather variables in the projection calculation are shown in Table 7.4. However, it must be noted that the weather variables were found to be insignificantly associated with the IR of meningococcal disease in New Zealand from 1997 - 2007, during the past association modelling in Chapter 6. Therefore, it is statistically plausible that there may be in fact no change in the IR of meningococcal disease in 2015, 2040 and 2090 in comparison to that seen in the study period (1997 - 2007, inclusive). However, our projections still represent our "best guess" based on the data and the fitted model.

Meningococcal Disease Projections for 2015

A2 and A1B Climate Scenarios

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2015 under the A2 and A1B climate scenarios are mapped in Figures 7.28 and 7.29, respectively. The spatial patterns in the projected percentage change of meningococcal disease IR for each season were found to be the same in both scenarios, and are outlined below:

- In summer, the map suggests the projected change in meningococcal disease IR will range between -0.97% and -0.16% throughout New Zealand.

- In autumn, the range of projected change in IR meningococcal disease throughout New Zealand lies between -0.24% to 0.48%, with a change of -0.97% to -0.35% projected in parts of the South Canterbury DHB.

- In winter and spring, the projected change in IR ranges from -0.97% to -0.25% throughout New Zealand, with a range of -0.15% to 0.48% in parts of the Nelson-Marlborough, Canterbury and Southland DHBs.

B1 Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2015 under the B1 climate scenario are mapped in Figure 7.30. The projected percentage changes for each season are outlined below:

- In summer, the range of projections for the change in IR in New Zealand predominantly lie between -0.34% and -0.16%, with a range from -0.15% to 0.48% in parts of the Northland, Waitemata, Auckland, Counties Manakau, Canterbury, and Southland DHBs.
- In autumn, the projected change in IR throughout in New Zealand ranges from -0.24% to 0.48%.
- In winter and spring, the projected change in IR throughout the North Island ranges from -0.97% to -0.16%, except in the Nelson-Marlborough, Canterbury and Southland DHBs (-0.24% to 0.44%) and parts the South Canterbury and Otago DHBs (-0.97% to -0.35%).

Meningococcal Disease Projections for 2040

A2 Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2040 under the A2 climate scenario are mapped in Figure 7.31. The projected percentage changes for each season are outlined below:

- In summer, the projected change in meningococcal disease IR throughout New Zealand ranges from -1.25% to -0.49%.
- In autumn, the change in IR throughout New Zealand is projected to ranged between -0.81% and 2.13%.
- In winter and spring, the projected change in meningococcal disease IR in New Zealand ranges from -1.25% to -0.49%, with a range from -0.48% to 2.13% projected in parts of the Nelson Marlborough, Canterbury and Southland DHBs.

A1B Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New

Zealand in 2040 under the A1B climate scenario are mapped in Figure 7.32. The projected percentage changes for each season are outlined below:

- In summer, the projected change in IR ranges from -4.04% to -1.26% throughout New Zealand.
- In autumn and spring, the projected change in IR predominantly ranges from -4.04% to -1.26% throughout New Zealand, except in parts of Nelson-Marlborough, Canterbury and Soutland DHBs (-0.48% to 2.13%).
- In winter, the projected change in IR throughout New Zealand ranges between -1.25% and -0.49%.

B1 Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2040 under the B1 climate scenario are mapped in Figure 7.33. The projected percentage changes for each season are outlined below:

- In summer, the projected change in meningococcal disease IR ranges from -1.25% to 2.13% throughout New Zealand.
- In autumn, the projected change in IR throughout New Zealand predominantly ranges from -0.48% to 2.13%.
- In winter and spring, the projected change in IR ranges from -1.25% to 2.13% throughout New Zealand, with a change ranging between -4.04% and -1.25% in parts of the South Canterbury, Otago and Southland DHBs.

Meningococcal Disease Projections for 2090

A2 Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2090 under the A2 climate scenario are mapped in Figure 7.34. The projected percentage changes for each season are outlined below:

- In summer and autumn, the projected change in meningococcal disease IR throughout New Zealand ranges from -3.64% to -1.62%.
- In winter, the projected change in IR throughout New Zealand ranges from -5.96% to -2.81%.

- In spring, the change in IR throughout New Zealand is projected to range between -3.64% and -2.81%.

A1B Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2090 under the A1B climate scenario are mapped in Figure 7.35. The projected percentage changes for each season are outlined below:

- In summer and winter, the projected change in IR throughout New Zealand ranges from -5.96% to -3.65%.

- In autumn, the projected change in meningococcal disease IR ranges between -5.96% and -2.81% throughout New Zealand.

- In spring, the projected change in IR ranges from -5.96% to -3.65% in the North Island and from -5.96% to -2.81% in the South Island.

B1 Climate Scenario

The seasonal projected percentage change in meningococcal disease IR throughout New Zealand in 2090 under the B1 climate scenario are mapped in Figure 7.36. The projected percentage changes for each season are outlined below:

- In summer, winter and spring, the projected change in meningococcal disease IR thought New Zealand ranges between -2.80% and 0.69%.

- In autumn, the projected change in IR ranges between -1.61% and 0.69% throughout New Zealand.

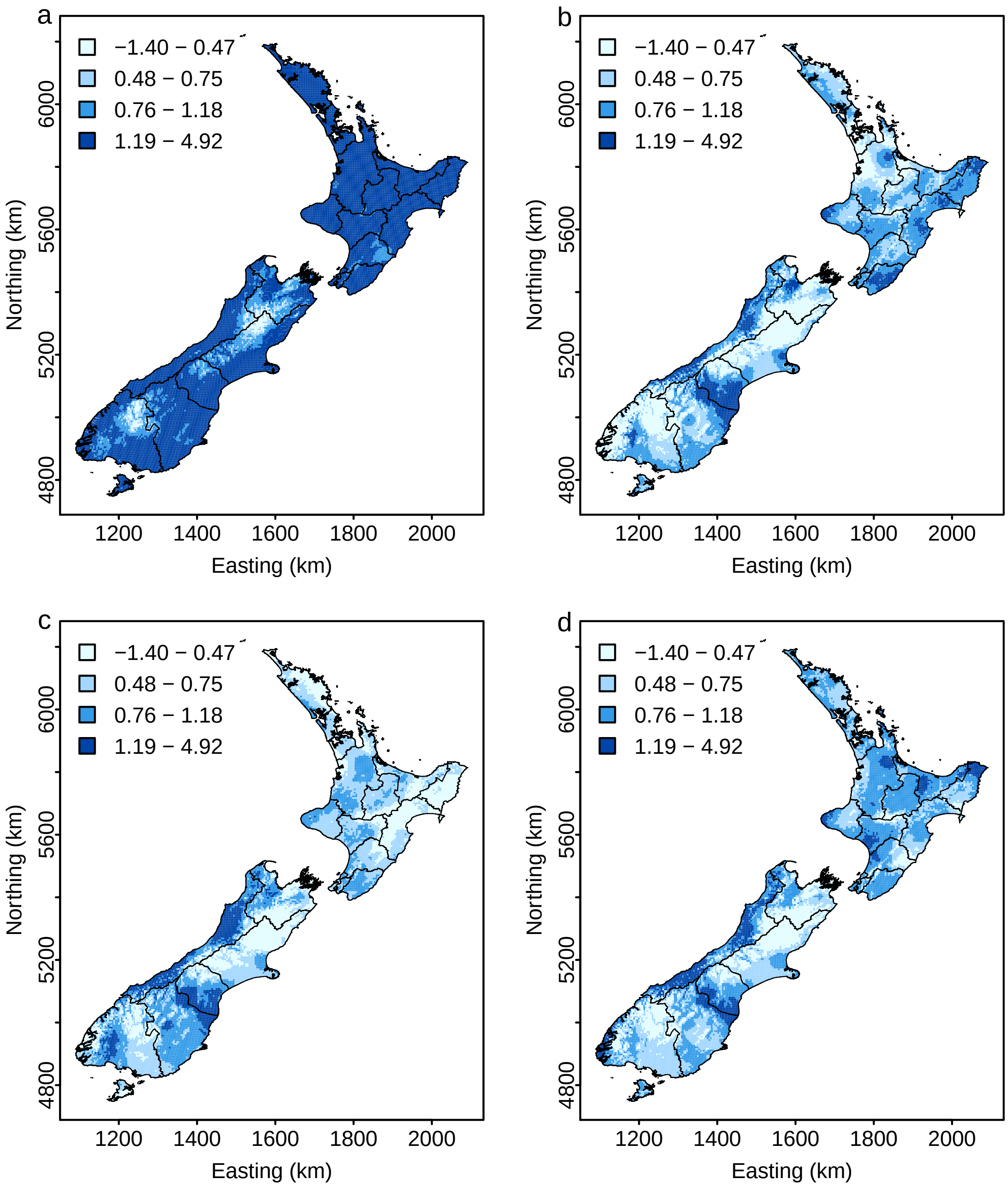


Figure 7.1: The projected change in campylobacteriosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

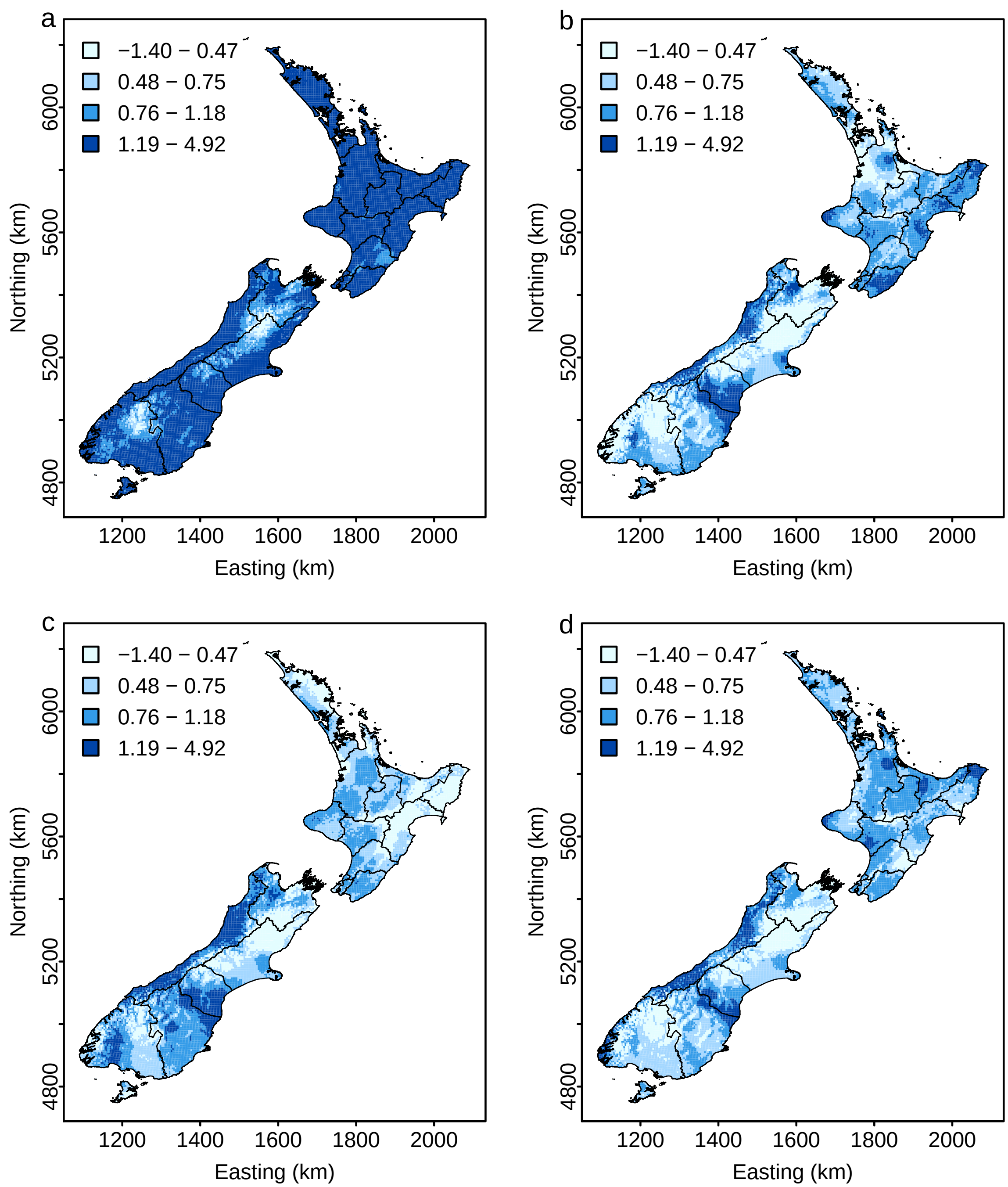


Figure 7.2: The projected change in campylobacteriosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

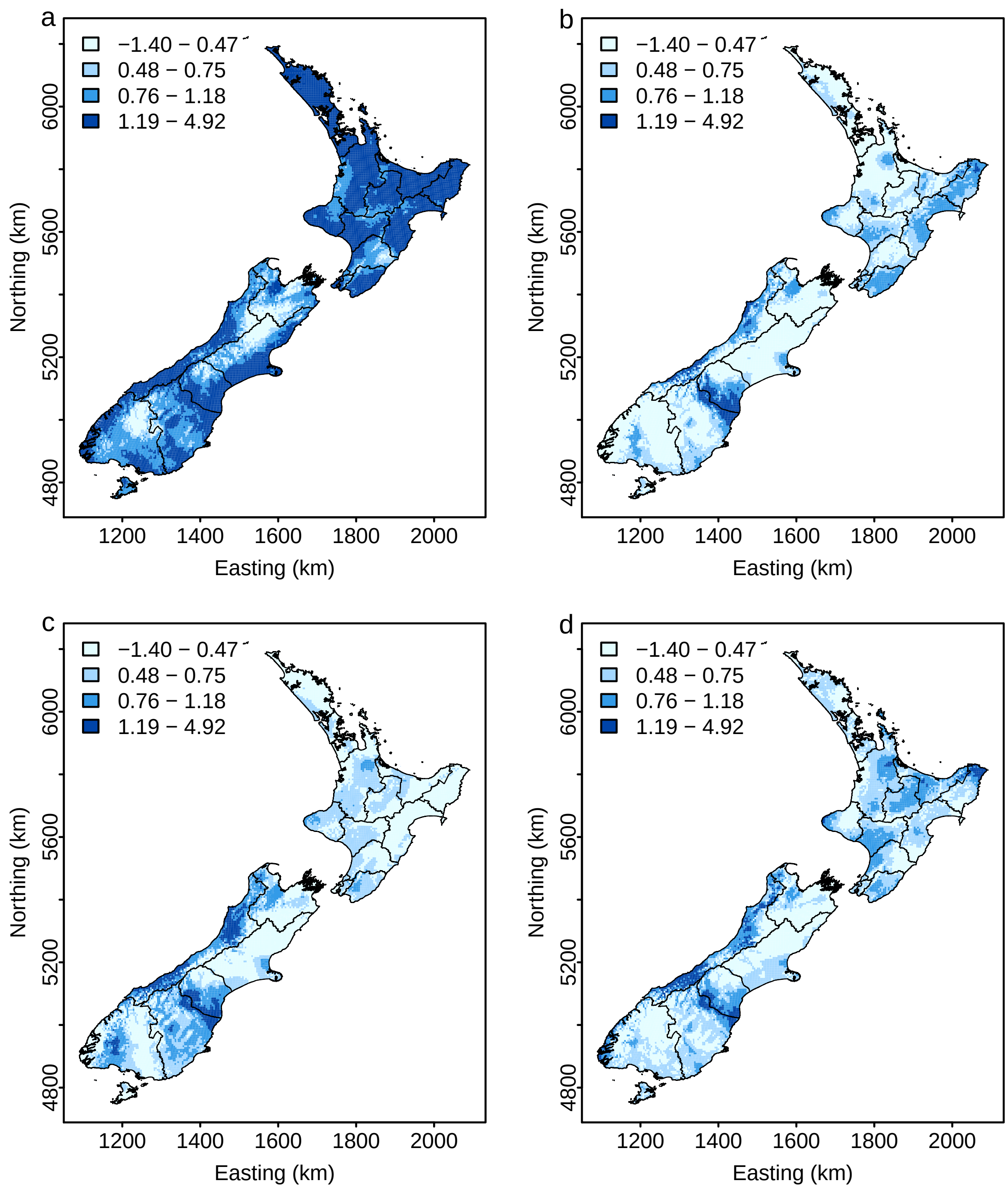


Figure 7.3: The projected change in campylobacteriosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

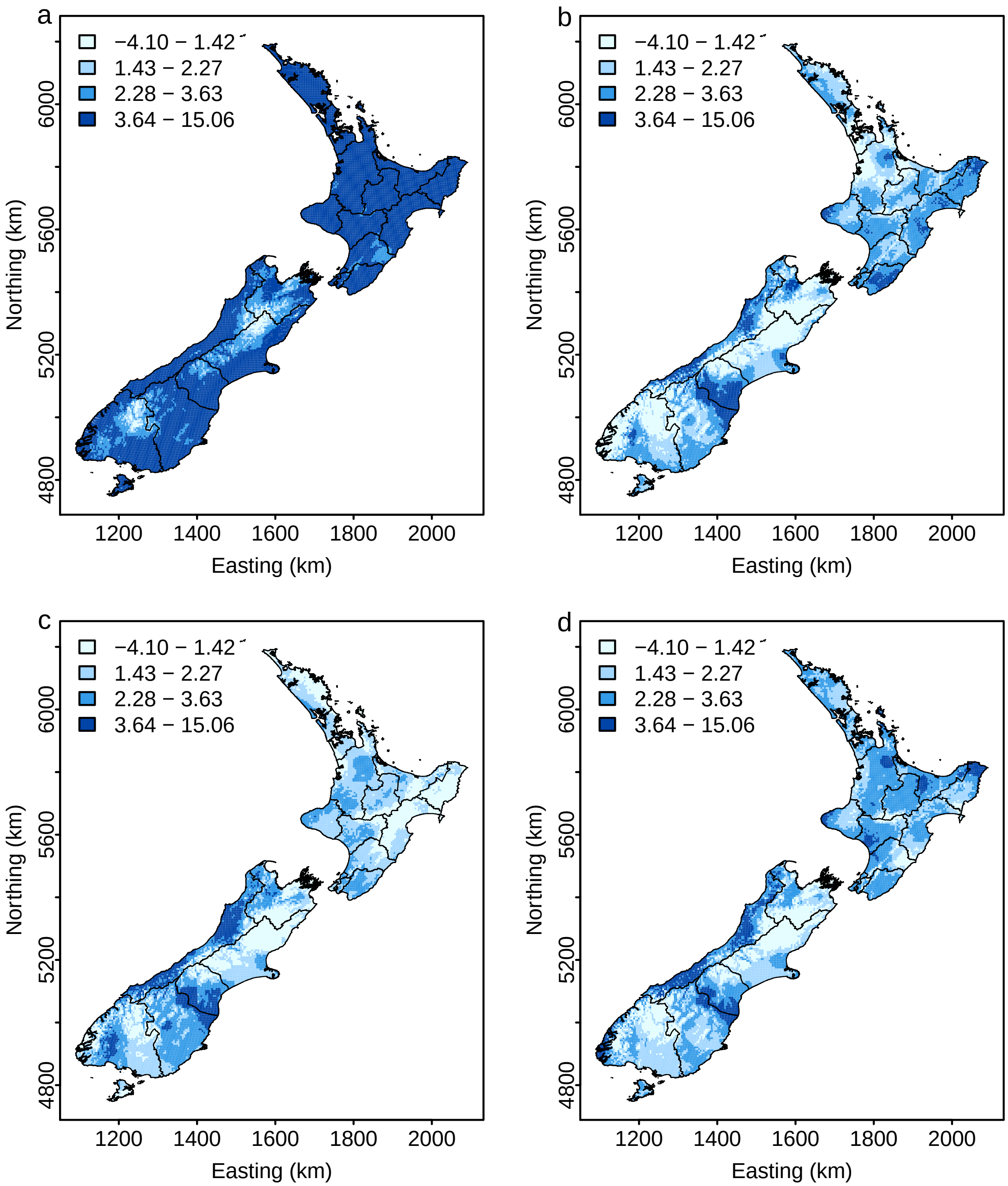


Figure 7.4: The projected change in campylobacteriosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

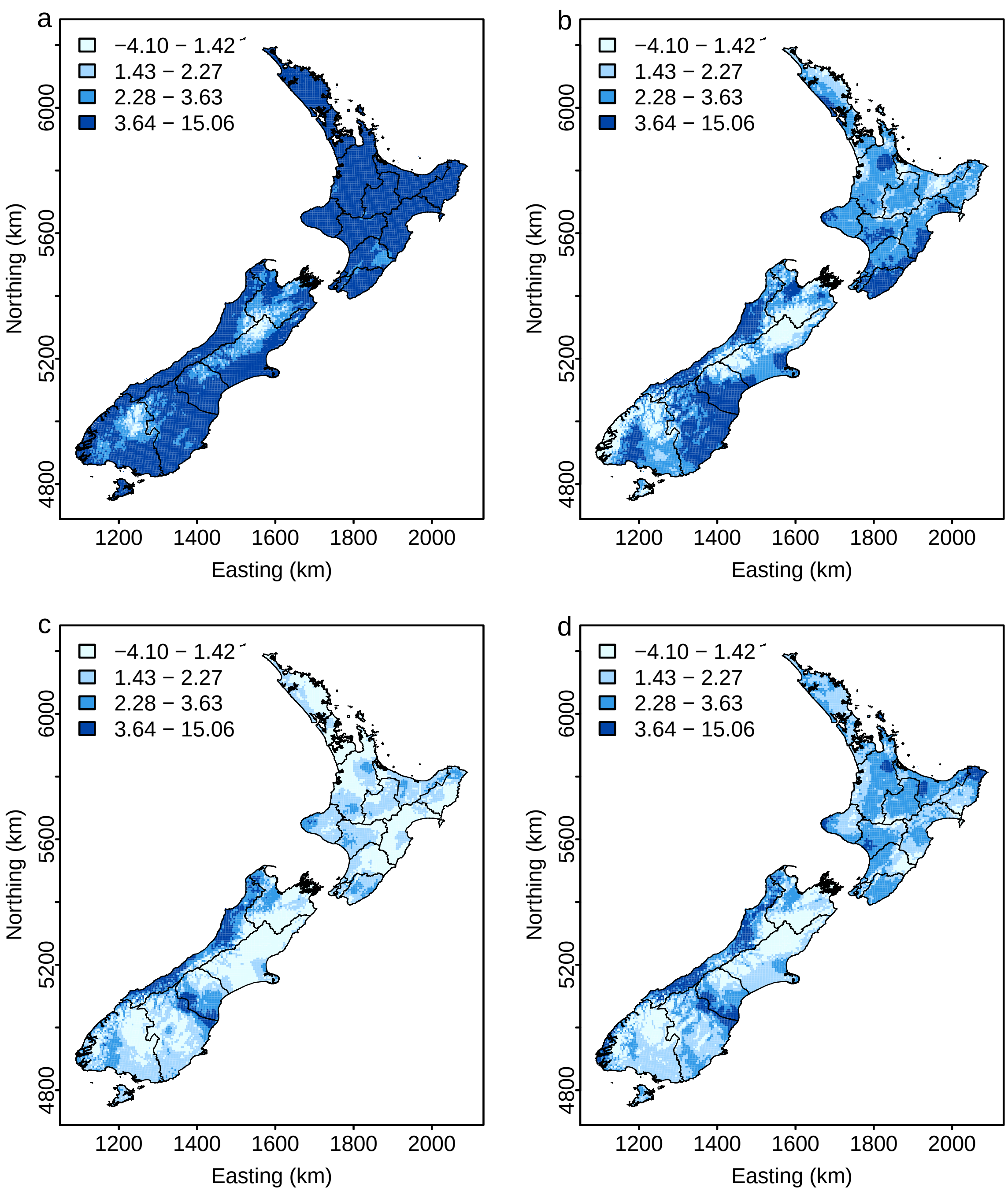


Figure 7.5: The projected change in campylobacteriosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

