Durham University

Durham E-Theses

A robust Bayesian land use model for crop rotations

PATON, LEWIS,WILLIAM

How to cite:

PATON, LEWIS,WILLIAM (2016) *A robust Bayesian land use model for crop rotations*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/11484/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

A robust Bayesian land use model for crop rotations

Lewis Paton

A Thesis presented for the degree of
Doctor of Philosophy

Statistics and Probability
Department of Mathematical Sciences
University of Durham
England

June 2015

Dedicated to

My family

A robust Bayesian land use model for crop rotations

Lewis Paton

Submitted for the degree of Doctor of Philosophy
June 2015

Abstract

Often, in dynamical systems, such as farmers' crop choices, the dynamics are driven by external non-stationary factors, such as rainfall and agricultural input and output prices. Such dynamics can be modelled by a non-stationary stochastic process, where the transition probabilities are functions of such external factors. We propose using a multinomial logit model for these transition probabilities, and investigate the problem of estimating the parameters of this model from data.

We adapt the work of Chen and Ibrahim to propose a conjugate prior distribution for the parameters of the multinomial logit model. Inspired by the imprecise Dirichlet model, we will perform a robust Bayesian analysis by proposing a fairly broad class of prior distributions, in order to accommodate scarcity of data and lack of strong prior expert opinion.

We discuss the computation of bounds for the posterior transition probabilities, using a variety of calculation methods. These sets of posterior transition probabilities mean that our land use model consists of a non-stationary imprecise stochastic process. We discuss computation of future events in this process.

Finally, we use our novel land use model to investigate real-world data. We investigate the impact of external variables on the posterior transition probabilities, and investigate a scenario for future crop growth. We also use our model to solve a hypothetical yet realistic policy problem.

Declaration

The work in this thesis is based on research carried out at the Statistics and Probability Group, the Department of Mathematical Sciences, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Copyright © 2015 by Lewis Paton.
"The copyright of this thesis rests with the author. No quotations from it should be published without the author's prior written consent and information derived from it should be acknowledged".

Acknowledgements

I am indebted to my supervisor Matthias Troffaes for all of his guidance, assistance and extreme patience throughout the course of my PhD. I would like to thank Nigel Boatman, Mohamud Hussein and Andy Hart for all of their input, and for the many interesting discussions which have helped shape this work. I would also like to thank all those who have helped me, both at Durham and at Fera, in particular Paul Brown for his assistance in producing some of the figures which appear in this thesis.

I would like to acknowledge the financial support I received for this work, from both the Engineering and Physical Sciences Research Council and the Department for Environment, Food and Rural Affairs through the Fera seedcorn fund.

Finally, thank you to my friends and family for all the encouragement they have given me. In particular, a special thanks to my parents, for their constant support and motivation.

Contents

Abstract **iii**

Declaration **iv**

Acknowledgements **v**

1 Introduction **1**

2 An overview of land use modelling **5**

2.1 Drivers of crop choice 5

2.2 The importance of land use modelling 8

2.3 Difficulties 9

2.4 Comparison with other land use models 11

2.5 Chapter summary 13

3 The model **15**

3.1 Stochastic processes 15

3.2 A multinomial logit model 20

3.3 A conjugate prior for the multinomial logit model 23

3.4 Chapter summary 29

4 Robust Bayesian analysis **31**

4.1 Why perform a robust Bayesian analysis? 32

4.2 Sets of conjugate distributions 34

4.2.1 Choice of s 35

4.2.2 Choice of $\mathfrak{T}$ 38

4.2.3 Choice of $\mathfrak{X}$ 41
4.3 Posterior inference 43
4.3.1 Markov chain Monte Carlo 43
4.3.2 Maximum a posteriori estimation 45
4.3.3 Other possible approaches 53
4.4 An imprecise stochastic process 54
4.5 Chapter summary 56

5 Analysis of the National Character Area data 58
5.1 Initial analysis 58
5.1.1 Crop data 59
5.1.2 Soil data 61
5.1.3 Economic data 65
5.1.4 Rainfall data 69
5.2 Results 73
5.2.1 A simple model 73
5.2.2 Break crop model 78
5.2.3 Including repeated wheat 83
5.3 Model validation 88
5.3.1 A non-parametric estimate 89
5.3.2 Cross validation 96
5.3.3 Comparison with the imprecise Dirichlet model 97
5.4 Chapter summary 99