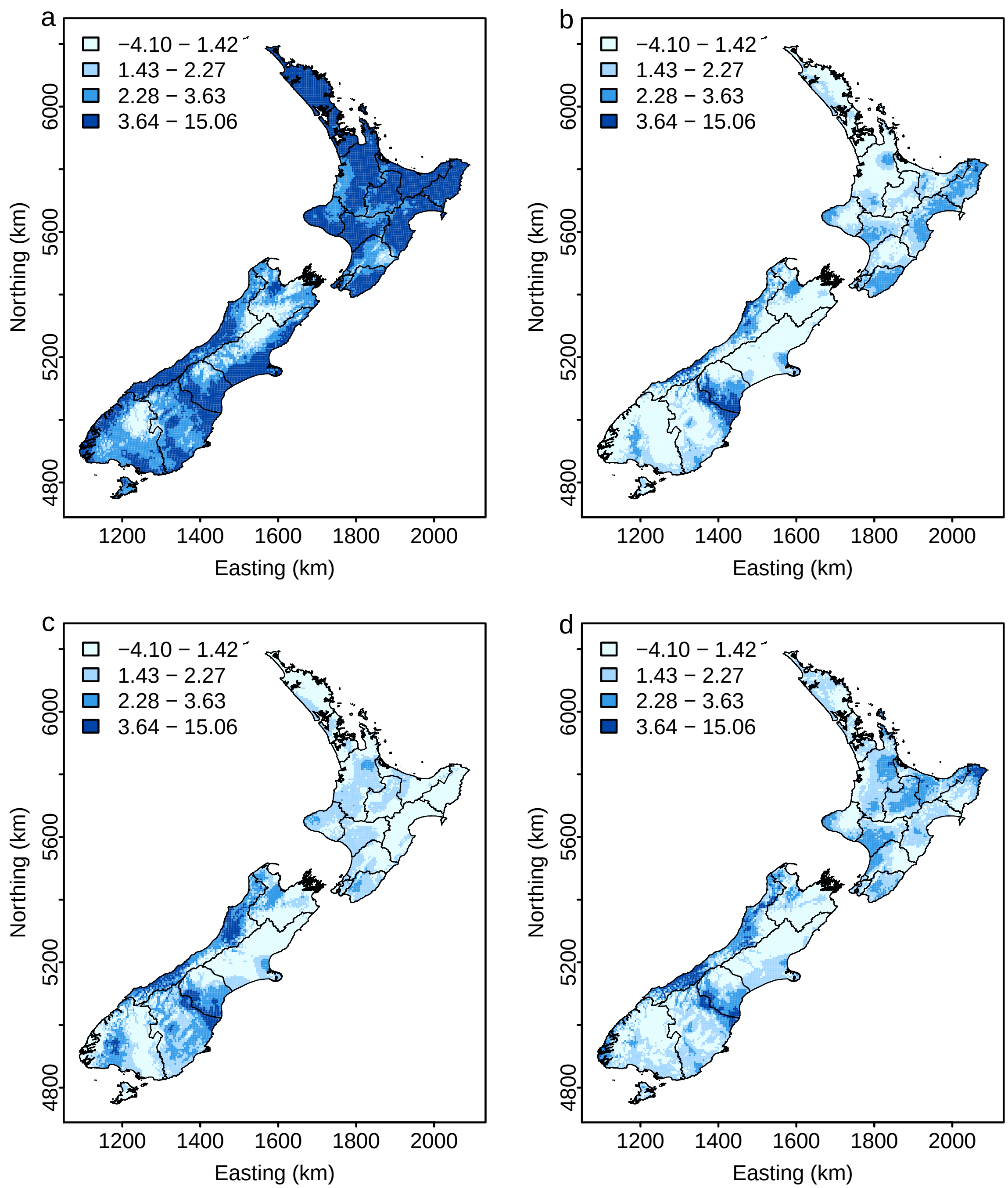


Figure 7.6: The projected change in campylobacteriosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

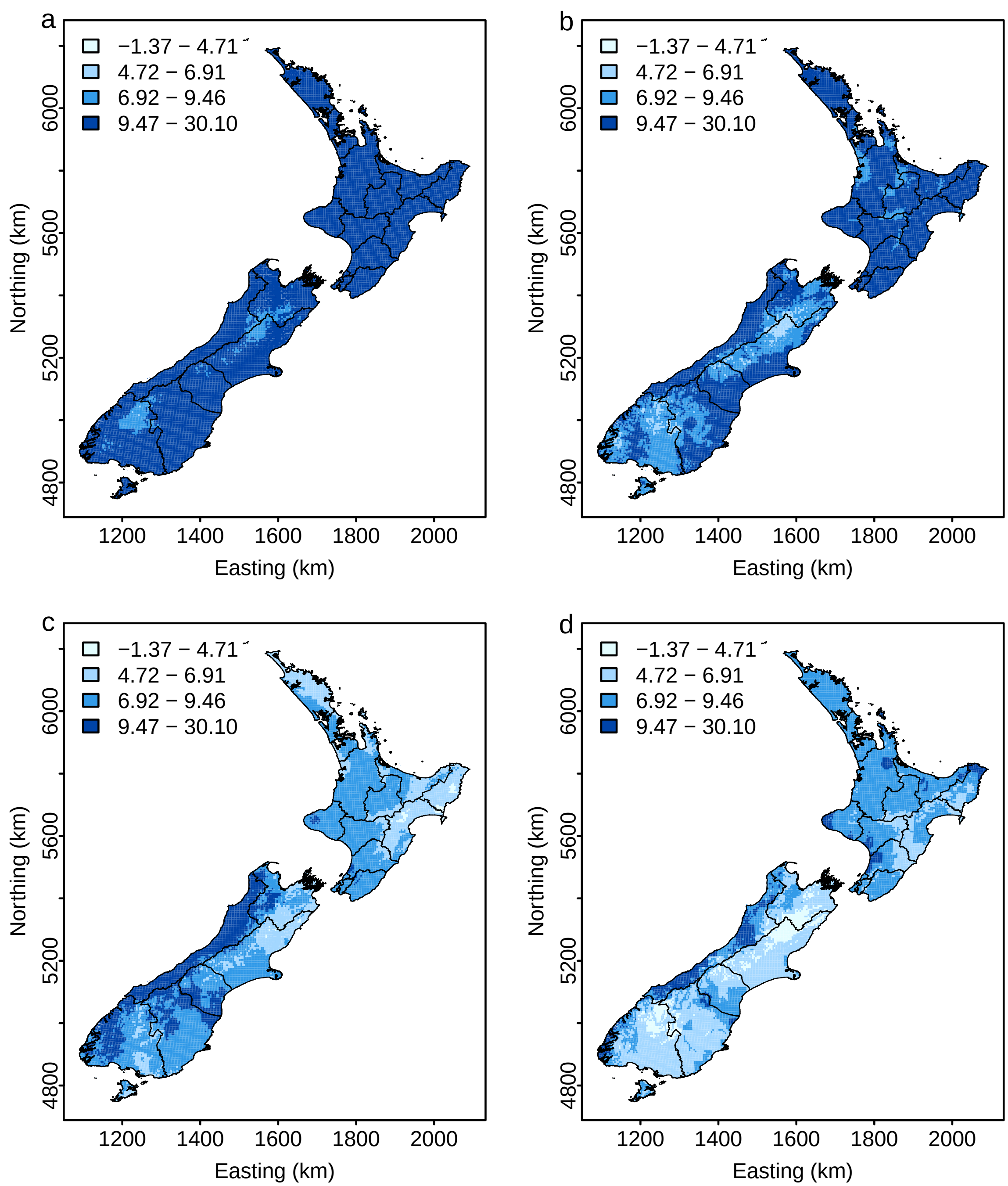


Figure 7.7: The projected change in campylobacteriosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

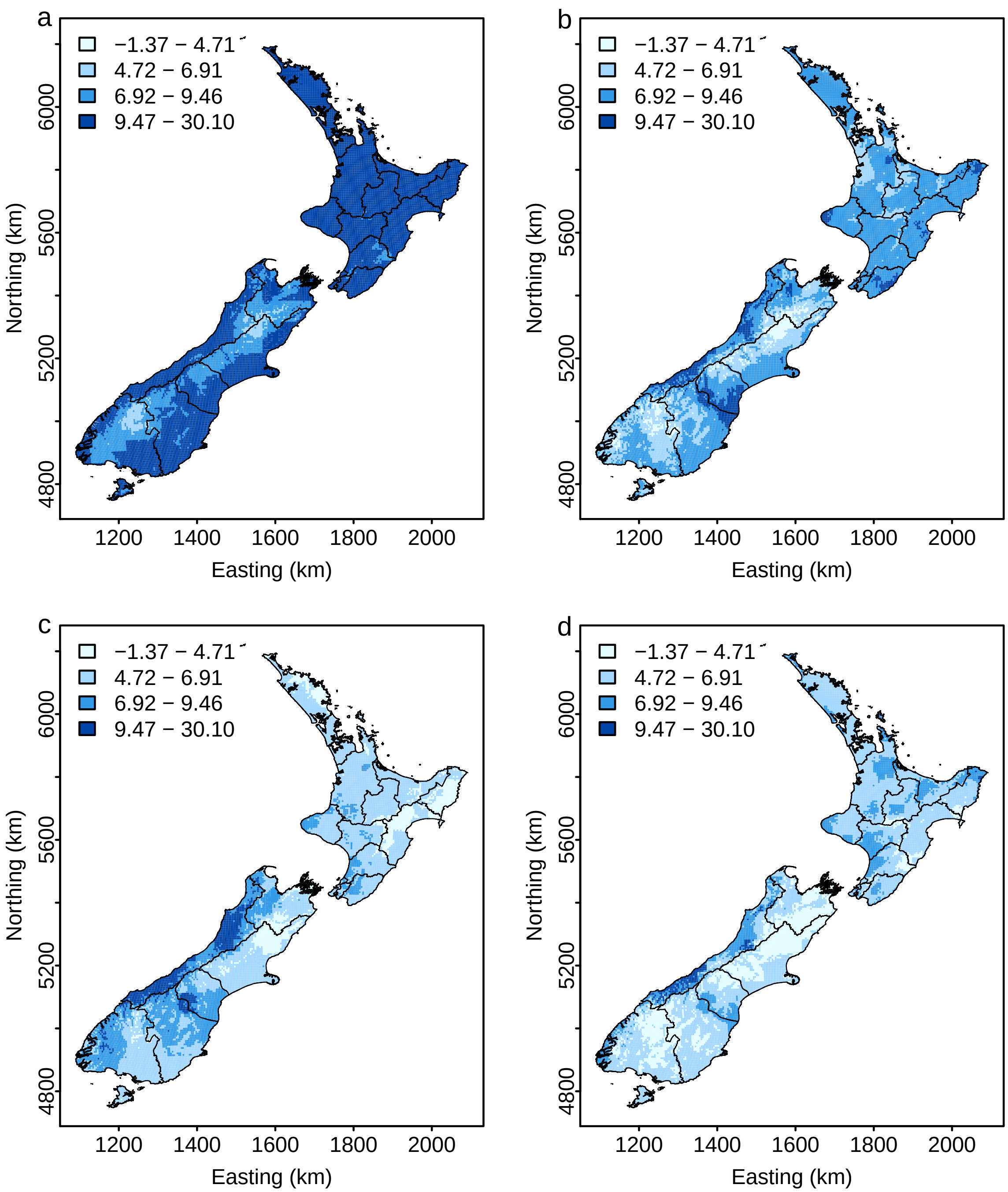


Figure 7.8: The projected change in campylobacteriosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

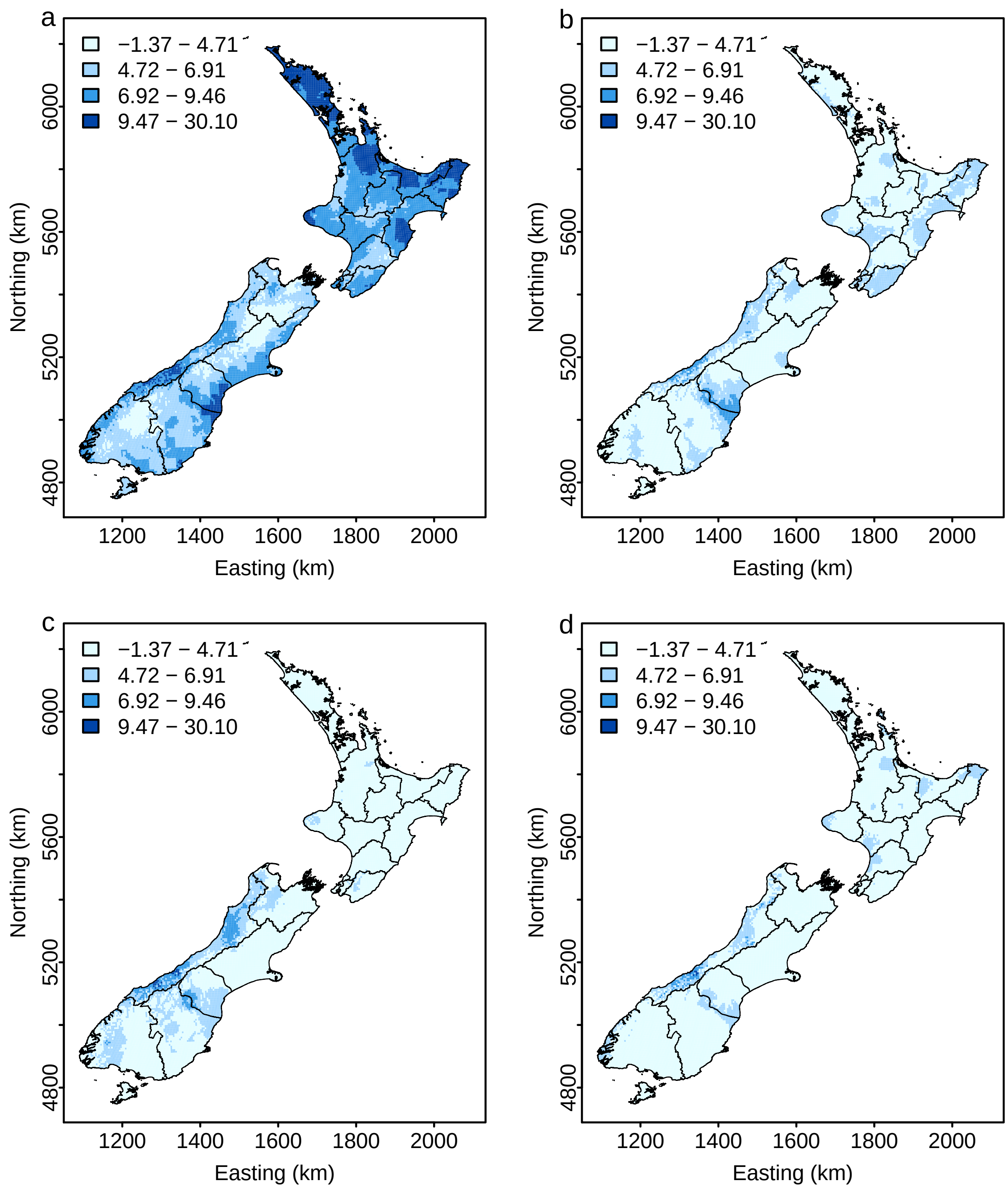


Figure 7.9: The projected change in campylobacteriosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

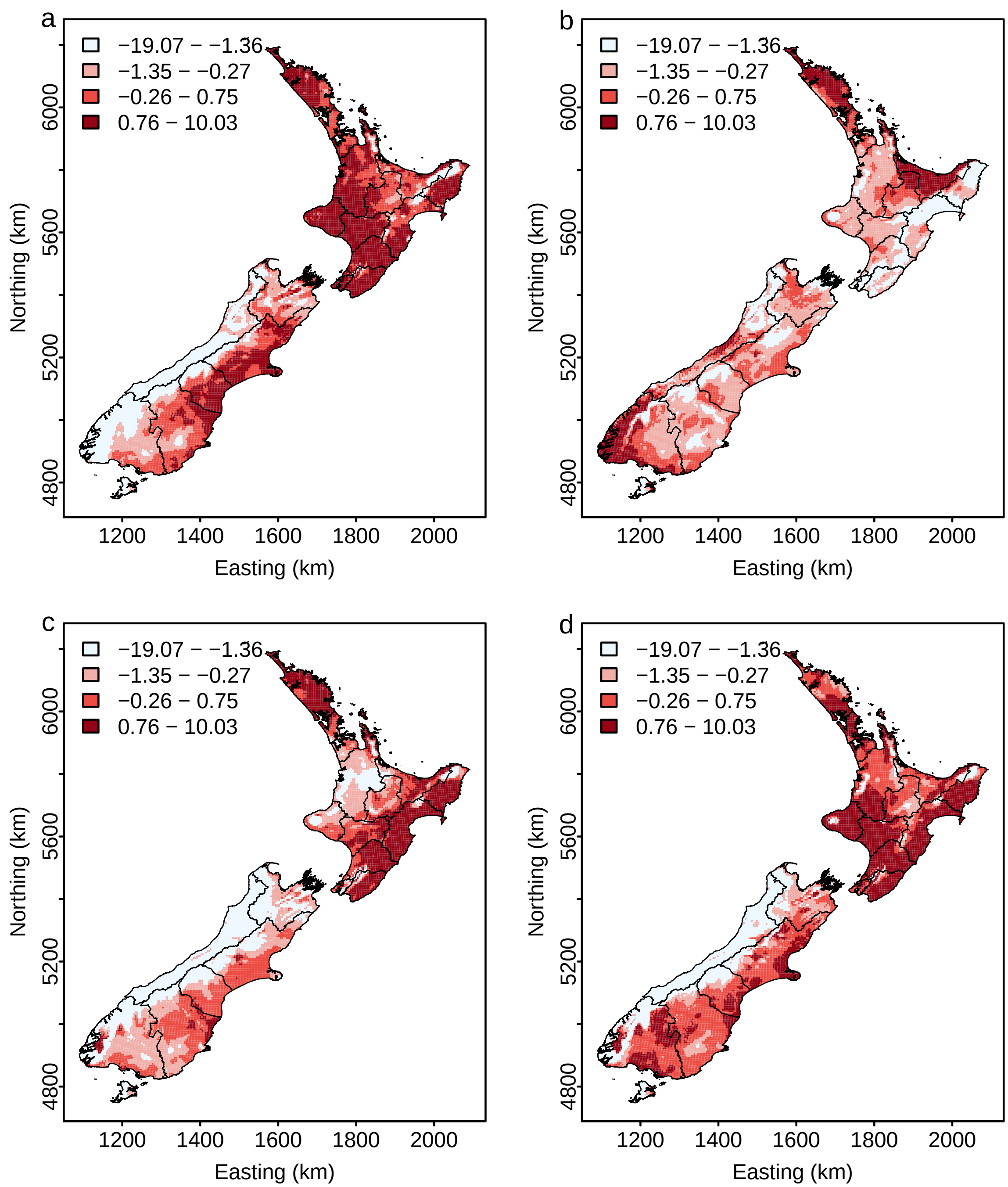


Figure 7.10: The projected change in cryptosporidiosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

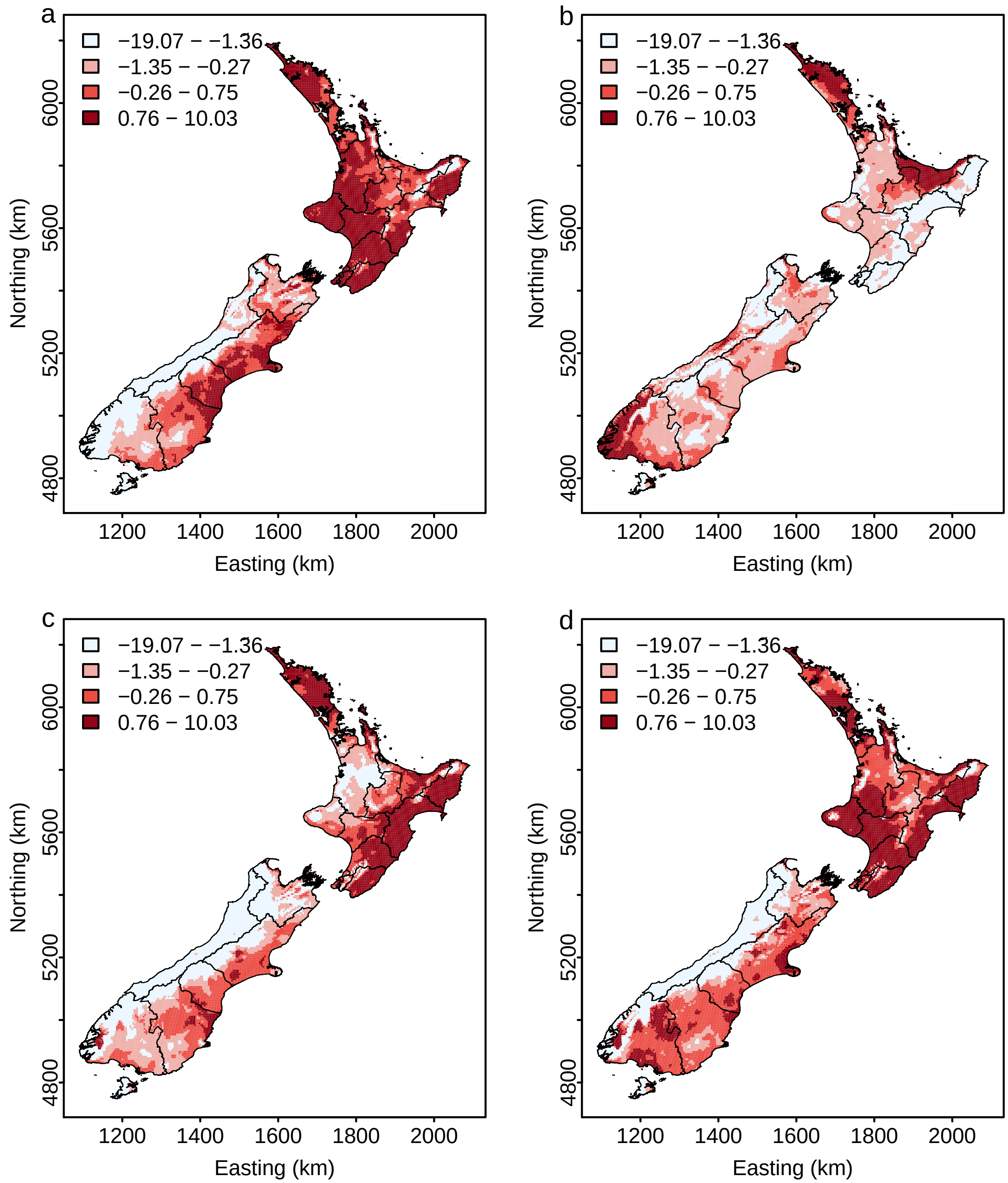


Figure 7.11: The projected change in cryptosporidiosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

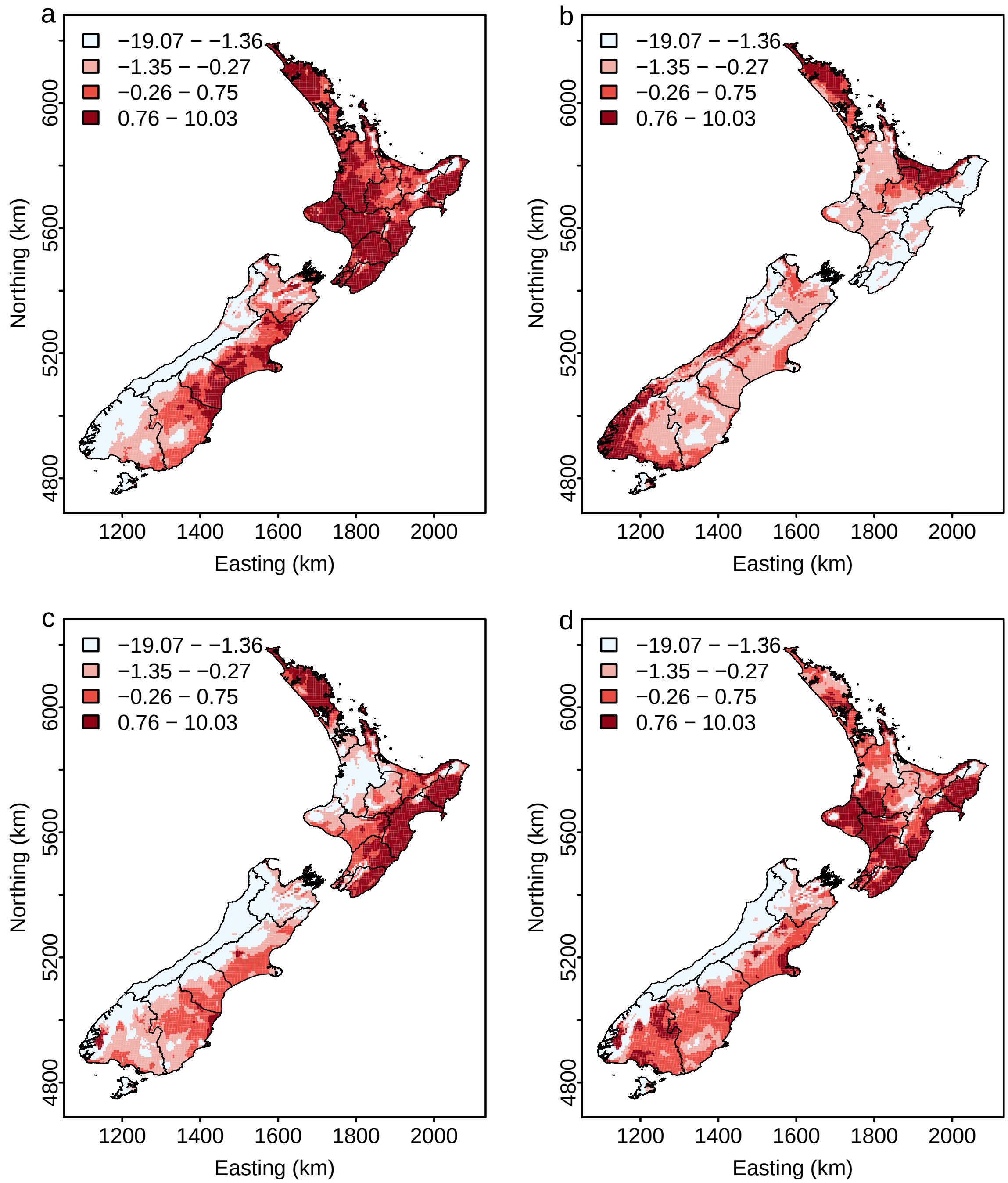


Figure 7.12: The projected change in cryptosporidiosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

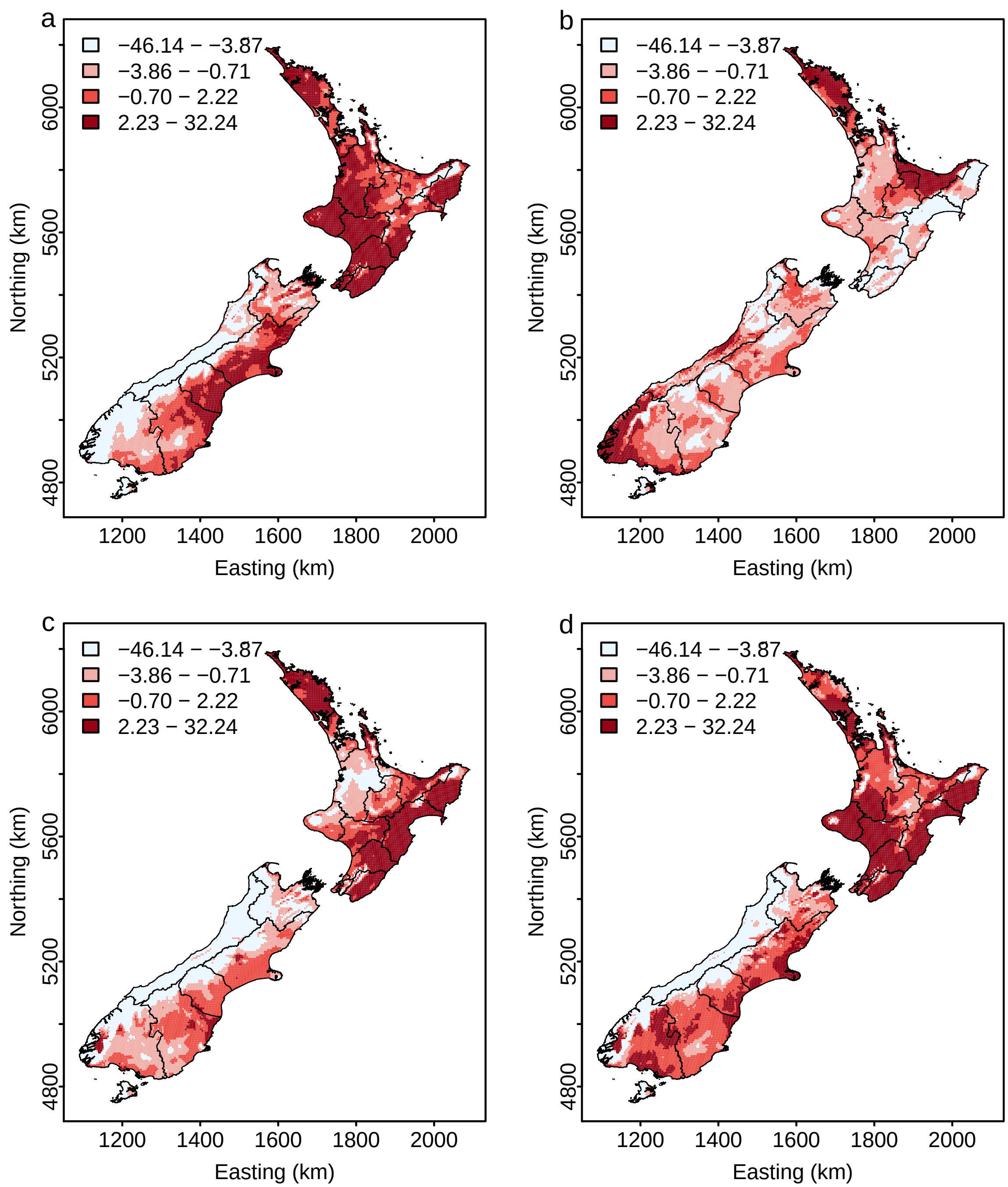


Figure 7.13: The projected change in cryptosporidiosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

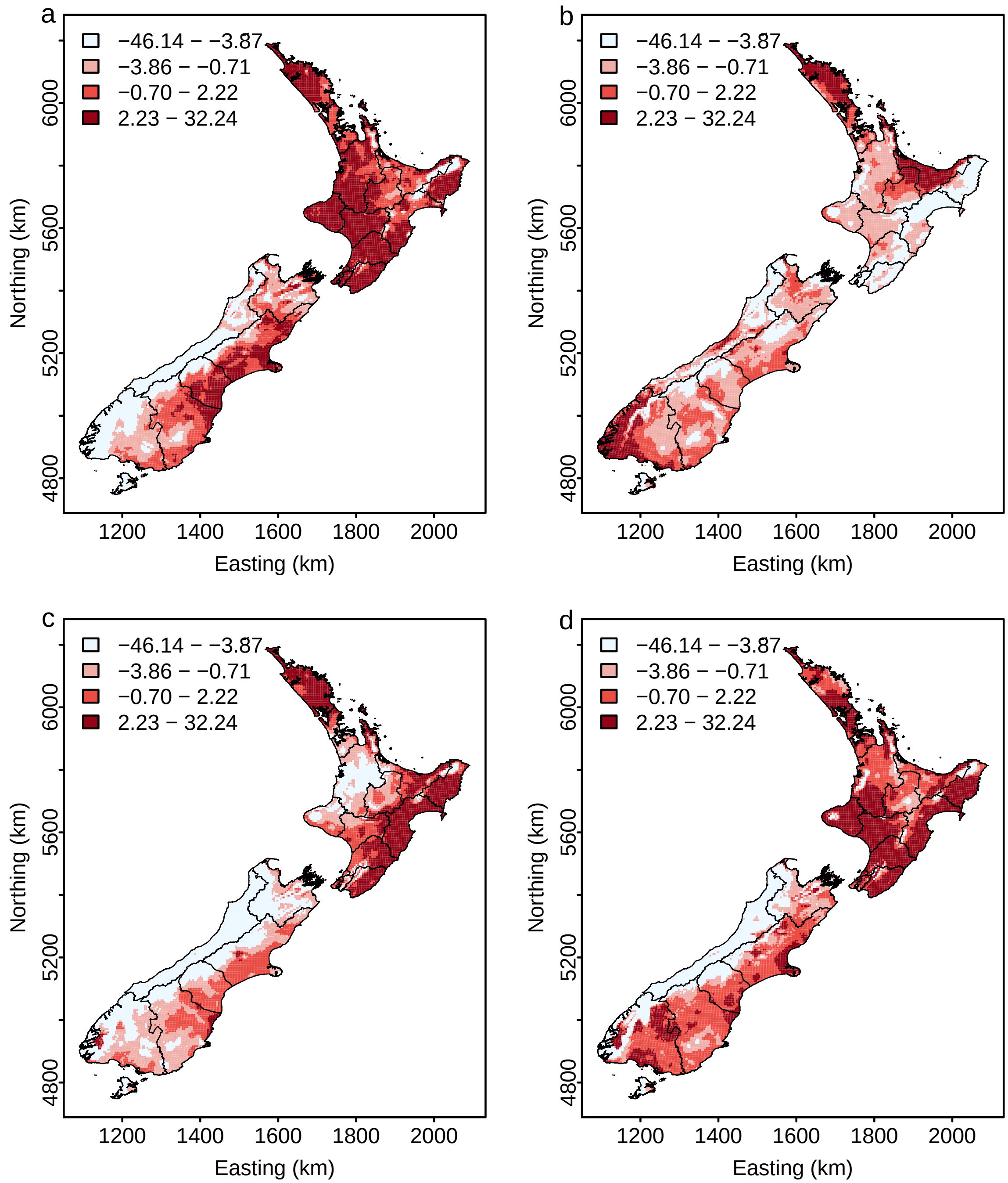


Figure 7.14: The projected change in cryptosporidiosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

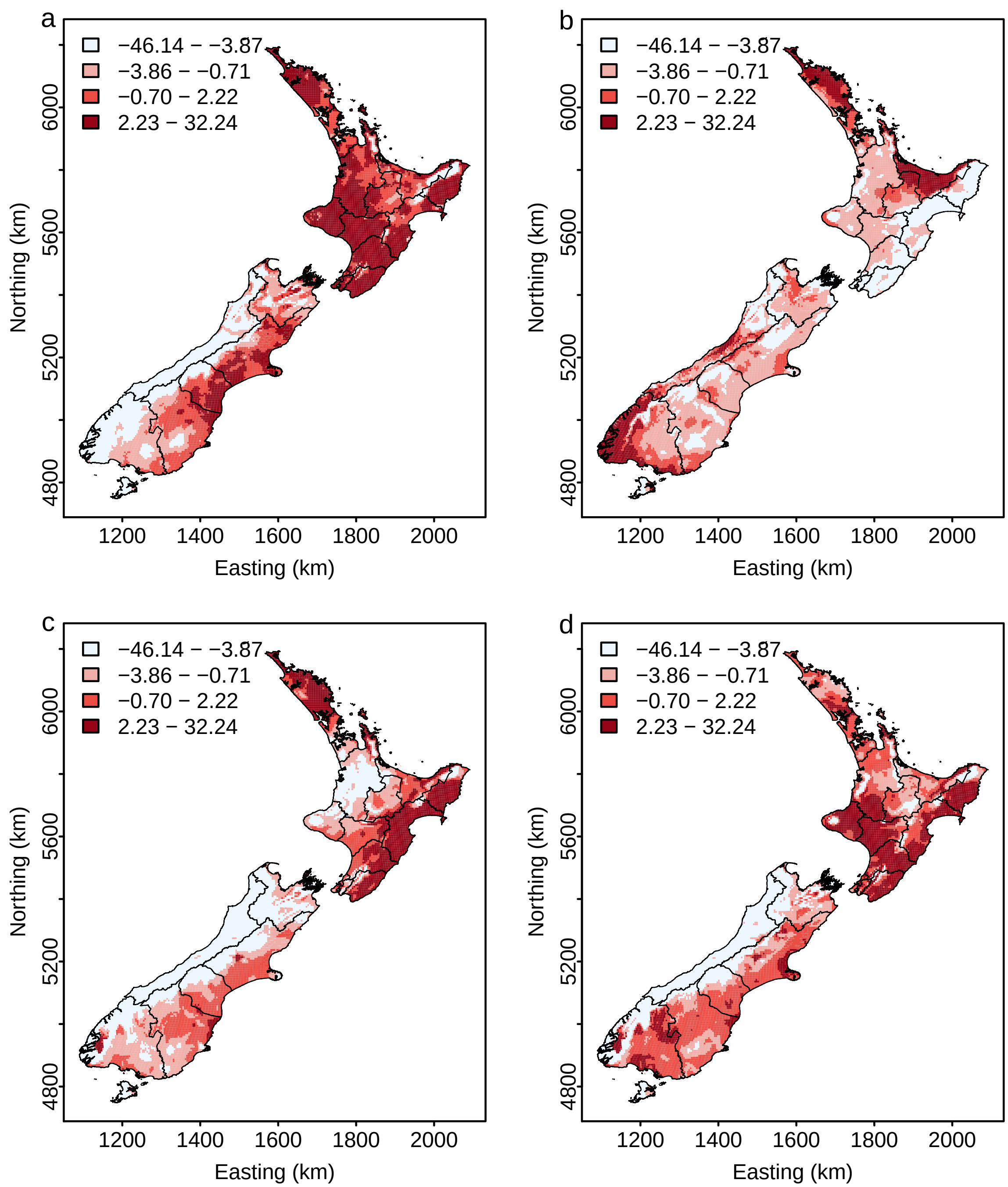


Figure 7.15: The projected change in cryptosporidiosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

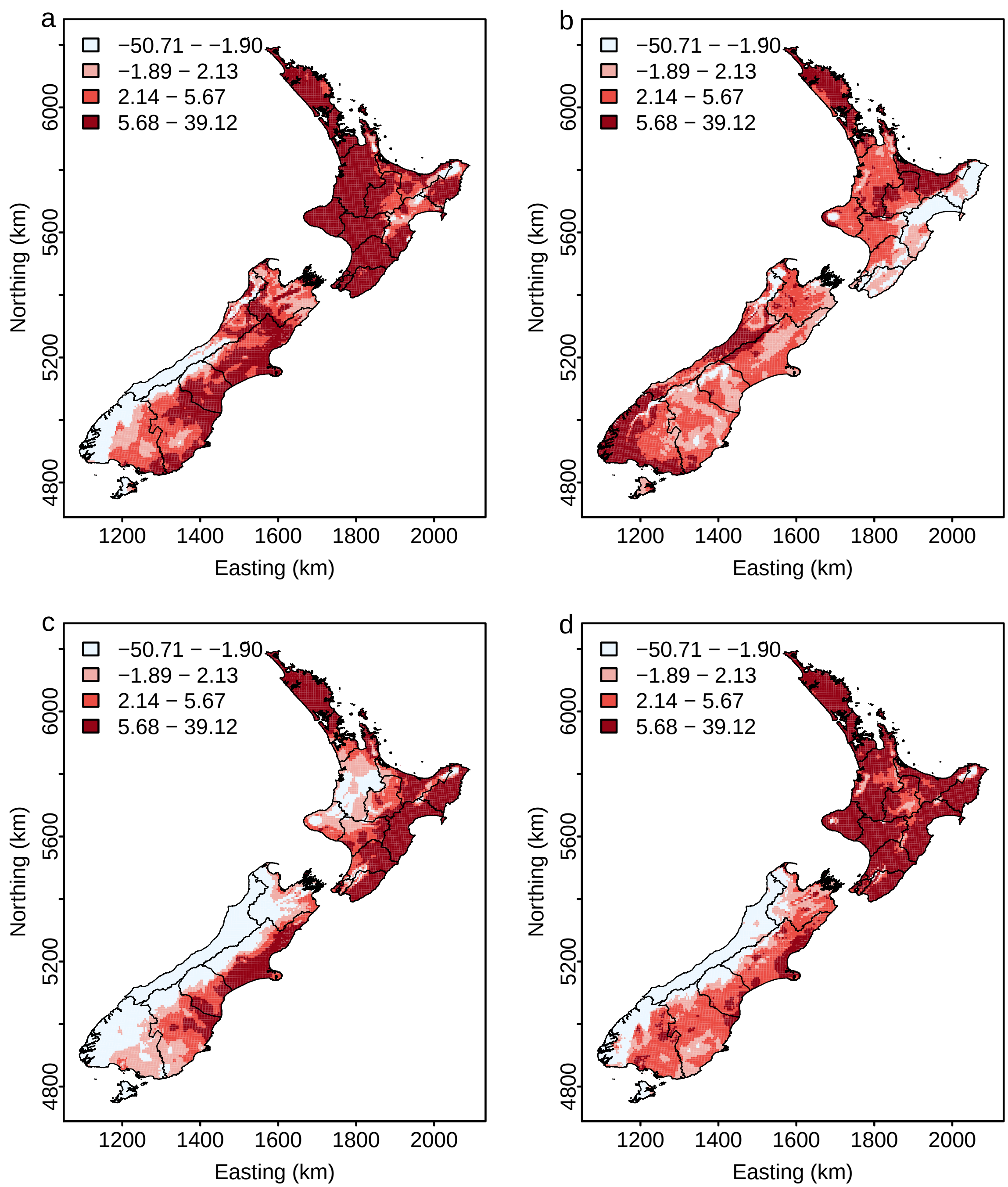


Figure 7.16: The projected change in cryptosporidiosis incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

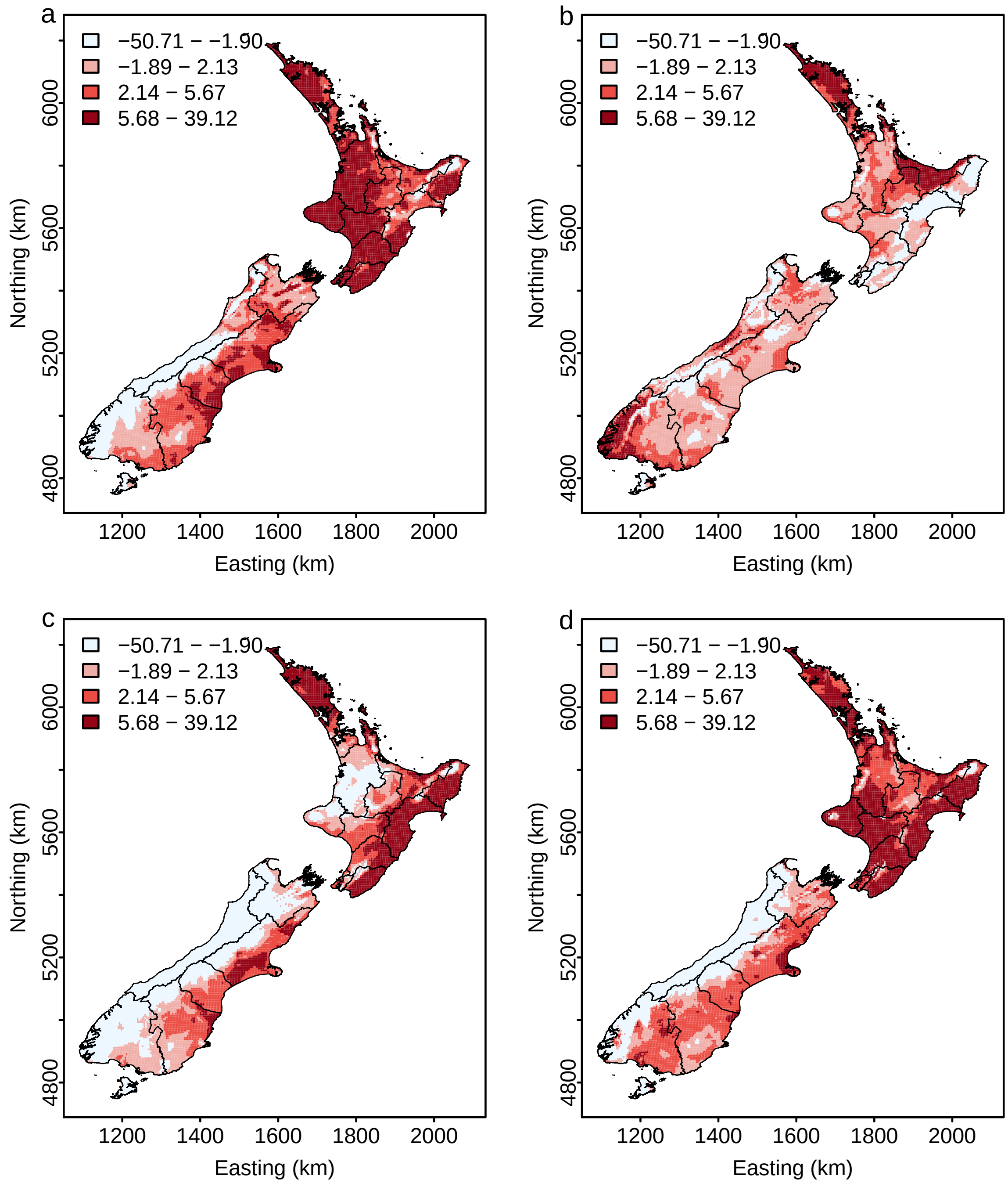


Figure 7.17: The projected change in cryptosporidiosis incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

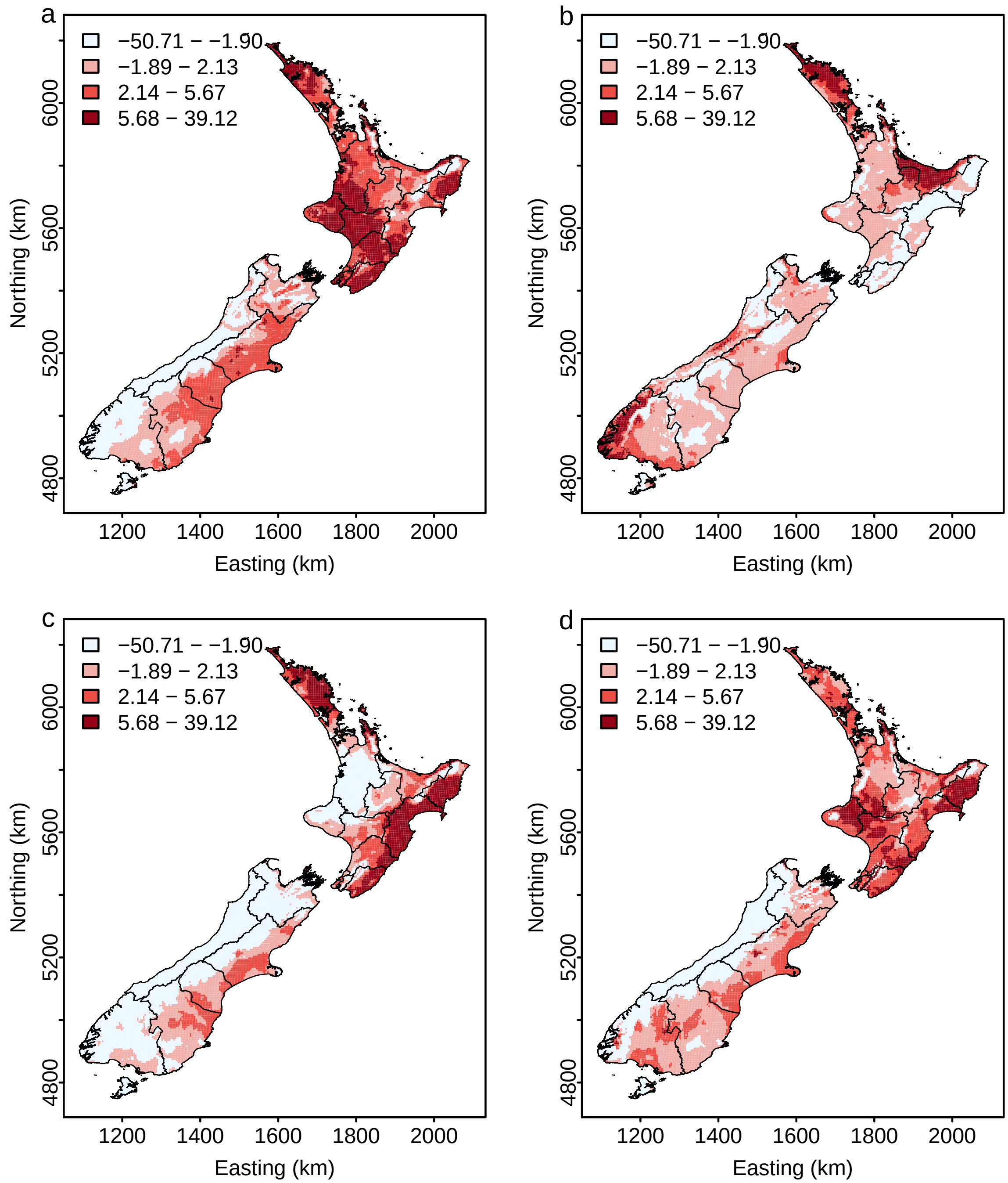


Figure 7.18: The projected change in cryptosporidiosis incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