6 Future crop distributions 102
6.1 Comparison with observed crop distributions 102
6.2 A hypothetical example 110
6.3 Future data sets 112
6.4 Chapter summary 113

7 Potential relevance to policy making 115
7.1 Utility functions 115

7.2 Decision making under uncertainty 117
7.3 Application to increasing legumes 119
7.4 Chapter summary 126

8 Concluding discussion 127

8.1 Summary of the thesis 127
8.2 Potential areas of future work 130

List of Figures

3.1 State transition diagrams for two Markov chains, representing two different specifications of the state space $\mathcal{Y}$ 18

4.1 A comparison of different s values 36

4.2 A comparison of posterior transition probabilities in a simple binary model under two different sets of priors. 39

4.3 Inferences under two different $\mathfrak{X}$ 42

4.4 A comparison of the posterior transition probability bounds generated by the MAP estimate and an MCMC simulation, and the range of all 90% credible intervals generated by the MCMC simulations 49

4.5 A selection of autocorrelation plots from the MCMC simulation. . . . 51

4.6 Trace plot of the simulation shown in plot **E** of figure 4.5 51

5.1 The location within England of the East Anglian Chalk NCA and the Mease/Sence NCA 60

5.2 Soil distributions for our two NCAs of interest. 64

5.3 Historic crop distributions for the main crops in our NCAs, split by soil type. 66

5.4 Historic profit margins for the main crops of interest 67

5.5 Historic profit margins of beans and peas, and the assumed profit margin of legumes 68

5.6 Historic profit margins of wheat and repeated wheat 69

5.7 The historic mean yearly indexed nitrogen price 70

5.8 The observed rainfall locations for the East Anglian NCA. Each black circle represents one weather station, each of which provides measurements of monthly rainfall. 71
5.9 The bounded rainfall prediction for September 1994 in the East Anglian NCA, based on kriging. 72
5.10 The state transition diagram for the simple model 74
5.11 Transition probabilities from wheat in Anglia on all soil types, as monthly rainfall varies, for the simple model 75
5.12 Transition probabilities from barley in Anglia on all soil types, as monthly rainfall varies, for the simple model 76
5.13 Transition probabilities from other in Anglia on all soil types, as monthly rainfall varies, for the simple model 77
5.14 The state transition diagram for the break crop model 79
5.15 Transition probabilities from wheat for the break crop model with rainfall and profit margin difference as the regressors in Anglia 81
5.16 Transition probabilities from wheat for the break crop model with rainfall and profit margin difference as the regressors in Mease 82
5.17 Transition probabilities from wheat for the break crop model with rainfall and nitrogen price as the regressors in Anglia 84
5.18 The state transition diagram for the break crop model when repeated wheat is included 85
5.19 Transition probabilities from wheat for the break crop model, with repeated wheat, with rainfall and nitrogen price as the regressors in Anglia 86
5.20 Transition probabilities from repeated wheat for the break crop model, with repeated wheat, with rainfall and nitrogen price as the regressors in Anglia 87
5.21 Parametric and non-parametric estimates of the transition probabilities for all soil types in Anglia, from rapeseed and wheat, as nitrogen price varies, for rainfall in September fixed at the mean 90

5.22 Parametric and non-parametric estimates of the transition probabilities for all soil types in Anglia, from legumes and other, as nitrogen price varies, for rainfall in September fixed at the mean 91
5.23 Parametric, maximum likelihood and non-parametric estimates of the transition probabilities for all soil types in Anglia, from rapeseed and wheat, as nitrogen price varies, for rainfall in September fixed at the mean 93
5.24 Parametric, maximum likelihood and non-parametric estimates of the transition probabilities for all soil types in Anglia, from legumes and other, as nitrogen price varies, for rainfall in September fixed at the mean 94
5.25 Smoothed $n_i(x)$ as a function of x 95

6.1 A prediction of crop distributions in 2005–2009, in Anglia on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–2004 103
6.2 A prediction of crop distributions in 2000–2004, in Anglia on light soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999 105
6.3 A prediction of crop distributions in 2000–2004, in Anglia on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999 106
6.4 A prediction of crop distributions in 2000–2004, in Anglia on heavy soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999 107
6.5 A prediction of crop distributions in 2000–2004, in Mease/Sence on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999 . 108
6.6 A prediction of crop distributions in 2000–2004, in Mease/Sence on heavy soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999 . 109