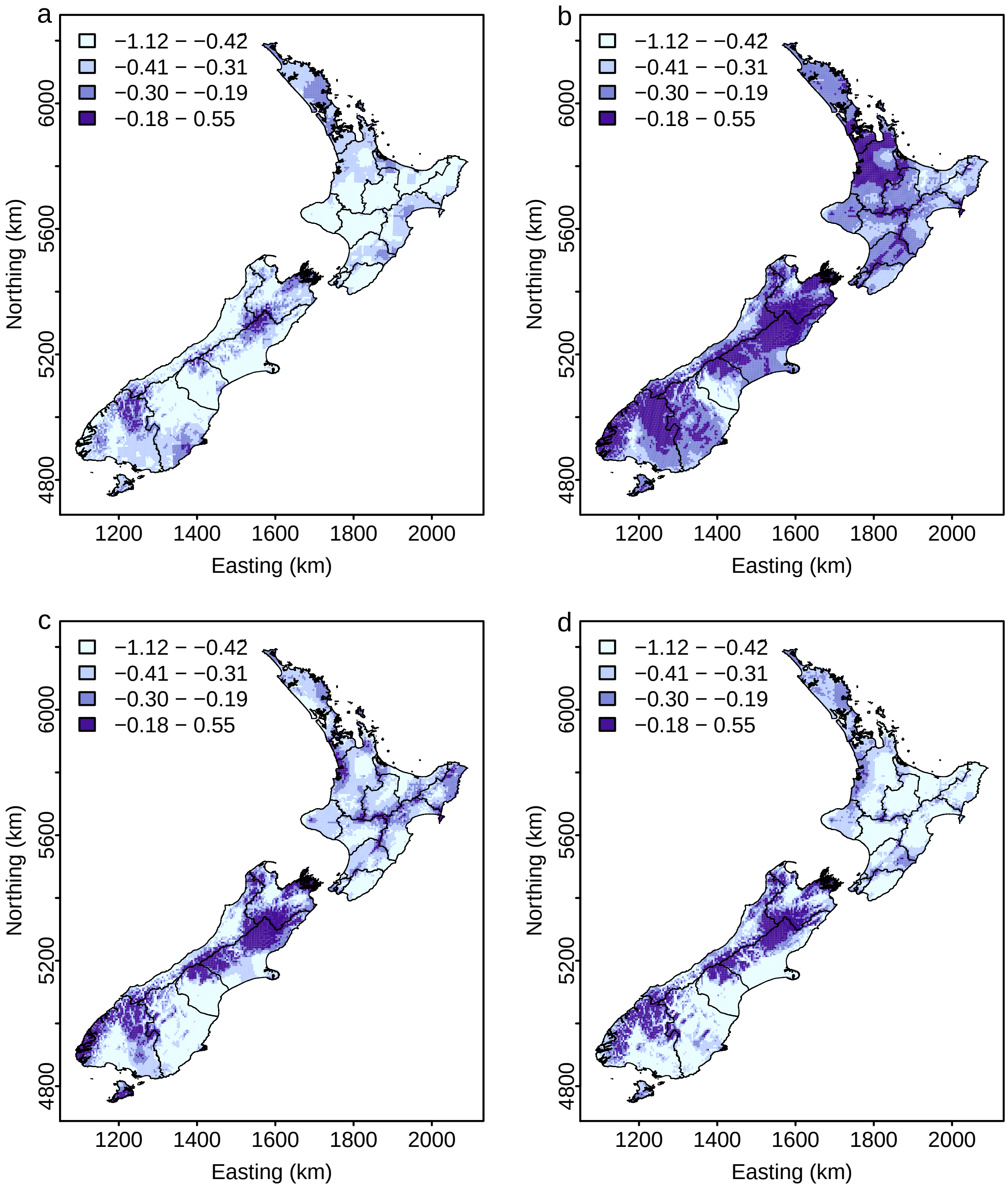


Figure 7.19: The projected change in influenza hospitalisations incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

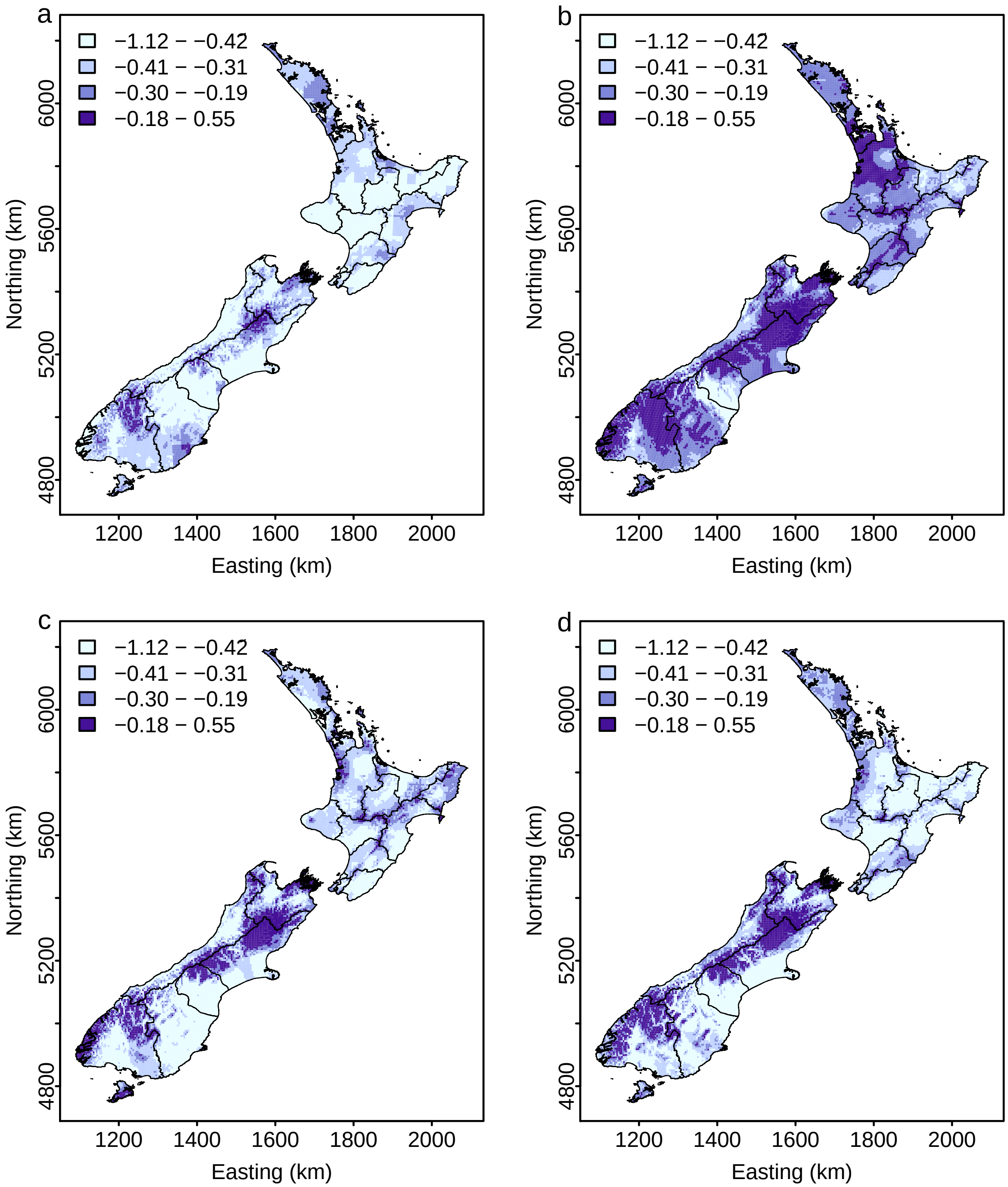


Figure 7.20: The projected change in influenza hospitalisations incidence risk the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

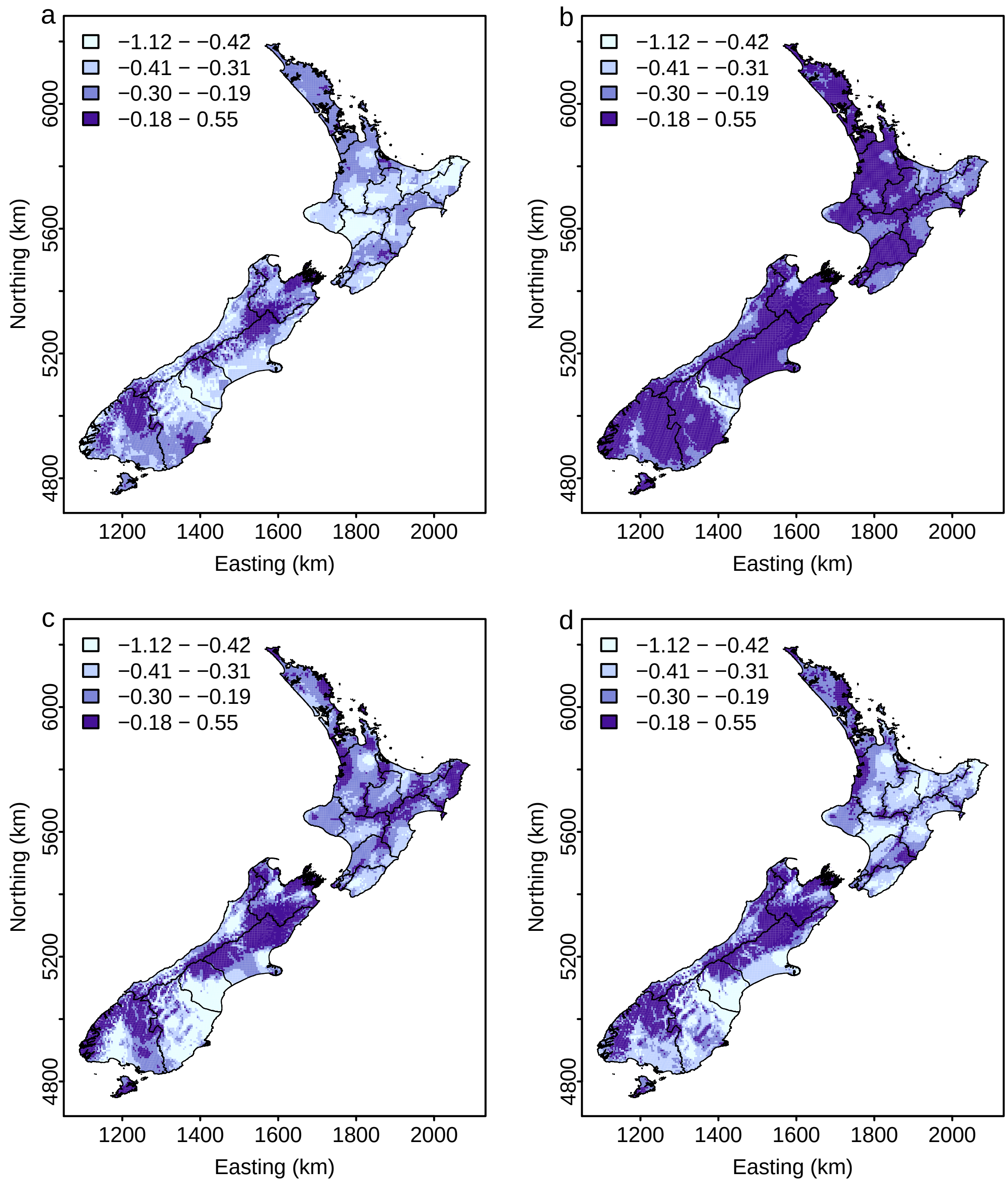


Figure 7.21: The projected change in influenza hospitalisations incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

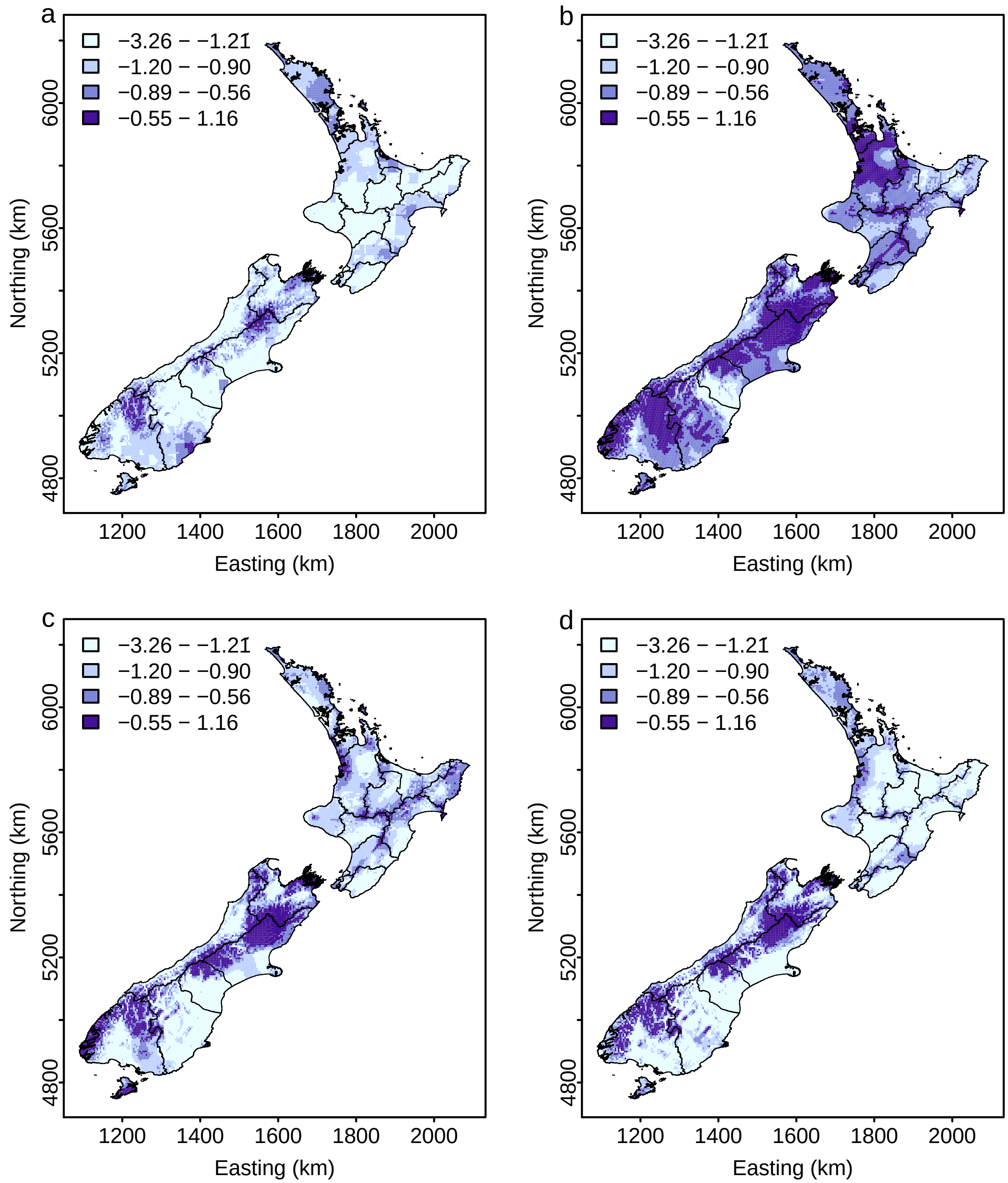


Figure 7.22: The projected change in influenza hospitalisations incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

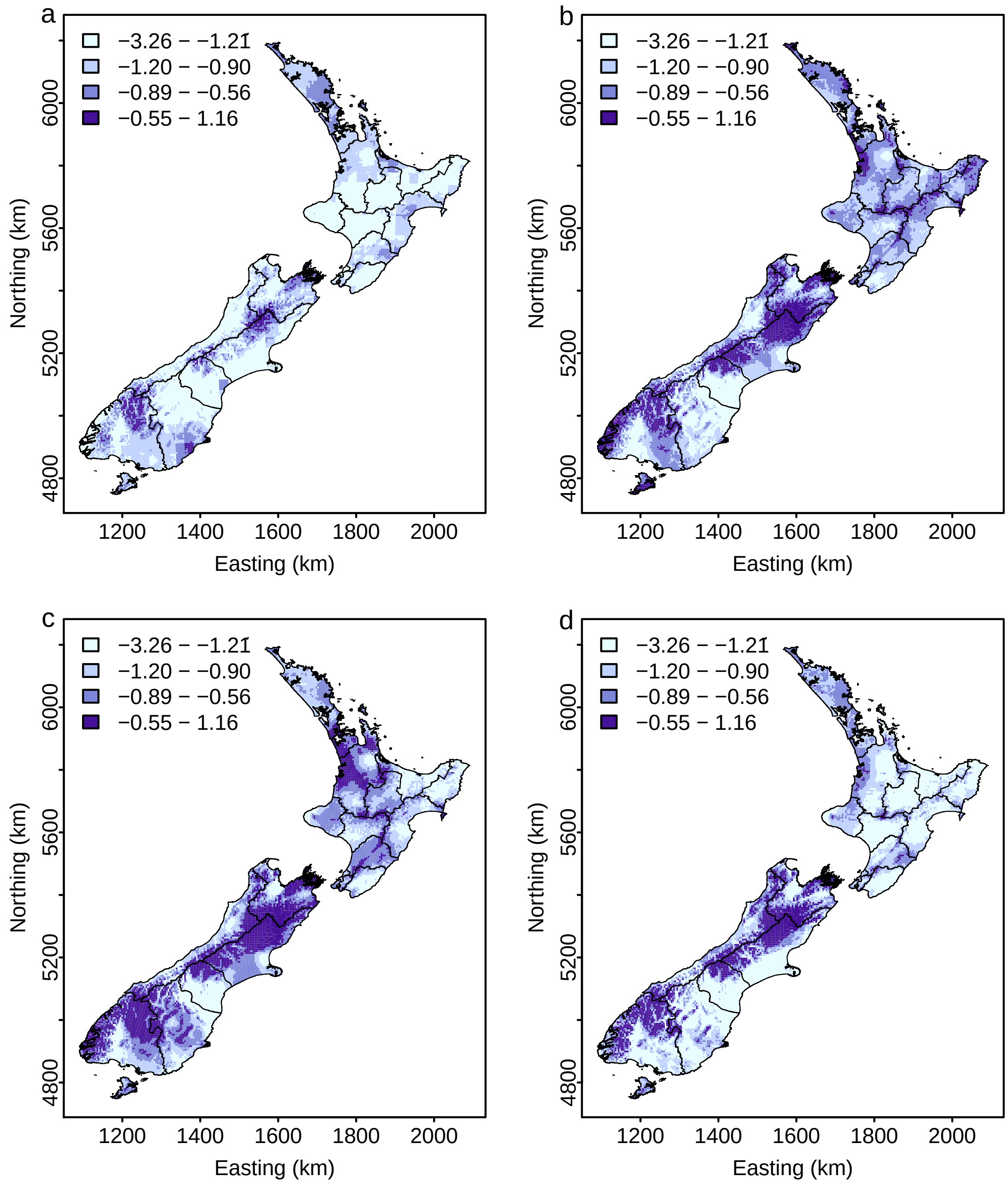


Figure 7.23: The projected change in influenza hospitalisations incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

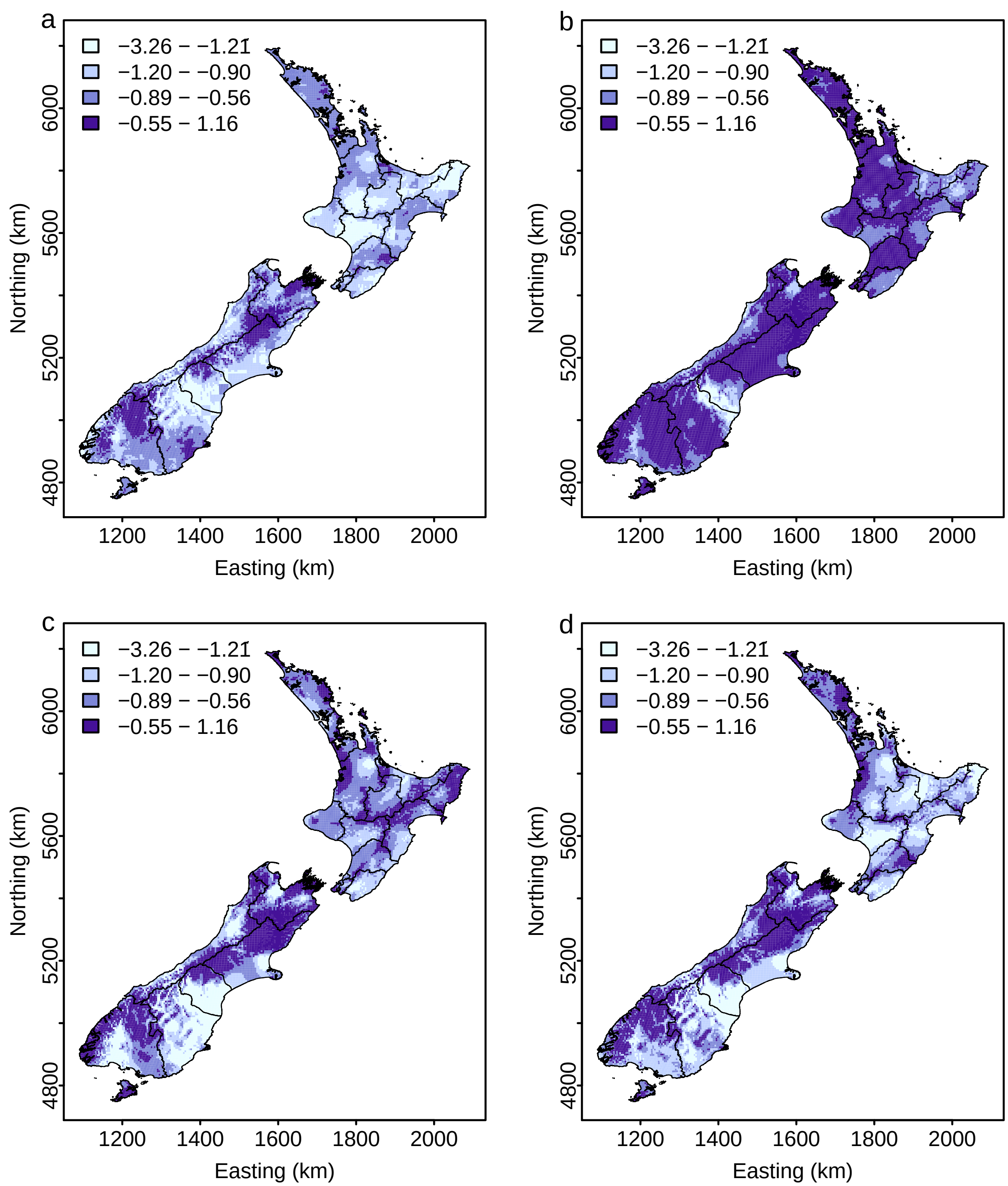


Figure 7.24: The projected change in influenza hospitalisations incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

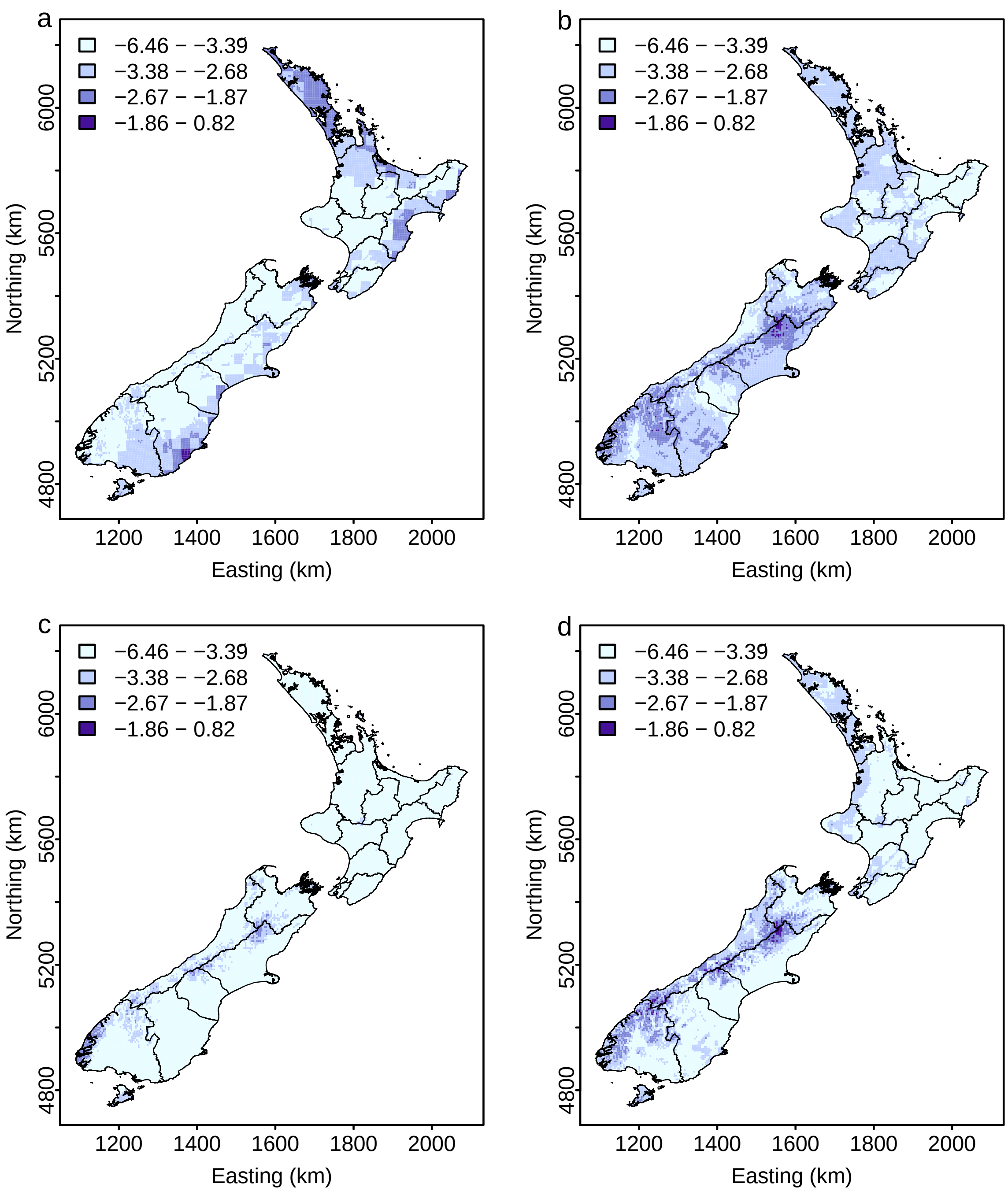


Figure 7.25: The projected change in influenza hospitalisations incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

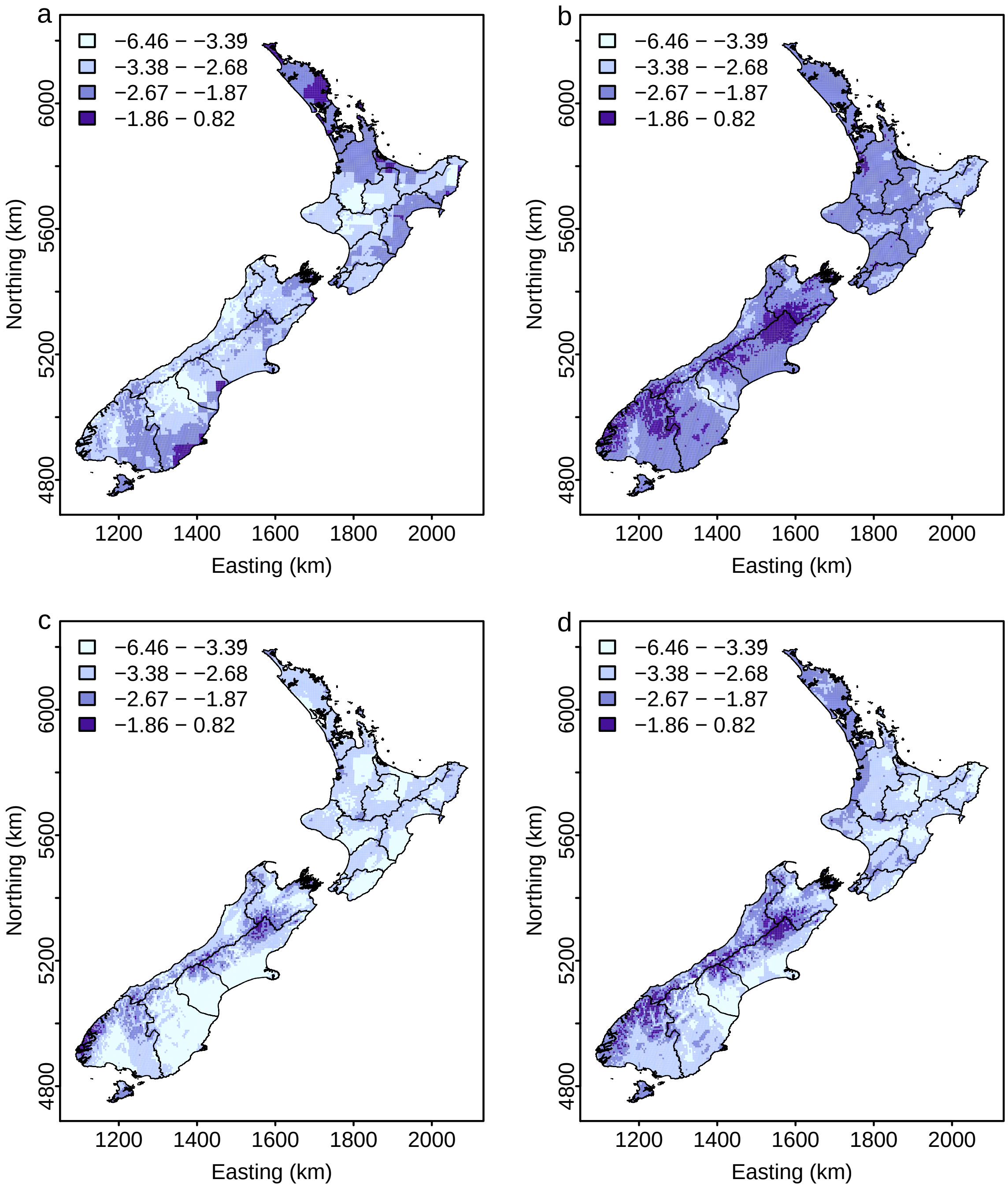


Figure 7.26: The projected change in influenza hospitalisations incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

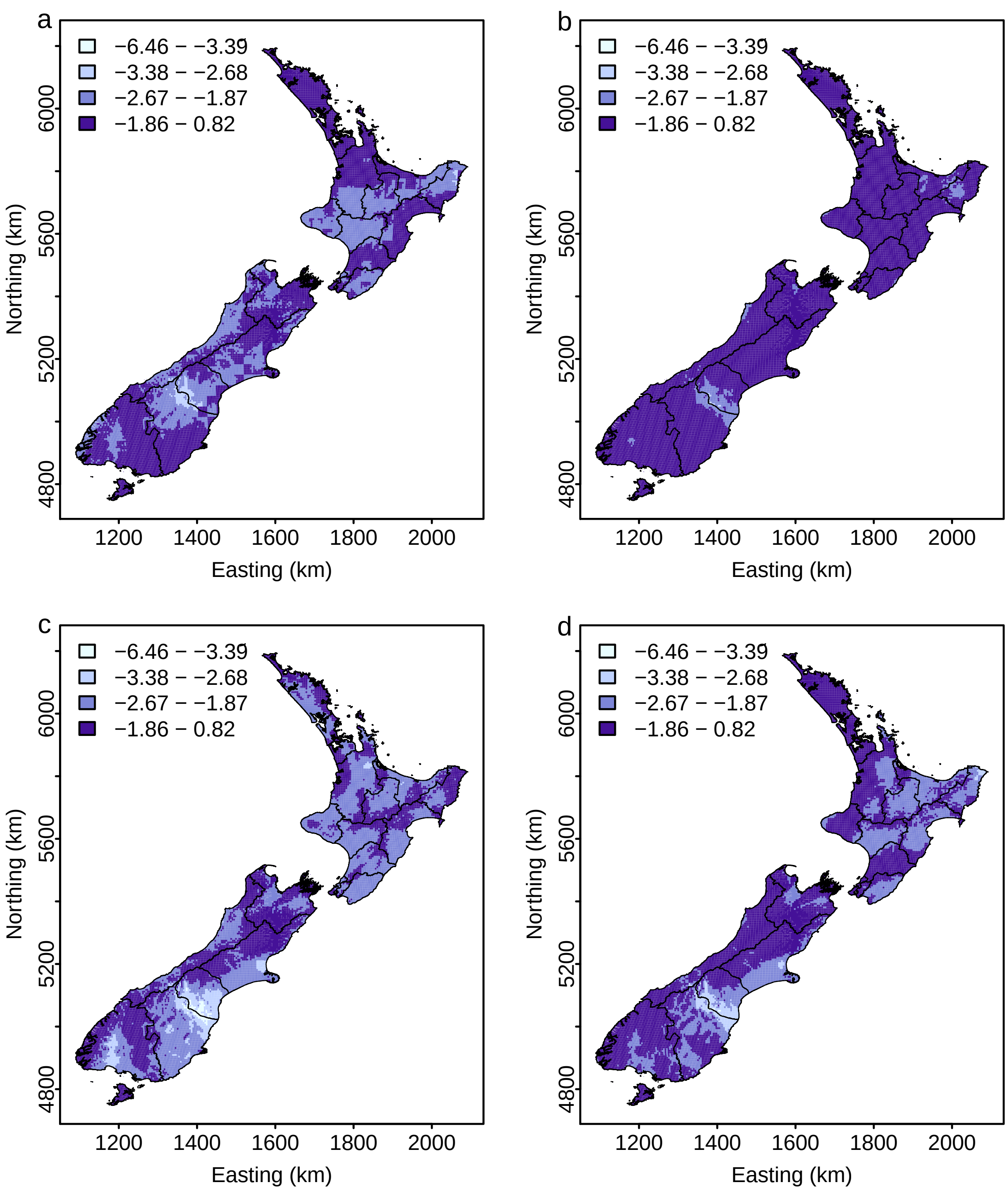


Figure 7.27: The projected change in influenza hospitalisations incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

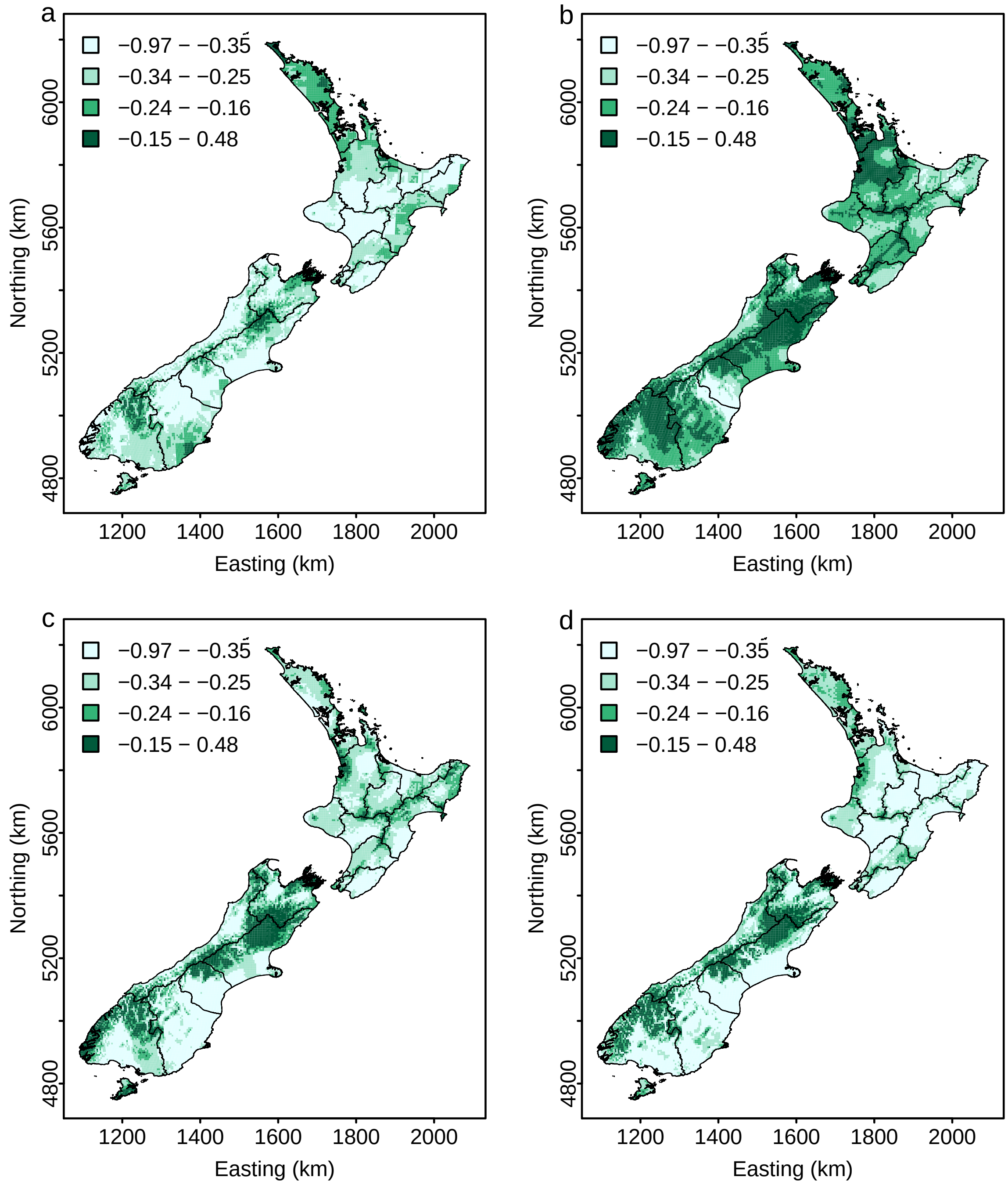


Figure 7.28: The projected change in meningococcal disease incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

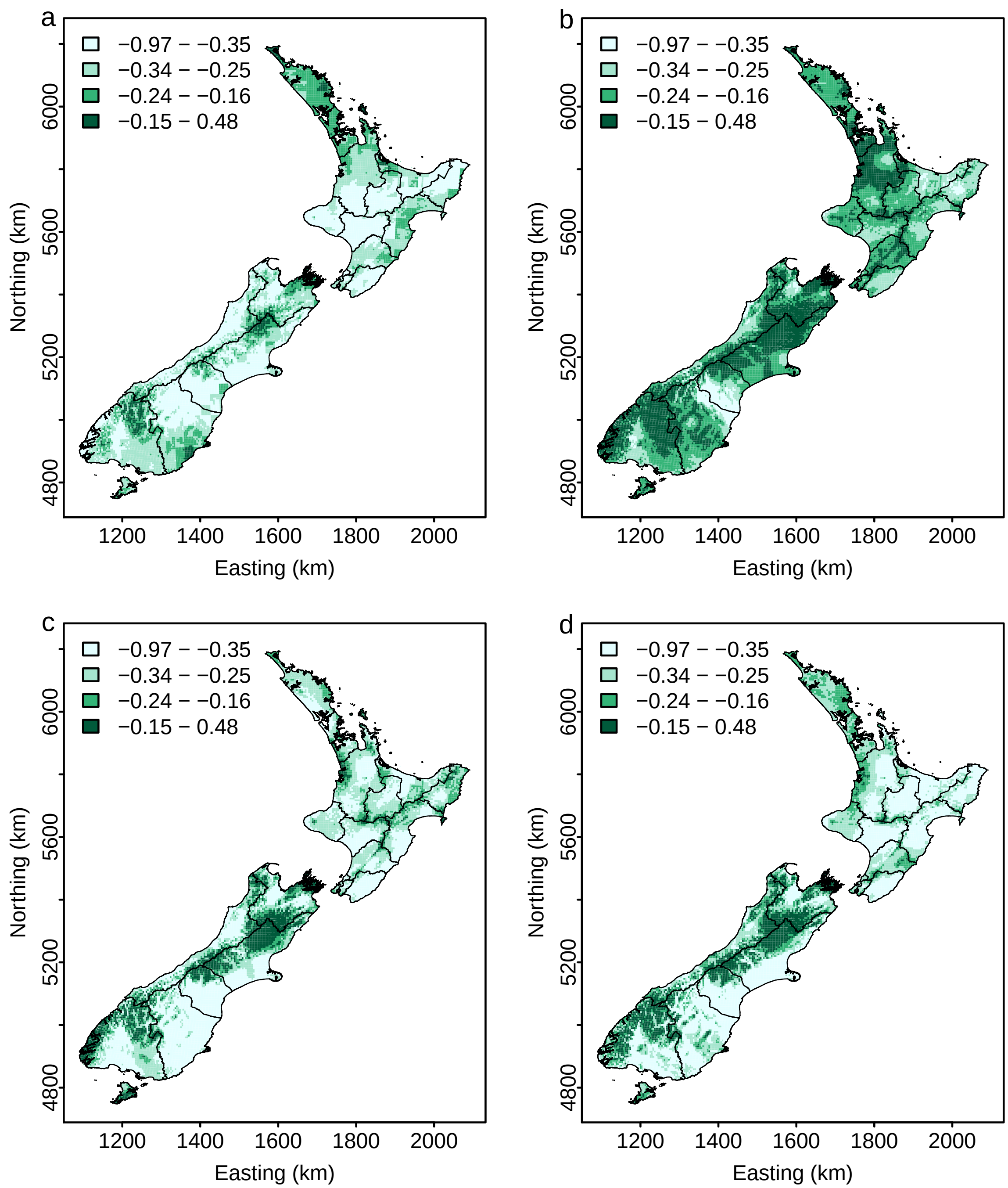


Figure 7.29: The projected change in meningococcal disease incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

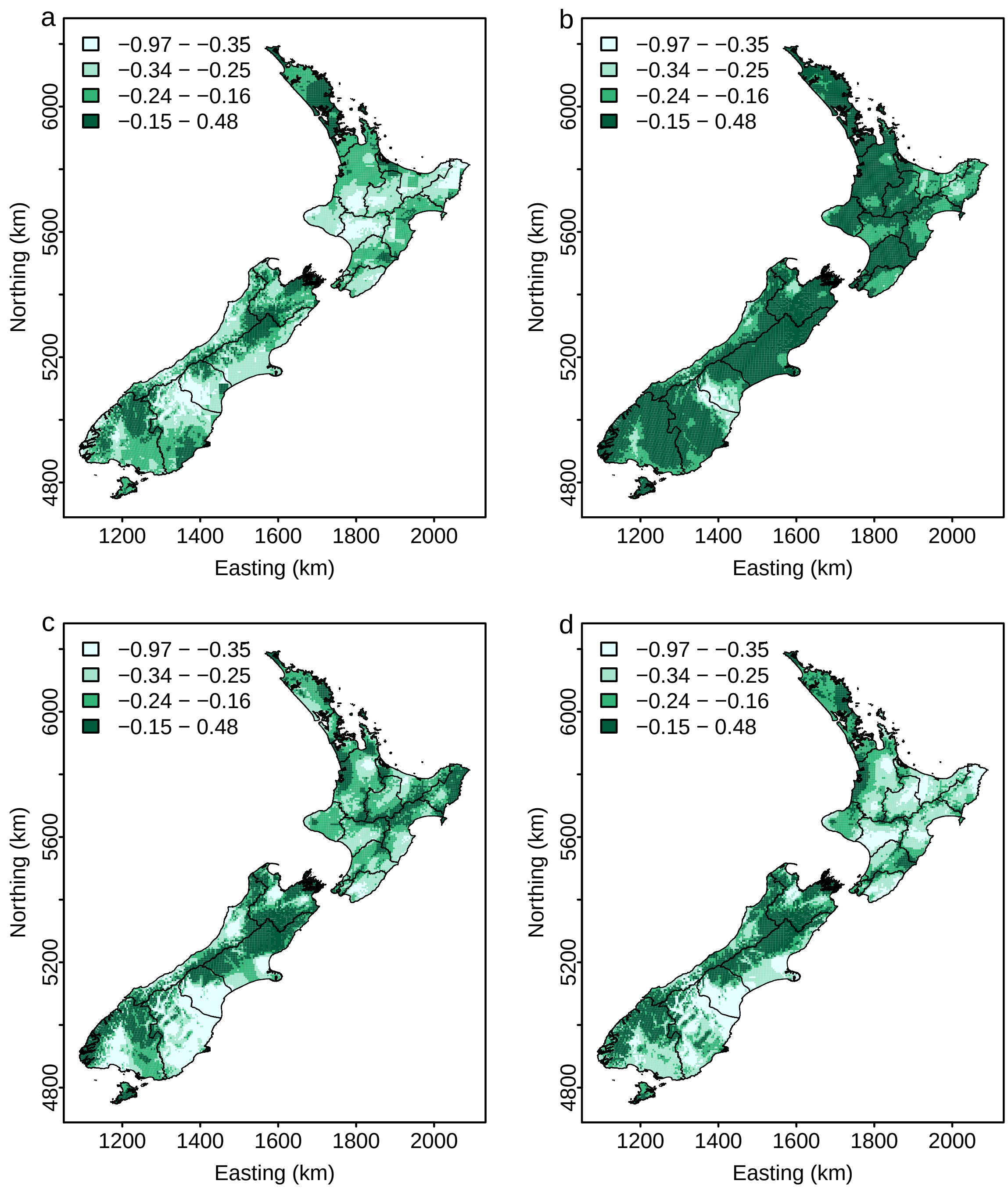


Figure 7.30: The projected change in meningococcal disease incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2015.