6.7 A prediction of future crop distributions in 2005–2009 in Anglia on heavy soil, given values of rainfall and nitrogen price for that period. 111

7.1 Utility functions and decision criteria for a model which contains legume subsidy directly 121

7.2 Utility functions and decision criteria for a model which includes nitrogen price, on light soil in Anglia 122

7.3 Utility functions and decision criteria for a model which includes nitrogen price, on medium soil in Anglia 123

7.4 Utility functions and decision criteria for a model which includes nitrogen price, on heavy soil in Anglia 124

List of Tables

3.1 An extract of crop rotation data 24

5.1 Main crop code classification for the NCAs 61

5.2 Soil classification for the NCAs 63

5.3 Cross validation results for East Anglia, Mease/Sence, and the combined data set 97

5.4 Cross validation results for East Anglia, Mease/Sence, and the combined data set, split by previous crop grown 98

Chapter 1

Introduction

This thesis is concerned with modelling and predicting agricultural land use. That is, building a mathematical model that can predict which crop is grown in a particular field at a particular time. This is a research question of great importance. Both the global climate [50] and the global economy [59, 108] are changing, and being able to predict how future crop distributions may react to these changes is important to ensure food security [1]. Furthermore, being able to model what drives a farmer's crop choice can be useful to both agricultural scientists and policy makers in governments.

Many land use models have been developed over the last 40 years, with a wide range of inputs and outputs depending on the purpose of the model. However, there is generally little acknowledgement of the uncertainty in the modelling process. Throughout this thesis we will develop a novel approach to land use modelling that is flexible with regards to the inputs, and that captures the uncertainty of the predictions. A basic version of this model was first presented in [93], while a more general version was published in [77].

It is well known that farmers often grow crops in rotations. This means that there is a dependence from year to year on which crop is planted in a field (or a collection of fields). A simple model that captures some of this behaviour is a Markov chain, where the state at any given time is the crop choice at that time. There are existing land use models – Castellazzi et al. [22] and Luo [65] – which favour this approach, and this will be the starting point for our model.

However, these rotations are not entirely deterministic. Crop choices are not only influenced by what has been planted previously, but also by a wide variety of economic and environmental factors. To model the effect of these variables, Luo [65] suggested modelling the transition probabilities of the Markov chain as linear functions of said variables. We will follow a similar approach, but we will propose using a generalised linear model – specifically a multinomial logit model – for the transition probabilities. Multinomial logit models generalise the classic binary logistic regression model to allow for more than two categorical outcomes. Thus, it is an obvious choice to consider for modelling the transition probabilities.

A key challenge of any regression model is to estimate its parameters. This thesis is written from a Bayesian viewpoint, which means that we, or perhaps a land use expert, can specify prior beliefs about these parameters in the form of a probability distribution. This distribution is then updated with observed data and posterior inferences which take into account both evidence and prior opinion are made.

If the prior distribution and the posterior distribution are from the same family then we call these distributions conjugate. The property of conjugacy makes the updating process in a Bayesian analysis simple. Chen and Ibrahim [24], building on the pioneering work of Diaconis and Ylvisaker [32], proposed a general form of a conjugate prior for generalised linear models. Their work assumes the locations of the data are known a priori. We will adapt the work of Chen and Ibrahim to remove this dependence, specifically for the multinomial logit model. This work was first presented in [77].

A robust Bayesian approach [8] argues that specifying a single prior distribution is in most cases impossible, and that instead of trying to specify one prior distribution, we should include all that are reasonable. Therefore we specify a set of prior distributions, which generates a set of posterior inferences. In this setting, we will argue that this is a sensible approach to take. We will follow a similar approach to Walley's imprecise Dirichlet model [100], by identifying a reasonably vacuous set of conjugate prior distributions, and calculating posterior bounds. Thus, within our land use model, we will also present a novel approach to robust Bayesian multinomial logistic regression, based on conjugate analysis. Efficiency of calculation is an

important consideration in a robust Bayesian analysis due to the number of distributions we are considering, and we will use the computationally efficient MAP estimate of our posterior parameters due to this.