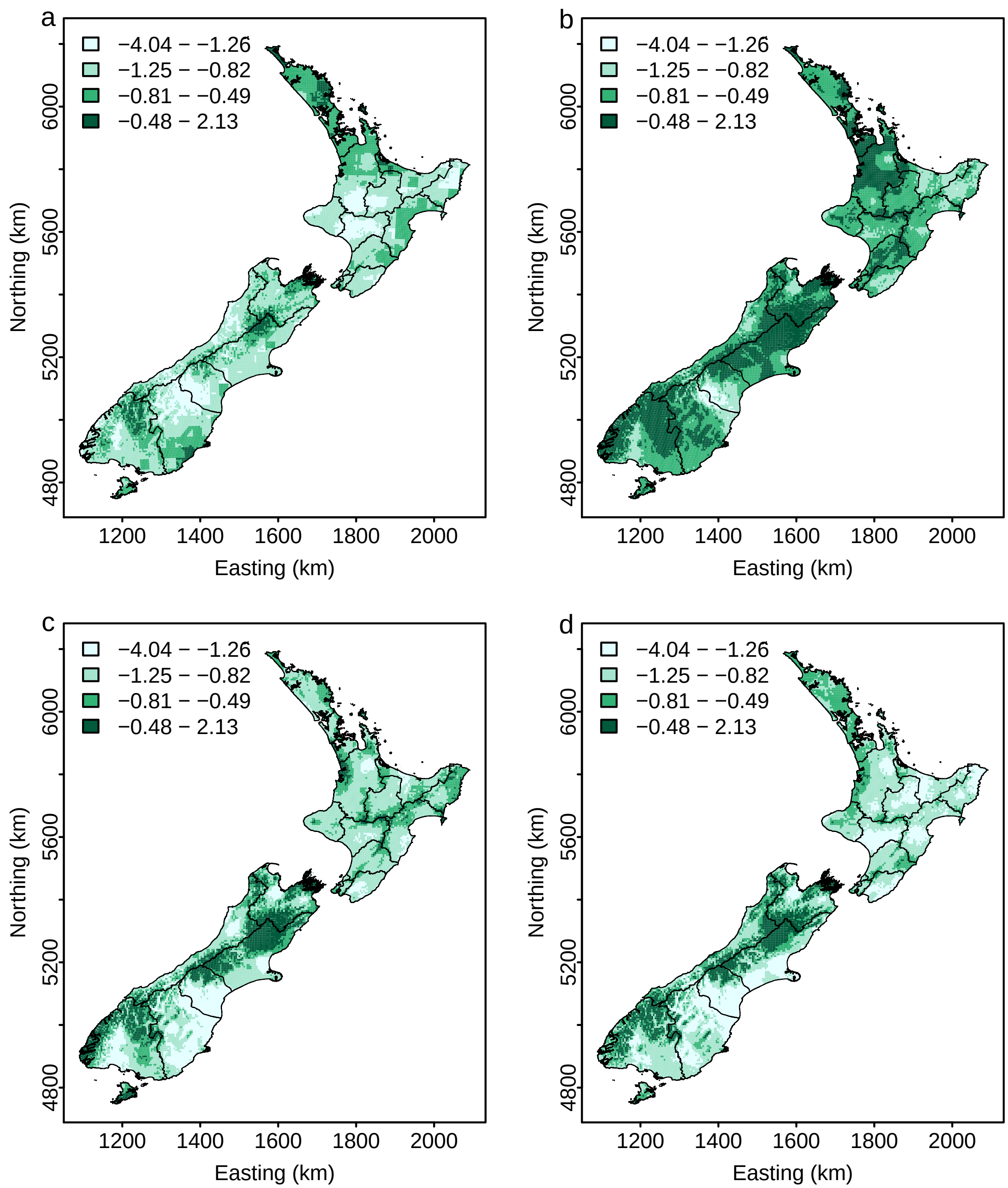


Figure 7.31: The projected change in meningococcal disease incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

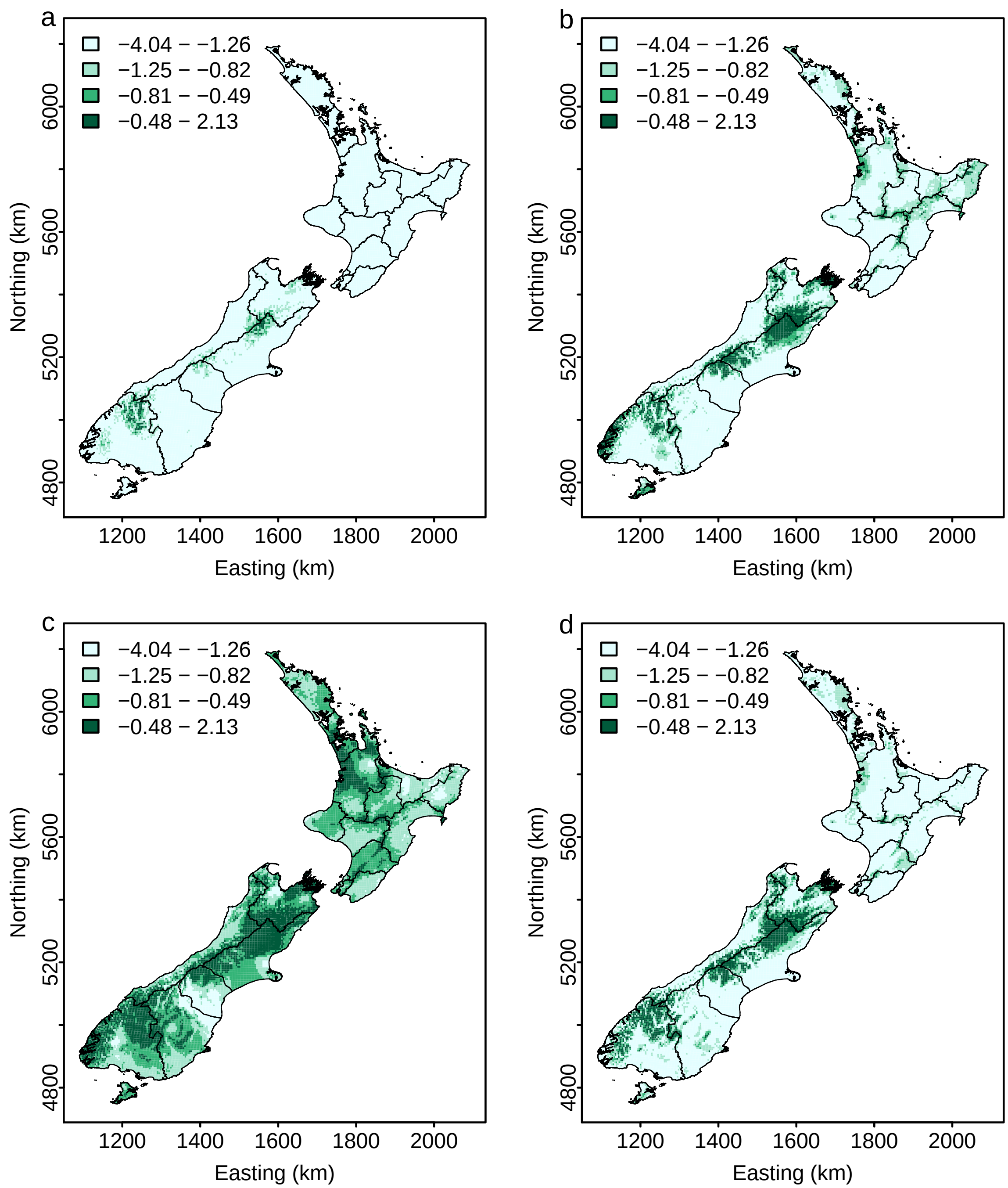


Figure 7.32: The projected change in meningococcal disease incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

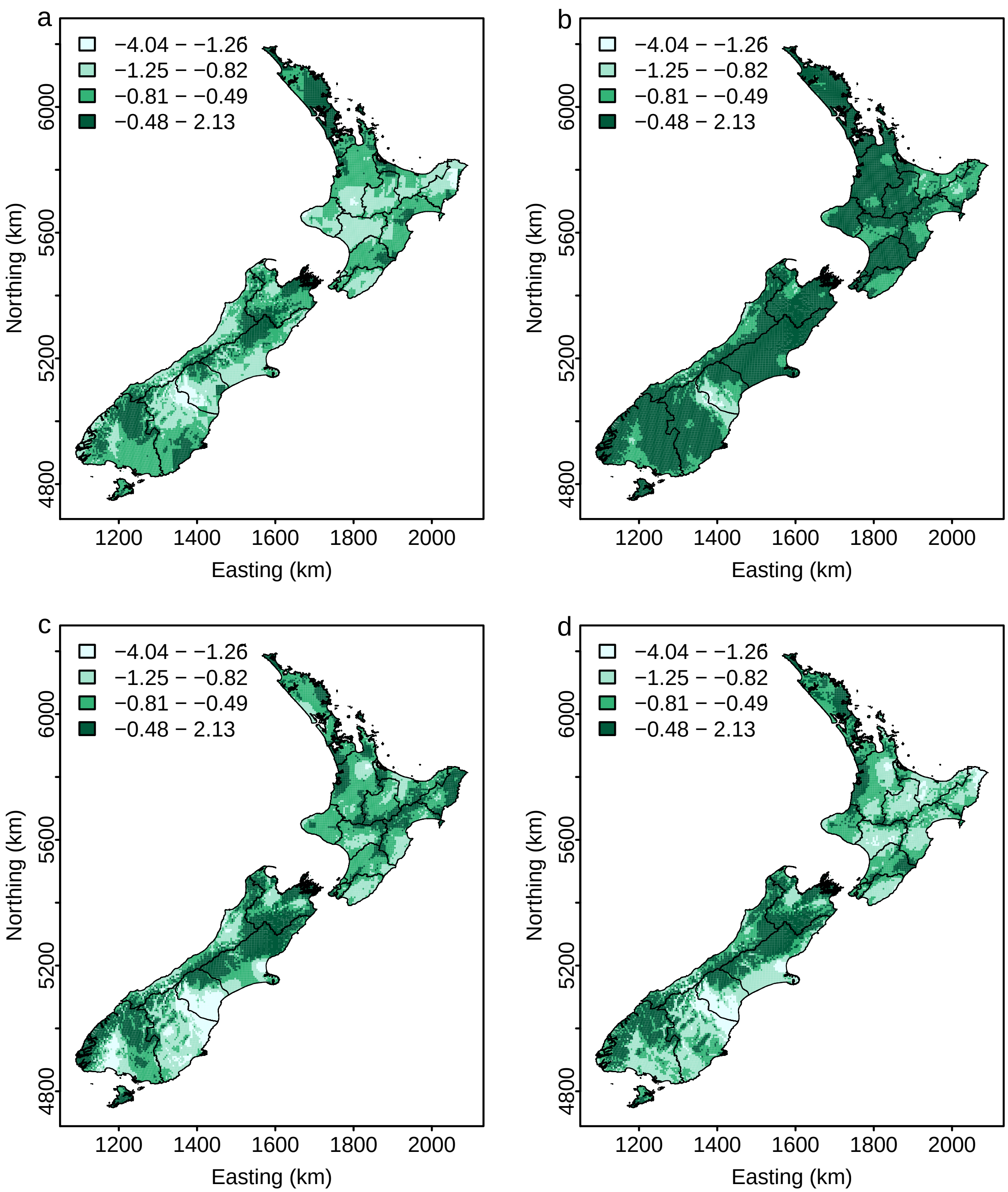


Figure 7.33: The projected change in meningococcal disease incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2040.

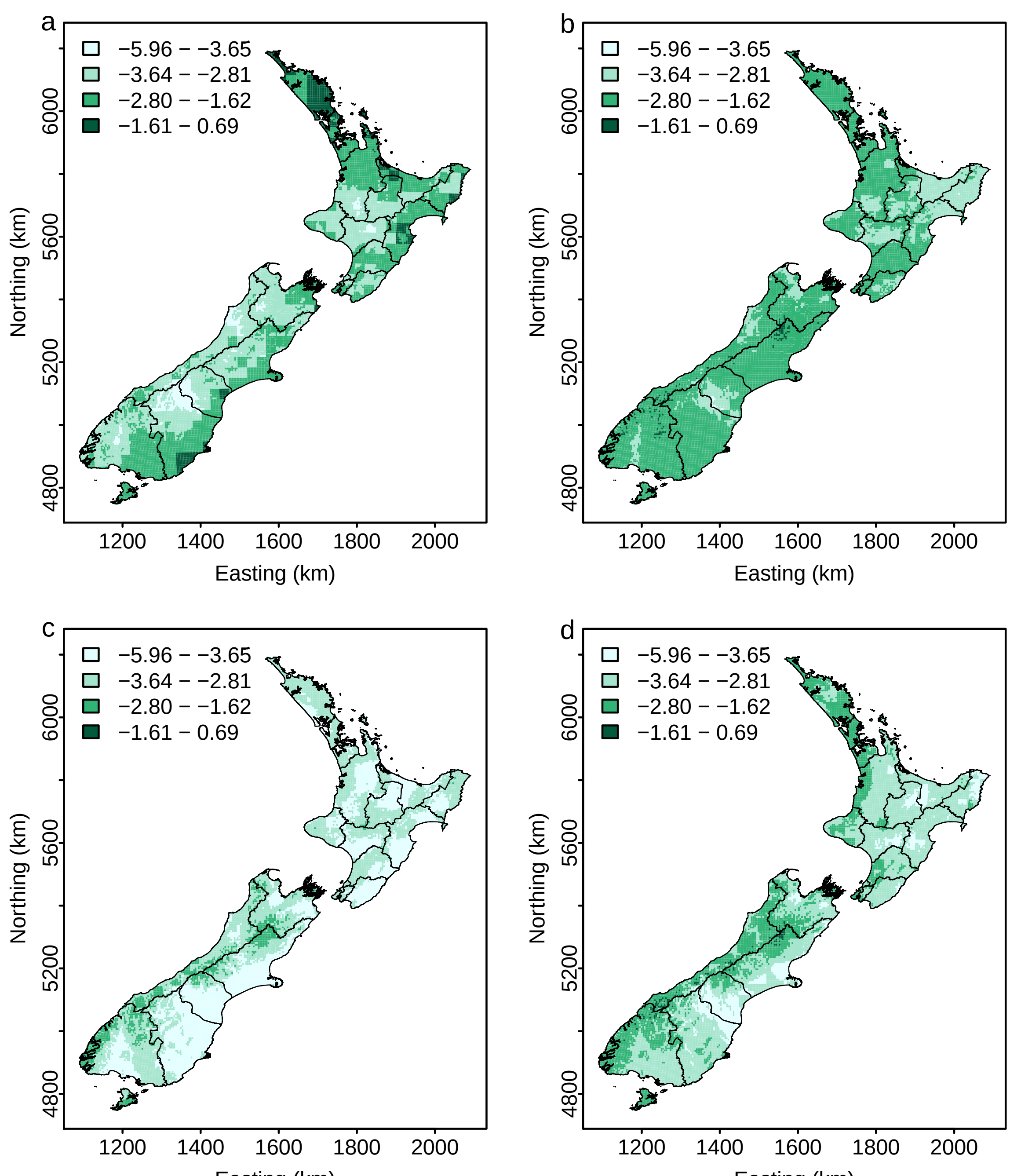


Figure 7.34: The projected change in meningococcal disease incidence risk under the A2 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

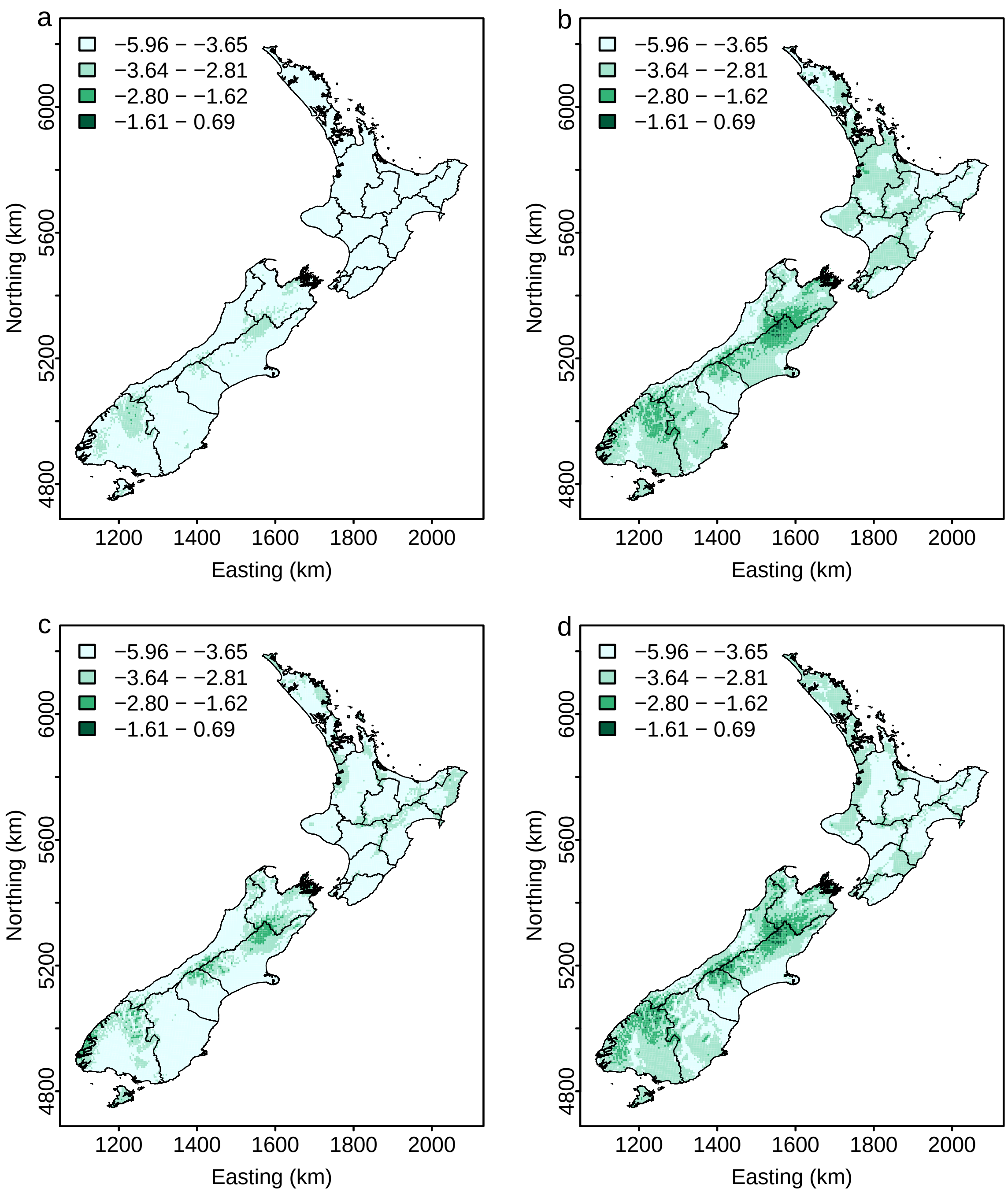


Figure 7.35: The projected change in meningococcal disease incidence risk under the A1B climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

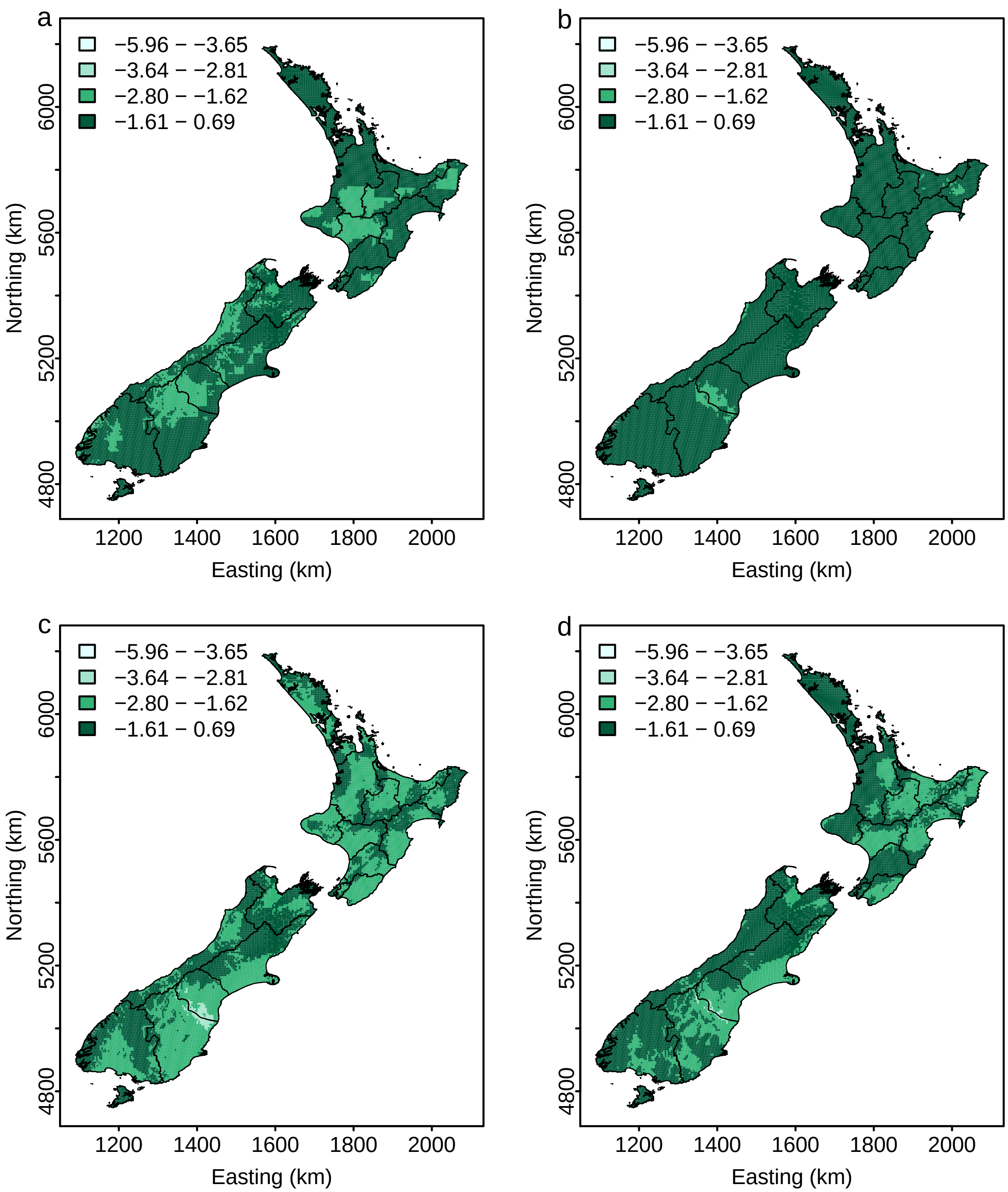


Figure 7.36: The projected change in meningococcal disease incidence risk under the B1 climate scenario for a) summer, b) autumn, c) winter, and d) spring 2090.

Chapter 8

Discussion

8.1 Introduction

The purpose of this chapter is to summarise the effect of past climate variation on infectious diseases in New Zealand and the implications these have on the burden of disease in New Zealand for the years 2015, 2040 and 2090, under three IPCC climate scenarios. The limitations experienced with data accuracy, complex datasets, modelling, bias and confounding will then be discussed. This chapter concludes with an outline of further research recommendations.

This thesis explored the association of past climate variation on the incidence risk (IR) of four infectious diseases throughout New Zealand at a fine spatial and temporal resolution from 1997 - 2007. The associations identified were then used to produce projection estimates for the percentage change in IR of disease for 2015, 2040 and 2090, under three IPCC climate scenarios.

The process started with the selection of four indicator diseases chosen to represent the effect of climate on infectious diseases throughout New Zealand (Chapter 4). Three locations in New Zealand were used in this process, namely Waikato, Manawatu and Canterbury. These regions were chosen as they contain a mixture of rural and urban areas, and represent a north-to-south gradient. The requirements of the selected indicator diseases were:

1. the combined average annual IR from 1997 - 2007 in the three locations was greater than 5 per 100,000 population, and

2. the disease exhibited seasonality.

The four indicator diseases chosen, and endorsed by the FRST Advisory Panel, were campylobacteriosis, cryptosporidiosis, influenza hospitalisations and meningococcal disease.

The past association of weather variables and other possible confounders with the IR of each indicator disease was then explored during a screening process in Chapter 5. The modelling technique used was a quasi-Poisson generalized linear model (GLM).

In Chapter 6, a modified Knorr-Held Richardson model was used to explore the effect of past climate variation on the four indicator diseases in New Zealand from 1997 - 2007, at a fine spatial and temporal level (574 weekly measurements for each of the 11,491 VCS grids), having adjusted for confounders. Expert opinion on epidemiological plausibility, alongside the quasi-Poisson GLM screening results were used to select the confounding variables for each model.

The posterior coefficient estimates from the extended Knorr-Held Richardson model formed the basis of the projection calculation (Chapter 7). The projection calculation was used to estimate the percentage change in incidence risk (IR) for each indicator disease in New Zealand for the years 2015, 2040 and 2090. This was carried out using a combination of three IPCC emission scenarios (A2, B2 and A1B) and 12 downscaled global climate models. The projection data for the weather variables was supplied by NIWA, with the confounding variables assumed to remain constant with respect to time.

This thesis began with an overview of the statistical model approaches considered for the assessment of the past association and projection of the effect of climate variation on infectious disease in New Zealand in Chapter 2, with a strong focus on techniques used in this report. Chapter 3 described the format of the data used in this thesis, obtained from various organisations throughout New Zealand.

8.2 Key Findings

This purpose of this section is to use the findings of the past association model and projection calculations to answer the aims of this thesis, along with supporting evidence from other sources of literature. The two aims of this study were to:

1. Investigate the effect of climate variation on four indicator diseases throughout New Zealand from 1997 - 2007 inclusive, at a fine spatial and temporal level, and

2. Use the identified associations to produce projection models for the percentage change in incidence risk of these diseases in 2015, 2040 and 2090.

8.2.1 Campylobacteriosis

The results from the past association model for the incidence of notified campylobacteriosis in New Zealand from 1997 - 2007, suggest that the weekly average absolute humidity weather variable is significantly associated ($\hat{\beta}_1$=0.143, p-value=0.000). An increase of one standard deviation in the weekly average absolute humidity in grid i changes the expected rate of campylobacteriosis notification in that grid by a factor of 1.154 ($\Delta\lambda_{it} = e^{0.143}$), when all other variables are held constant. In contrast, past studies have identified an association between temperature and sunshine hours with campylobacteriosis [78], where in contrast we found the average weekly temperature to be insignificantly associated (p-value=0.505) with the incidence of notified campylobacteriosis in New Zealand, having adjusted for other explanatory variables.

Other risk factors significantly associated with the rate of a campylobacteriosis notifications in grid i were:

- the proportion greater than 65 years old ($\Delta\lambda_{it}$=1.201, p-value=0.000),
- beef ($\Delta\lambda_{it}$=1.032, p-value=0.011) and dairy density ($\Delta\lambda_{it}$=1.043, p-value=0.000),
- intermediate ($\Delta\lambda_{it}$=1.057, p-value=0.000) and poor drinking water quality ($\Delta\lambda_{it}$=1.093, p-value=0.000), and
- the proportion who identify with Pacific Islander ($\Delta\lambda_{it}$=1.065, p-value=0.002) and Asian ethnicities ($\Delta\lambda_{it}$=1.024, p-value=0.002).

The association with animal density and untreated or contaminated drinking water with campylobacteriosis in New Zealand have been previously identified as risk factors in past studies [29, 30, 32, 58].

Protective factors identified with campylobacteriosis notification in grid i were:

- the proportion aged less than 4 years ($\Delta\lambda_{it}$=0.935, p-value=0.008), and
- the proportion who identify with Maori ethnicity ($\Delta\lambda_{it}$=0.667, p-value=0.000).

The beta coefficient estimates for the projection calculations for the average absolute humidity was $\tilde{\beta}_1$=85.05 (p-value = 0.000), the average rainfall was $\tilde{\beta}_2$=0.0009 (p-value = 0.387) and the weekly average temperature was $\tilde{\beta}_3$=0.0026 (p-value = 0.505). These beta coefficient estimates are all positive, indicating an increase in campylobacteriosis in New Zealand from the study period (1997 - 2007) to 2015, 2040 and 2090, is associated with projected increases in rainfall and absolute humidity. This increase is reflected the most part in the projected percentage change under each climate scenario, where the percentage change in IR during 2015 ranges from -1.40% - 4.92%, increasing to -4.10% - 15.06% in 2040 and -1.37% - 30.10% in 2090. For each projection year, the changes in summer indicate an increase in IR of campylobacteriosis notifications, with a small to no decrease in the change of IR during autumn, winter and spring.

8.2.2 Cryptosporidiosis

The results from the past association model for the incidence of notified cryptosporidiosis in New Zealand, from 1997 - 2007, suggest that the weekly average rainfall and temperature weather variables are significantly associated ($\hat{\beta}_1$=-0.094, p-value=0.002 and $\hat{\beta}_2$=0.124, p-value=0.004, respectively). An increase of one standard deviation in the weekly average rainfall in grid i, changes the expected rate of cryptosporidiosis notification in that grid by a factor of 0.910 ($\Delta\lambda_{it} = e^{-0.094}$), when all other variables are held constant. It follows that an increase of one standard deviation in the weekly temperature in grid i, changes the expected rate of cryptosporidiosis notification in that grid by a factor of 1.132 ($\Delta\lambda_{it} = e^{0.124}$), when all other variables are held constant. A positive association between temperature and cryptosporidiosis in New Zealand during summer and autumn has also been identified by [46].

Other risk factors identified with cryptosporidiosis notifications in grid i were:

- the proportion aged less than 4 years ($\Delta\lambda_{it}$=1.175, p-value=0.000),
- poor drinking water ($\Delta\lambda_{it}$=1.108, p-value=0.048), and
- living in rural areas ($\Delta\lambda_{it}$=1.570, p-value=0.000).

Protective factors identified with the rate of cryptosporidiosis notification in grid i were:

- unknown drinking water quality ($\Delta\lambda_{it}$=0.705, p-value=0.002), and
- the proportion who identify with Maori ethnicity ($\Delta\lambda_{it}$=0.711, p-value=0.000).

Similar findings were identified by Snel et al. (2009), where cattle density was found to be associated with an increase rate of cryptosporidiosis [71]. In our study, rurality was used as a proxy for animal density, and the rate of cryptosporidiosis notifications increased by 1.570 times in rural areas compared to urban areas. The association of farm animals and rural areas was also identified by Lake et al. (2007) [46] and Learmonth et al. (2004) [48]. Drinking water quality, especially when contaminated by cattle manure [81], has also been previously identified as a risk factor for cryptosporidiosis in New Zealand [24, 72].

The beta coefficient estimates for the projection calculations for the average rainfall was $\tilde{\beta}_1$=-0.0123 (p-value = 0.002) and the weekly average temperature was $\tilde{\beta}_2$=0.0276 (p-value = 0.004). The beta coefficient estimate for the average rainfall was negative, indicating a decrease in the projected cryptosporidiosis is associated with an increase in projected rainfall; whereas the beta coefficient for the average temperature was positive, indicating an increase in the projected cryptosporidiosis is associated with an increase in projected temperature. The projections for 2015, 2040 and 2090 suggest an overall increase in cryptosporidiosis IR throughout New Zealand, with an increase in magnitude of projected cryptosporidiosis IR ranging from -19.07% - 10.03% in 2015 to -50.71% - 39.12% in 2090. This increase is suggested during summer, winter and spring throughout New Zealand, with a decrease projected in the west of the South Island. In autumn, an increase in IR is suggested in the Northland, Waitemata, Auckland, Bay of Plenty and parts of the Southland DHBs, a decrease on the east coast of the North Island, and no change throughout the rest of the country.

8.2.3 Influenza Hospitalisations

The results from the past association model suggests that the incidence of influenza hospitalisations from 1997 - 2007, is not associated with variations in the absolute humidity, rainfall or temperature throughout New Zealand. An association with temperature was hypothesised by Lofgren et al. (2007), where it was suggested cooler temperatures cause behavioural patterns in humans that aid viral transmission [51].

The only risk factor identified with the rate of influenza hospitalisations in grid i were the proportion who identify with Maori ethnicity ($\Delta\lambda_{it}$=1.218, p-value=0.008). This finding is supported in a study carried out on hospitalisations for the H1N1 pandemic in New Zealand during 2009, where Maori and Pacific Islanders were identified to have a higher incidence of hospitalisation [77]. The higher incidence seen in Maori when compared to New Zealand European, could be driven by a greater readiness of Maori to self-refer for assessment at the emergency departments of a hospital if unwell [77]. In addition, Maori who may be subject to lower socio-economic conditions could be more likely to contract influenza that would lead to hospitalisations. There were no protective factors identified.

The beta coefficient estimates for the projection calculations for the average absolute humidity was $\tilde{\beta}_1$=27.47 (p-value = 0.918) and the average temperature was $\tilde{\beta}_2$=-0.0283 (p-value = 0.541). The beta coefficient for the average absolute humidity was positive, indicating an increase in projected influenza hospitalisations is associated with an increase in projected absolute humidity. The beta coefficient for average temperature was negative, indicating a decrease in the projected influenza hospitalisations is associated with an increase in projected temperature. The projected percentage change in influenza hospitalisations from the study period to 2015, 2040 and 2090 indicate a decrease in influenza hospitalisations, ranging from -1.12% - 0.55% in 2015 to -3.26% - -1.16% in 2090. Throughout the three projection years, the projected change in IR in summer and autumn indicated a decrease in influenza hospitalisation IR. Projections in winter and spring also indicate a decrease in IR, with no change suggested in parts of Nelson-Marlborough, Canterbury and Otago DHBs.

8.2.4 Meningococcal Disease

The results from the past association model suggests that the incidence of notified meningococcal disease from 1997 - 2007, is not associated with absolute humidity, rainfall or temperature variations throughout New Zealand. In contrast, a study carried out in Auckland on the effect of meteorological conditions on meningococcal disease identified associations with humidity, temperature and rainfall [49].

Other risk factors identified with meningococcal disease notifications in grid i were:

- the proportion aged less than 4 years ($\Delta\lambda_{it}$=1.142, p-value=0.017),
- the proportion who identify with Maori ethnicity ($\Delta\lambda_{it}$=1.185, p-value=0.000), and
- an increasing SDI score ($\Delta\lambda_{it}$=1.254, p-value=0.000).

The only protective factor identified with the rate of meningococcal disease notifications in grid i were the proportion who identify with Asian ethnicity ($\Delta\lambda_{it}$=0.924, p-value=0.044).

In past studies, infants, young children, and those of Maori and Pacific Island ethnicities have been identified to have an increased risk of meningococcal disease in New Zealand [31, 43]. Meningococcal disease outbreaks have been identified to be strongly associated with overcrowding [6, 5] and reported to occur in partly closed communities [43], which may lend explanation to the association identified with increasing SDI.

The beta coefficient estimates for the projection calculations for the average absolute humidity was $\tilde{\beta}_1$=28.72 (p-value = 0.257) and the weekly average temperature was $\tilde{\beta}_2$=-0.0260 (p-value = 0.166). The beta coefficient for the average absolute humidity is positive, indicating an increase in projected meningococcal disease is associated with an increase in projected absolute humidity. The beta coefficient estimate is negative for the average temperature, indicating a decrease in projected meningococcal disease is associated with an increase in projected temperature. The projected change in meningococcal disease IR from the study period to 2014, 2040 and 2090 indicate a decrease, where the magnitude of change was smaller in 2015 (from -0.97% - 0.48%)compared to 2090 (-5.96% - 0.69%). In 2015 and 2040, the summer projections indicate a decrease in meningococcal disease incidence, autumn suggests no change, and winter and spring indicate a decrease throughout New Zealand with no change in parts of Nelson-Marlborough, Canterbury and Otago DHBs. In 2090, a small decrease to no change is projected in summer and spring and no change projected in autumn and winter.

8.3 Issues Encountered

Throughout this section, limitations experienced with data uncertainty, the size and complexity of the dataset, statistical analysis, bias and confounding will be discussed.

8.3.1 Data Uncertainty

Weather Data

The weather data comprised of weekly measurements spanning 1997 - 2007, provided for 11,491 grids throughout New Zealand. The centroid of each grid contained a virtual climate station, derived from approximately 500 sites recording rainfall, 150 sites recording temperature and humidity, 120 sites recording wind speed and 100 sites recording solar radiation. The measurements were calculated using a second order derivative trivariate thin plate smoothing spline spatial interpolation model as described by Tait et al. (2006) [74]. The model was chosen as it had been shown to perform well in previous rainfall interpolation studies in New Zealand [74], however it must be noted that with any modelling, the interpolated weekly weather measurements will contain an uncertainty about their true value.

Disease Data

The weekly count of each indicator disease was derived by the summation of notifications from Wednesday through to Tuesday. Our aggregation to weekly data prevents the "day of week" effect, where notifications numbers drop off during the weekend. This can be due to the cost of emergency appointments after hours and during the weekends, or that the first signs of sickness of patients with appointments with their General Practitioners on Monday can occur anywhere between Friday night and Monday morning.

The true number of notifications for each indicator disease may be under-estimated due to under-reporting, although tends to be uncommon in New Zealand. Under-reporting can occur when patients do not seek medical help from a Medical Practitioner when symptoms occur. Other options for medical help can come in the form of pharmacies, herbal medicines and self diagnosis or treatment. Another source of under-reporting can be from the Medical Practitioner themselves not filling out the correct paperwork for notifications or requesting the correct diagnosis test. Cultural and deprivation differences can also affect the chances of notification. In New Zealand, under-ascertainment is more likely to be a limitation, which includes limitations of patients not presenting (mainly enteric diseases) to the limitation of clinical diagnosis and testing (mainly influenza).

Over-reporting, and hence the over-estimation of the true IR within New Zealand, may occur when a case is misclassified as the wrong disease or illness based on the clinical symptoms or exposures of the individual seeking medical attention. In this thesis, the inclusion of probable cases, where the necessary tests required for confirmation were not carried out, may introduce over-reporting.

The definitions for a confirmed and probable notified case of each indicator disease are listed below. The details and instructions for the notification for of a disease occurrence can be obtained from ESR at `http://www.surv.esr.cri.nz/episurv/crf.php`.

A confirmed case of campylobacteriosis or cryptosporidiosis requires a clinically compatible illness that is laboratory confirmed. A probable case is either a contact with a confirmed case of the same disease, or has had contact with the same common source i.e. part of an identified common source outbreak.

A confirmed case of meningococcal disease requires a clinically compatible illness with any one of the following:

- isolation of *Neisseria meningitidis* from blood, cerebrospinal fluid (CSF) or other normally sterile site, or
- detection of gram negative intracellular diplococci in blood, CSF or skin petechiae, or
- detection of meningococcal antigen in CSF, or
- a positive polymerase chain reaction (PCR).

A probable case is defined as one that is a clinically compatible illness.

The data for confirmed cases of influenza hospitalisations, where influenza is the primary diagnosis, are obtained from hospital records. The presence of the influenza virus is tested for via a nasopharyngeal swab or throat swab using PCR analysis [56], however only a small fraction of those hospitalised with influenza are laboratory confirmed in New Zealand.

The rate of under- and over-reporting is assumed to remain constant over the 11 year study period and therefore in the future projections. The true rate of influenza hospitalisations is thought to be higher and the medical tests carried out at the hospital are assumed to give the correct diagnosis. Under-reporting of meningococcal disease is also assumed to be captured with the inclusion of probable cases and therefore little to no effect of misclassification bias is expected. The rate of under-reporting is thought to be greater than the rate of over-reporting in campylobacteriosis and cryptosporidiosis. It is therefore considered that the true incidence of campylobacteriosis and cryptosporidiosis will be much higher (differential misclassification bias), as noted by Wheeler et al. (1999) in a study of campylobacteriosis in England [79].

Demographic Data

The demographic data was linearly interpolated from Census data collected in 1996, 2001 and 2006. The area units (AU) from Statistics New Zealand and VCS grids from NIWA were overlaid and the VSC grids assigned the corresponding value for each variable. The demographic data for the years 1997 - 2000, 2002 - 2005 and 2007 represent the best estimate of the true status with any misclassification bias caused is considered non-differential.

The proportion in each grid of the five groups of ethnicity: European or Other Ethnicity (including New Zealander); Maori; Pacific Peoples; Asian; and Middle Eastern, Latin American or African. The ethnic groups that people identify with can change over time, therefore caution must be exercised when interpreting results involving ethnicity as differential misclassification bias may be evident.

The District Health Boards (DHB) were established in New Zealand in 2001 and did not change from 2001 to 2006. Any effect in DHB reporting practices from 1997 - 2000 may be seen by chance. In 2007, the Otago and Southland DHBs were joined. This joining and restructuring was not taken into account during the screening of the DHB against the IR of each indicator disease. However, this joining will have no effect on the past association model or projection calculation as the variable was excluded due to its sheer number of levels (20 DHBs entered as factors, with Counties Manukau as the baseline).

The social deprivation index (SDI) is a measure of deprivation on an ordinal scale from 1 – 10, developed at the Department of Public Health, University of Otago, Wellington. The original data was provided to ESR at the meshblock level and deprivation scores were allocated to the grids used in this work using a weighted average of New Zealand Deprivation score values, as suggested in the NZDep2006 Manual [80]. The weighted average for each grid-week (it) spanning the 11 year study period was assigned a NZDep score from the closest census year. A SDI score of 0 was allocated to areas that are predominately oceanic, sea inlets or river estuaries and contain very few people in total. There appear to be little to no limitations or bias identified in using this variable.

Environmental Data

The animal density data was obtained at the farm-address level from AgriBase. The VCS grid corresponding to the farm centroid were allocated the total number of beef, diary and sheep animals present on that farm, and not the VCS grids that span the farm. This method will cause misclassification of the number of animals per km^2 as the farm polygon boundaries were not used. The effect of this method on misclassification bias is considered to be non-differential as the misclassification will occur randomly throughout New Zealand.

The drinking water quality for each VCS grid was recorded for 1,333 out of 11,491 VCS grids. The 10,158 VCS grids with missing drinking water quality scores were assigned a value of 3 representing "unknown drinking water quality". These grids could be speculated as areas with little to no PAR resulting in no need to monitor the drinking water quality, or grids where the residents have private drinking water sources such as roof supplies or rural ground supplies. Both speculations could be a source of misclassification bias, non-differential in the first instance and differential in the second.

The land use data used in this thesis was derived from the best source available and collected in summer 2001/2002. Throughout the 11 year study period, it was assumed that the land use in each VCS grid remained constant. Any misclassification bias caused by the land use data is thought to be non-differential.

Other Data

The original influenza vaccination data was obtained from three different sources, in three different formats. The calculation of the influenza vaccination coverage data in this thesis assumed that the number of individuals in each grid vaccinated each year was proportional to the district health board (DHB) rate of vaccination. This estimation of the number of vaccinations per 100,000 people is not an accurate representation of the true rate in each VCS grid and results must be treated as such.

Meningococcal disease vaccination (MeNZB) data from the New Zealand Ministry of Health was not obtained for the analysis carried out in this thesis, as the nation-wide coverage was not achieved until the end of our study period (2007) and data from 1997 - 2007 was found to be poor spatially.

Data on school holiday dates from 1997 - 2007 were obtained from the New Zealand Ministry of Education. These accurate dates were then matched to the weekly measures from 1997 - 2007 and therefore considered to introduce no forms of bias or limitations.

Climate Projection Data

Projection data will always incur uncertainties around the measures. The climate projection data used in this thesis was the best available at the time this work was carried out. However, the use of any climate projections for calculating the projected change in disease incidence will increase the level of uncertainty about the model output.

When using climate projections, the multiple number of models and emission scenarios available need to be considered. In future years, the accuracy of the projections will depend on the choice of emission scenarios used. The B1, A1B and A2 emission scenarios were selected by the HAIFA team as they represent a range of conservative to non-conservative estimates of population growth and behavioural changes to the effects of climate change.

The variable climate in New Zealand makes the climate projection estimates even more complicated, once again highlighting the uncertainty about model outputs.

8.3.2 Size and Complexity of Dataset

The data for each disease, weather and confounding variable (including the levels within) consisted of 11,491 VCS grids containing 574 weekly measurements per grid. The sheer size of this dataset, along with memory restrictions in the current versions of statistical software, created the room to learn new skills involved with data manipulation, model coding and the numerous operating systems available. The limitations experienced with computing power on a 32-bit operating system were caused primarily due to the RAM, and memory limitations within the statistical package R [62].

Screening

The quasi-Poisson models implemented in the screening of weather variables and possible confounders, required the use of a 64-bit linux machine, due to the memory restrictions in the statistical package R on a Microsoft 32-bit operating system.

Past Association Models

The requirement for the past association model to adequately model a large, complex dataset, resulted in a complicated, yet rewarding challenge. To further complicate the already present issue with data complexity, the spatial and temporal structure of each disease and explanatory covariate needed to be adjusted for. This ensured the Poisson distribution assumption of independent outcomes was upheld, i.e. the occurrence of a notified case of disease was independent.

To resolve the computing power issue, two disease models were moved to a 64-bit linux operating system with four CPU's (Xeon X5482 CPU with 16 Gb RAM and 4 Tb hard drive). The other two disease models were run on a 64-bit Windows XP operating system (i5-650 CPU with 8Gb RAM and 1 Tb hard drive). Both the 64-bit Linux and Windows XP operating systems allowed an increase in R memory allocation.

The length of time for the completion of each MCMC iteration increased linearly with the number of variables entered into the model. With the number of iterations required for convergence in the order of 10,000, this proved to be time-consuming and traditional methods for model building could not be used. The model building process was carried out in two steps. In the first step, the weather and other possible confounding variables were screened for possible associations with the IR of disease. The computing limitations of such a complex model resulted in a restriction of the maximum number of weather and possible confounding variables to be included in each model set to eight. The decision was made to include all significant weather explanatory variables. The second step involved ranking the confounding variables found to be significantly associated with the IR of disease during screening, using expert opinion as to the epidemiologically plausibility of an association.

8.3.3 Statistical Analysis

Issues encountered during the statistical analysis used for each modelling phase are outlined below.

Screening

A quasi-Poisson generalized linear model (GLM) was used to screen the association of each weather variable, having adjusted for the effects of seasonality (sine and cosine). Significant associations with the weather and seasonal variables identified were used to form the base model. Possible confounding variables were then screened for associations with the IR of disease, having adjusted for the effects of the base model.

The offset of the GLM was set to the PAR in order to account for differing population sizes in each VCS grid. This excluded grids with no PAR from the screening phase, as the offset term ($\log(PAR)$) is undefined when PAR=0.

The significance of association required to be considered for selection into the past association and projection calculation stages was set to a p-value ≤ 0.05. In epidemiology, a more common significance level during screening is p ≤ 0.20, however, as discussed earlier, the quasi-Poisson GLM does not adjust for spatial or temporal correlation over and above that contained in the covariates. This required us to be conservative in our interpretation of p-values during screening in the hope to reduce the chances of taking variables through to the past association modelling stages which are in fact only significant due to the large number of "repeated measures" (6,448,316 measures of "one location" derived from the 574 weekly measurements per grid with PAR > 0).

The quasi-Poisson GLM did not allow for the spatial and temporal dependencies of the data but did account for overdispersion.

A traditional model building process could not be carried out during screening, due to the size of the dataset and memory restrictions in R, within the Microsoft and Linux software. Screening for possible weather and confounding variable associations with the IR of disease was carried out in two phases. Phase One involved screening the weather and seasonal terms, resulting in the formation of a base model (B) for each disease. Phase Two involved screening possible confounders to check for significant associations with the incidence of disease, having adjusted for the base model.

Past Association Models

A modified Knorr-Held Richardson model was used to explore the effect of past climate variation on the four indicator diseases in New Zealand from 1997 - 2007, at a fine spatial and temporal level, having adjusted for the effect of confounders. Expert opinion on epidemiological plausibility, alongside screening results were used to select the confounding covariates for each model.

Traditional model building processes could not be carried out to identify the "best" model for the effect of climate variation on the IR of disease from 1997 - 2007. The weather and confounding variables identified to be possibly associated with the IR of disease during screening were then ranked using expert opinion on epidemiological plausibility. Once entered into the past association model, variables identified to be insignificantly attributing to the model were not removed, as with backwards selection model building. This method has identified drawbacks:

1. There was no room to explore the best model using different combinations of confounding variables.
2. Some of the significant variables during screening were no longer associated with the IR of disease once the effects of the other covariates and spatial-temporal correlations were accounted for.
3. The chances of identifying new, un-researched risk factors could not be explored.
4. Interactions could not be included.
5. Fitted lagged variables could also not be included, for example weather or disease variables at time $t-1$.

The effect of climate variation on the IR of disease used weather covariates from virtual climate stations that were estimated from real climate stations. For this reason, the posterior beta coefficients ($\hat{\beta}_k$) can only represent our "best guess" of the true past association of climate variation on the IR of disease from 1997 - 2007, at a fine spatial and temporal level.

At the recent InFER conference [36], it was discussed that a negative binomial model should always be used to account for overdispersion in Poisson distributed data. It can be argued that the spatial, temporal and unstructured random effects in the Knorr-Held Richardson model already adjust for overdispersion, with the unstructured component essentially providing an overdispersed Poisson model (the same way that a quasi-Poisson model does).

The modified Knorr-Held Richardson model is supposed to ensure the posterior spatial effects (U_i) are normally distributed with a mean of zero. If the mean is not zero, this can skew the distribution. The mean of the posterior spatial effects for the past association model of campylobacteriosis was -4.880 x 10^{-18}, cryptosporidiosis was -7.270 x 10^{-18}, influenza hospitalisations was 1.474 x 10^{-18} and meningococcal disease was -3.401 x 10^{-18}. We reasoned that the mean posterior spatial effect for each disease was extremely close to zero, and therefore the distribution would not be skewed.

Projection Calculations

The posterior beta coefficients ($\hat{\beta}_k$) for the covariates in the past association model were used to project the percentage change in IR for each indicator disease, under the 3 IPCC climate scenarios. Uncertainty around these posterior beta coefficients identified in the section above, are reinforced in the projection calculations. In an additional note, the projections based around the posterior beta coefficients identified from the past association of climate and confounding variables on the IR of disease throughout New Zealand from 1997 - 2007, did not take into account any industry level regulations or interventions, with are an universal feature of notifiable diseases. An example of this has already been seen with the 2007/2008 drop in campylobacteriosis notifications attributed to the introduction of a range of voluntary and regulatory interventions in poultry [67]. Any future interventions, such as new influenza or meningococcal disease vaccinations, could greatly reduce and possibly eliminate these diseases in the near future, making the effect of climate change factors minor in comparison.

It was assumed that the extended Knorr-Held Richardson model provided the "best estimate" for the true modelling of the past association of weather variables on the IR of each indicator disease, having adjusted for possible confounders. The inclusion of weather and confounding variables that were no longer found to be significantly associated with the IR of disease, having taken into account other variables, reinforces the uncertainty in the projection estimates. It therefore must be noted that the projections for the change in campylobacteriosis IR include one insignificantly associated weather variable and the projections for the change in influenza hospitalisations and meningococcal disease IR are based on two insignificantly associated weather variables. In the case of influenza hospitalisations and meningococcal disease, it is statistically plausible that there may be no change in the IR from the study period to the projection years 2015, 2040 and 2090.

Prediction intervals to describe the magnitude of uncertainty about these estimations were not calculated. The projection uncertainty from the past association model parameter estimates would be incredibly wide, and would not account for the uncertainty in the projected weather variables. Once again, it must be reinforced that the projections presented in this thesis represent our "best guess".

8.3.4 Bias

Bias could occur in this study when a systematic error results in an incorrect estimate of the true association between climate variation and IR of disease throughout New Zealand. Two classes of error identified in this study are selection bias and misclassification bias.

Selection bias could have occurred where disease notification were obtained from EpiSurv. This naturally selects a cohort of the New Zealand population who are willing to seek medical help from a General Practitioner or the likes.

Misclassification bias could occur in under- or over-reporting disease notifications to

EpiSurv. As outlined in Section 8.3.1, under-reporting results in an under-estimation of the true disease incidence, and over-reporting results in an over-estimation of the true disease incidence throughout New Zealand. The rate of under- and over-reporting in each disease was assumed to remain constant in the projections of the percentage change in IR throughout New Zealand.

Some of the confounding variables used were possibly incorrectly associated with the wrong VCS grids. This possibly occurred in the following variables:

- data linearly interpolated from census data (i.e. population at risk estimates, age and ethnicity),
- the farm allocation of animal numbers in the animal density data,
- the land use data assumed no change throughout the study period, using data collected in summer 2001/2002, and
- influenza vaccination coverage was obtained in three different formats and assumed the proportion in each VCS was proportional to the DHB vaccination rate.

8.3.5 Confounding

Confounding occurs when an unaccounted for variable is associated with both the outcome (disease counts) and explanatory (weather) variables of interest. It is very difficult to control for confounding at the population level, however commonly identified risk factors of each indicator disease were included throughout the modelling processes where possible.

Confounders we sought data for but were unavailable at the fine spatial level required for this project were the meningococcal disease vaccination coverage, from the 2004 MeNZB vaccination roll-out; air quality; and recreational water quality.

In a perfect model, all confounding is accounted for. It is possible that unmeasured confounders were not adjusted for in the models. Additionally, the limit on the number of confounding variables included in the past association models due to the inability to carry out traditional model building practices, makes this a likely source of error.

8.4 Recommendations for Further Research

8.4.1 Past Association Models

A traditional model building process (i.e. forward or backward variable selection) was not carried out during the past association modelling phase. The preliminary work carried out in this thesis could be extended to find the model which best describes the association between the climate variation and IR of disease in New Zealand from 1997 - 2007, having adjusted for confounders. It could then follow that projection calculations could be re-estimated with more certainty.

The model building process could then be extended to explore the possibilities of new risk factors associated with the IR of disease throughout New Zealand.

A subset of data could be used to estimate the goodness-of-fit of the past association models. Following this, the accuracy of the posterior beta coefficients in the past association model obtained from the 1997 - 2007 data could be estimated using collected weather and confounding data from 2008 - 2011. The calculated estimates from the modified Knorr-Held Richardson model could be compared to the actual number of notified cases on EpiSurv.

8.4.2 Projection Calculations

The projection calculations assumed the confounding variables remained constant throughout the future years (2015, 2040 and 2090). Further research could be carried out to estimate the projected change in IR throughout New Zealand using data on the projected change in confounders as well as weather variables. These projection estimates would give a more accurate picture of the expected change in IR of each indicator disease in New Zealand.

8.5 Conclusion

The work carried out in this thesis has demonstrated an association between past climate variation on the IR of campylobacteriosis and cryptosporidiosis in New Zealand from 1997 - 2007. Campylobacteriosis was found to be positively associated with the weekly absolute humidity. Additionally, campylobacteriosis notification risk factors were increasing beef and dairy density, intermediate and poor drinking water quality, being over 65 years of age and identifying with Pacific Island or Asian ethnicity. Protective factors were being less than 4 years of age and identifying with Maori ethnicity. Cryptosporidiosis was found to be positively associated with the weekly average temperature and negatively associated with the weekly average rainfall. Risk factors for cryptosporidiosis notifications were poor drinking water quality, being less than 4 years of age and living in rural areas. Protective factors were identifying with Maori ethnicity and unknown drinking water quality.

No association between past climate variation and the IR of influenza hospitalisations and meningococcal disease notifications in New Zealand from 1997 - 2007 was found in the work carried out in this thesis. Identifying with Maori ethnicity was found to be a risk factor for influenza hospitalisations, with no protective factors identified. Risk factors for meningococcal disease notifications were an increasing social deprivation index (SDI) score, being less than 4 years of age and identifying with Maori ethnicity. Identifying with Asian ethnicity was a protective factor for meningococcal disease.

Projections calculations estimating the change in disease incidence from the study period (1997 - 2007) to 2015, 2040 and 2090 were carried out under 3 IPCC climate scenarios (A2, B1 and A1B). The projected change in campylobacteriosis and cryptosporidiosis suggested increases in the rate of notifications, and a small decrease to no change in influenza hospitalisations and meningococcal disease notifications. It was noted that the projection calculations must be interpreted with caution, as the weather and confounding variables used in the coefficient estimates were not all statistically significant.

Bibliography

[1] Afifi, A. A., Kotlerman, J. B., Ettner, S. L. and Cowan, M. [2007]. Methods for Improving Regression Analysis for Skewed Continuous or Counted Responses, *Annual Review of Medical Health* .

[2] Ajjampur, S. S. R., Gladstone, B. P., Selvapandian, D., Muliyil, J. P., Ward, H. and Kang, G. [2007]. Molecular and Spatial Epidemiology of Cryptosporidiosis in Children in a Semiurban Community in South India, *Journal of Clinical Microbiology* pp. 915–920.

[3] Anderson, H., Spix, C., Medina, S., Schouten, J., Castellsague, J., Rossi, G., Zmirou, D., Touloumi, G., Wojtyniak, B., Ponka, A., Bacharova, L., Schwartz, J. and Katsouyanni, K. [1997]. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project, *European Respiratory Journal* **10**: 1064–1071.

[4] Bae, S., Famoye, F., Wulu, J., Bartolucci, A. and Singh, K. [2005]. A rich family of generalized Poisson regression models with applications, *Mathematics and Computers in Simulation* **69**: 4–11.

[5] Baker, M., Martin, D., Kieft, C. and Lennon, D. [2001]. A 10-year serogroup B meningococcal disease epidemic in New Zealand: Descriptive epidemiology, 1991-2000, *Journal of Paediatrics and Child Health* **37**(S13-S19).

[6] Baker, M., McNicholas, A., Garrett, N., Jones, N., Stewart, J., Koberstein, V. and Lennon, D. [2000]. Household crowding a major risk factor for epidemic meningococcal disease in Auckland children, *The Pediatric infectious disease journal* **19**(10).

[7] Baker, M., Sneyd, E. and Wilson, N. [2007]. Is the major increase in notified campylobacteriosis in New Zealand real?, *Epidemilogy and Infection* **135**: 163–170.

[8] Benschop, J. [2011]. Senior Lecturer at the EpiCentre, Massey University. Personal Communication.

[9] Benschop, J., Meade, A. and French, N. [2010]. Review of Statistical Modelling Approaches for the Impact of Climate on Infectious Diseases, *Technical report*, Population and Environmental Health Group unpublished report for the FRST HAIFA Project.

[10] Berk, R. and MacDonald, J. [2008]. Overdispersion and Poisson Regression, *Journal of Quantitative Criminology* **4**: 269–284.

[11] Besag, J. and Mollie, A. [1991]. Bayesian mapping of mortality rates, *Bulletin of the International Statistical Institute* **53**: 127–128.

[12] Best, N., Richardson, S. and Thomson, A. [2005]. A comparison of Bayesian spatial models for disease mapping, *Statistical Methods for Medical Research* **14**: 35–59.

[13] Bi, P., Cameron, A., Zhang, Y. and Parton, K. [2008]. Weather and notified *Campylobacter* infections in temperatre and sub-tropical regions of Australia: an ecological study, *Journal of Infection* **57**: 317–323.

[14] Bowman, A. and Azzalini, A. [1997]. *Applied Smoothing Techniques for Data Analysis*, Oxford Statistical Sciences Series, Oxford University Press, Oxford.

[15] Britton, E. [2008]. *The impact of climate variability and change on selected enteric diseases in New Zealand*, Master's thesis, University of Otago.

[16] Cazelles, B., Chavez, M., de Magny, G. C., Guegan, J.-F. and Hales, S. [2007]. Time-dependent spectral analysis of epidemiological time-series with wavelets, *Journal of the Royal Society Interface* **4**: 625–636.

[17] CBG [2006]. Evaluation of Meningococcal B Immunisation National Roll-Out: Prepared for the Ministry of Health, *Technical report*, CBG Health Research Limited.

[18] Chagneau, P., Mortier, F., Picard, N. and Bacro, J.-N. [2011]. A Hierarchical Bayesian Model for Spatial Prediction of Multivariate Non-Gaussian Random Fields, *Biometrics* **67**(1): 97–105.

[19] Chatfield, C. [2004]. *The Analysis of Time Series: An Introduction*, Sixth edn, Chapman and Hall, London.

[20] Cleveland, W. [1979]. Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**: 829–836.

[21] Congdon, P. [2001]. *Bayesian Statistical Modeling*, London: John Wiley and Sons Ltd.

[22] DeGooijer, J. G. and Hyndman, R. J. [2006]. 25 years of time series forecasting, *International Journal of Forecasting* **22**: 443–473.

[23] Dockery, D. W. and III, C. A. P. [1997]. Acute respiratory effects of particulate air pollution, *Annual Review of Public Health* **15**: 107–132.

[24] Duncanson, M., Russell, N., Weinstein, P., Baker, M., Skelly, C., Hearnden, M. and Woodward, A. [2000]. Rates of Notified Cryptosporidiosis and Quality of Drinking Water Supplies in Aotearoa, New Zealand, *Water Research* **34**(15): 3804–3812.

[25] Ebi, K. [2009]. Managing the changing health risks of climate change, *Current Opinion in Environmental Sustainability* **1**: 107–110.

[26] Fisman, D. N. [2007]. Seasonality of Infectious Diseases, *Annual Review of Public Health* **28**: 127–143.

[27] French, N. [2011]. Professor of Food Safety and Veterinary Public Health, Massey University. Personal Communication.

[28] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. [2004]. *Bayesian Data Analysis*, Chapman and Hall, CRC.

[29] Gilpin, B. J., Scholes, P., Robson, B. and Savill, M. G. [2008]. The Transmission of Thermotolerant *Campylobacter* spp. to People Living or Working on Dairy Farms in New Zealand, *Zoonoses Public Health* **55**: 352–360.

[30] Gilpin, B., Thorrold, B., Scholes, P., Longhurst, R., Devane, M., Nicol, C., Walker, S., Robson, B. and Savill, M. [2008]. Comparison of *campylobacter jejuni* genotypes from dairy cattle and human sources from the Matamata-Piako District of New Zealand, *Journal of Applied Microbiology* **105**: 1354–1360.

[31] Harrison, L. H., Trotter, C. L. and Ramsay, M. E. [2009]. Global epidemiology of meningococcal disease, *Vaccine* .

[32] Hearnden, M., Skelly, C., Eyles, R. and Weinstein, P. [2003]. The regionality of campylobacteriosis seasonality in New Zealand, *International Journal of Environmental Health Research* **13**(4): 337–348.

[33] Held, L., Natario, I., Fenton, S. E., Rue, H. and Becker, N. [2005]. Towards joint disease mapping, *Statistical Methods for Medical Research* **14**: 61–82.

[34] Hu, W., Tong, S., Mengersen, K. and Connell, D. [2007]. Weather variability and the incidence of cryptosporidiosis: Comparison of time series Poisson regression and SARIMA models, *Animals of Epidemiology* **17**(9): 679–688.

[35] Hussain, I., Pilz, J. and Spoeck, G. [2010]. Hierarchical Bayesian space-time interpolation versus spatio-temporal BME approach, *Advances in Geosciences* **25**: 97–102.

[36] *InFER Conference* [2011]. Warwick Unverisity, United Kingdom.

[37] Intergovernmental Panel on Climate Change [2000]. IPCC Special Report, Emissions Scenarios, *Technical report*, Intergovernmental Panel on Climate Change.

[38] Ismail, N. and Jemain, A. A. [2007]. Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models, *Casualty Actuarial Society Forum* .

[39] Jeffery, C., Ozonoff, A. and Pagano, M. [n.d.]. Multivariate Disease Mapping, Harvard and Boston Schools of Public Health.

[40] Kardos, J. [2005]. *Visualising attribute and spatial uncertainty in choropleth maps using hierarchical spatial data models*, PhD thesis, University of Otago.

[41] Keene, O. N. [1995]. The Log Transformation is Special, *Statistics in Medicine* **14**: 811–819.

[42] Kelly, C., Arnold, R., Galloway, Y. and O'Hallohan, J. [2007]. A Prospective Study of the Effectiveness of the New Zealand Meningococcal B Vaccine, *American Journal of Epidemiology* **166**(7).

[43] Khatami, A. and Pollard, A. J. [2010]. The epidemiology of meningococcal disease and the impact of vaccines, *Expert Review of Vaccines* **9.3**.

[44] Knorr-Held, L. [1999]. Conditional prior proposals in dynamic models, *Scandinavian Journal of Statistics* **26**(1): 129–144.

[45] Knorr-Held, L. and Richardson, S. [2003]. A hierarchical model for space-time surveillance data on meningococcal disease incidence, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **52**(2): 169–183.

[46] Lake, I., Pearce, J. and Savill, M. [2007]. The seasonality of human cryptosporidiosis in New Zealand, *Epidemiology and Infection* .

[47] Lawson, A. B. and Williams, F. L. R. [2001]. *An Introductory Guide to Disease Mapping*, John Wiley and Sons, Inc., New York.

[48] Learmonth, J., Ionas, G., Ebbett, K. and Kwan, E. [2004]. Genetic characterization and transmission cycles of *Cryptosporidium* species isolated from humans in New Zealand, *Applied and Environmental Microbiology* **70**(7): 3973–3978.

[49] Lindsay, A. P., Hope, V., Marshall, R. J. and Salinger, J. [2002]. Meningococcal disease and meteorological conditions in Auckland, New Zealand, *Australian and New Zealand Journal of Public Health* **26**: 212–218.

[50] Lindsey, J. K. [1999]. A Reivew of Some Extensions to Generalized Linear Models, *Statistics in Medicine* **18**: 2223–2236.

[51] Lofgren, E., Fefferman, N., Naumov, Y., Gorski, J. and Naumova, E. [2007]. Influenza seasonality: underlying causes and modeling theories, *Journal of Virology* pp. 5429–5436.

[52] Marshall, J., Spencer, S. and French, N. [2009]. Final Report: SCIG-MAS-001. 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. Prepared for New Zealand Food Safety Authority, *Technical report*, Hopkirk Institute, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand.

[53] McMichael, A. [2009]. Human population health: sentinel criterion of environmental sustainability, *Current Opinion in Environmental Sustainability* **1**: 101–106.

[54] McMichael, A. J. [2003]. *Climate Change and Human Health - Risks and Responses Summary*, First edn, World Health Oragnisation.

[55] McMillan, N., Bortnick, S. M., Irwin, M. E. and Berliner, L. M. [2005]. A hierarchical Bayesian model to estimate and forecast ozone through space and time, *Atmospheric Environment* **39**: 1373–1382.

[56] Ministry of Health [2006]. National Laboratory Guidelines for Pandemic Influenza: Collection and handling of human specimens for laboratory diagnosis of influenza with pandemic potential, *Technical report*, Wellington: Ministry of Health.

[57] Ng, J., Eastwood, K., Durrheim, D., Massey, P., Walker, B., Armson, A. and Ryan, U. [2008]. Evidence supporting zoonotic transmission of *cryptosporidium* in rural new south wales, *Experimental Parasitology* **119**: 192–195.

[58] Pattrick, M., Christiansen, L., Waino, M., Ethelberg, S., Madsen, H. and Wegener, C. [2004]. Effects of climate on incidence of *Campylobacter* spp. in humans and prevalence in broiler flocks in Denmark, *Applied and Environmental Microbiology* **70**(12): 7474–7480.

[59] Population and Environmental Health Group [2009]. Notifiable and other disease in New Zealand. Annual Report 2008, *Technical report*, Institute of Environmental Science and Research Limited.

[60] Population and Environmental Health Group [2011]. Notifiable and other disease in New Zealand. Annual Report 2010, *Technical report*, Institute of Environmental Science and Research Limited.

[61] R Development Core Team [2009]. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. **URL:** *http://www.R-project.org*

[62] R Development Core Team [2010]. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. **URL:** *http://www.R-project.org/*

[63] Ranta, J., Makaela, P., Takala, A. and Arjas, E. [1999]. Predicting the course of meningococcal disease outbreaks in closed subpopulations, *Epidemilogy and Infection* **123**: 359–371.

[64] Rind, E. and Pearce, J. [2010]. The spatial distribution of campylobacteriosis in New Zealand, 1997 - 2005, *Epidemilogy and Infection* **138**: 1359–1371.

[65] Schmidt, A. M. and Pereira, J. B. M. [2011]. Modelling Time Series of Counts in Epidemiology, *International Statistical Review* **79**(1): 48–69.

[66] Schwartz, J., Spix, C., Touloumi, G., Bacharova, L., Barumandzadeh, T., le Tertre, A., Piekarksi, T., de Leon, A. P., Ponka, A., Rossi, G., Saez, M. and Schouten, J. P. [1996]. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions, *Journal of Epidemiology and Community Health* **50**: S3–S11.

[67] Sears, A., Baker, M. G., Wilson, N., Marshall, J., Muellner, P., Campbell, D. M., Lake, R. J. and French, N. P. [2011]. Marked Campylobacteriosis Decline after Interventions Aimed at Poultry, New Zealand, *Emerging Infectious Diseases* **17**(6).

[68] Semanza, J. and Menne, B. [2009]. Climate change and infectious diseases in Europe, *The Lancet* **9**: 365–375.

[69] Shumway, R. H. and Stoffer, D. S. [2011]. *Time Series Analysis and Its Applications, With R Examples*, third edn, Springer Science+Business Media, LLC 2011.

[70] Simmons, G., Martin, D., Stewart, J., Jones, N., Calder, L. and Bremner, D. [2001]. Carriage of *neisseria meningitidis* among household contacts of patients with meningococcal disease in new zealand, *European Journal of Clinical Microbiology & Infectious Diseases* **20**: 237–242.

[71] Snel, S., Baker, M., Kamalesh, V., French, N. and Learmonth, J. [2009]. A tale of two parasites: the comparative epidemiology of cryptosporidiosis and giardiasis, *Epidemiology and Infection* .

[72] Snel, S., Baker, M. and Venugopal, K. [2009]. The epidemiology of cyrptosporidiosis in New Zealand, 1997-2006, *The New Zealand Medical Journal* **122**(1290): 47–60.

[73] Spencer, S. E., Marshall, J., Pirie, R., Campbell, D. and French, N. P. [2011]. The detection of spatially localised outbreaks in campylobacteriosis notification data, *Spatial and Spatio-temporal Epidemiology* **2**(3): 173–183.

[74] Tait, T., Henderson, R., Turner, R. and Zheng, X. [2006]. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface, *International Journal of Climatology* **26**: 2097–2115.

[75] Tompkins, D. [2010]. Climate Change and Human Health in New Zealand: Modelling the Future Impact of Vector-Bourne Diseases., *Technical report*, Landcare Research unpublished report for the FRST HAIFA Project.

[76] *Understanding and Working with Ethnicity Data* [2005]. *Technical report*, Statistics New Zealand.

[77] Verrall, A., Norton, K., Rooker, S., Dee, S., Olsen, L., Tan, C. E., Paull, S., Allen, R. and Blackmore, T. K. [2010]. Hospitalisations for Pandemic (H1N1) 2009 among Maori and Pacific Islanders, New Zealand, *Emerging Infectious Diseases* **16**(1).

[78] Wallace, J., Slanley, K. N., Currie, J. E., Diggle, P. J. and Jones, K. [1997]. Seasonality of thermophillic *campylobacter* populations in chickens, *The Society for Applied Microbiology* **82**: 219–224.

[79] Wheeler, J. G., Sethi, D., Cowden, J. M., Wall, P. G., Rodrigues, L. C., Tompkins, D. S., Hudson, M. J. and Roderick, P. J. [1999]. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance, *British Medical Journal* **318**: 1046–1050.

[80] White, P., Gunston, J., Salmond, C., Atkinson, J. and Crampton, P. [2006]. *Atlas of Socioeconomic Deprivation in New Zealand NZDep2006*, Ministry of Health.
URL: *http://www.moh.govt.nz*

[81] Xiao, L. and Feng, Y. [2008]. Zoonotic cryptosporidiosis, *FEMS Immunology and Medical Microbiology* **52**(3): 309–323.

[82] Yang, L., Wong, C. M., Chan, K. P., Chau, P. Y. K., Ou, C. Q., Chan, K. H. and Peiris, J. M. [2009]. Seasonal effects of influenza on mortality in a subtropical city, *Bio Medical Infectious Diseases Journal* **9**.

[83] Zeileis, A., Kleiber, C. and Jackman, S. [2008]. Regression models for count data in r, *Journal of Statistical Software* **27**(8).

[84] Zhang, Y., Bi, P. and Hiller, J. [2008]. Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models, *International Journal of Biometeorology* **52**: 179–187.