A robust Bayesian analysis gives a set of posterior inferences, which turns the stochastic process which is the foundation of our model into an imprecise stochastic process. This process is influenced by non-stationary random variables, and we will adapt the methodology from imprecise Markov chain theory [30] to describe how we make forward predictions in this process. This work presents a first step at including imprecision in such a stochastic process.

Underpinning this entire thesis are the data sets that we have access to. We will use our proposed model on this data to investigate various properties of the farming system. Within this we will perform various model validation procedures, including a novel non-parametric estimate of the transition probabilities, using a Gaussian kernel. This work was first presented in [78].

Land use models are used to help agricultural scientists assess the impact of changes in future economic and climatic variables, and such models can also be used to aid policy makers in analysing the potential impact that changes in policy may have. We will use our model to look at both of these scenarios. As we have sets of inferences, we will use the theory of imprecise decision making [91] alongside our land use model, in order to formulate and answer a hypothetical yet realistic policy question. This represents a novel application of land use models. Within this analysis, we will present a novel approach to decision analysis, based on MAP estimation, which was first presented in [78].

This chapter has introduced the areas of work that are used and developed within this thesis. The rest of the thesis is structured as follows. In chapter 2 we will further expand on what is important within land use modelling. We will discuss the factors that influence a farmer's crop choice; why land use modelling is an important procedure; but also why it can be difficult. Some existing land use models will be discussed, and areas where we hope to improve on them will be introduced.

After chapter 2, the work can be broadly separated into two sections. Chapters 3

and 4 detail the mathematical theory behind the model, while chapters 5, 6 and 7 explore uses of the model on real data.

Chapter 3 will begin laying the theoretical ground work for our land use model. We will introduce stochastic processes and the multinomial logit model, and then explore how we can use these in conjunction to model crop rotations. We will then propose a conjugate prior distribution for the parameter of the multinomial logit model, and investigate some properties of this distribution.

Chapter 4 will focus on the robust Bayesian side of the model, and how we perform inferences in this framework. We will argue why it is a sensible approach in this setting, and explain how we actually go about performing it for this model. We will introduce the methodology needed to make forward inferences in our imprecise stochastic process. We will also discuss two approaches to calculation: an MCMC simulation, and using the MAP estimate. We will discuss their strengths and weaknesses with respect to our model, and also briefly mention other possible approaches to calculation.

In chapters 3 and 4 we will occasionally mention our data in the form of examples, but in chapter 5 we will introduce our data sets fully, discussing the structure of the crop data, and also the other data that we will use in the model. With spatial data, such as rainfall, we sometimes must interpolate, and we will briefly discuss one such method known as kriging. We will present some analysis of how the model transition probabilities are influenced by external variables, and discuss some approaches to model validation.

Chapters 6 and 7 look at applications of the model. Chapter 6 will focus on the use of the model in order to make forward predictions in time. We will discuss the types of data we need to do this, and we will analyse a hypothetical scenario with real-world meaning. Chapter 7 introduces some basic imprecise decision theory, and discusses how it can be used in conjunction with our land use model in a way that could prove useful to a policy maker. Using this theory, we will investigate a specific policy question.

Finally, in chapter 8 we will conclude the thesis by summarising our findings and by discussing potential areas for future research.

Chapter 2

An overview of land use modelling

We are interested in building a statistical model which we can use to predict agricultural land use, with a specific focus on the crop rotations which are commonly used by farmers. In order to do this, we must first understand what actually drives the crop choice that a farmer makes every year. These drivers include economic, climatic and geographical factors. Then, as with any statistical modelling procedure, we must have some concept of what others have done in the field, how we can improve upon it, and the challenges we may face along the way. This chapter introduces these ideas.

In section 2.1, we will discuss some of the factors which influence a farmer's crop choice. Section 2.2 will give some overview of why modelling this system is an important research area, while section 2.3 will describe some of the difficulties in modelling this process. Section 2.4 will briefly look at some other work in this domain, and will discuss how we aim to improve on them in this work.

2.1 Drivers of crop choice

Every year, a farmer makes a decision as to which crop to grow in each field he farms on. This decision is influenced by many factors; some are specific to the farmer, while some are regional, national or even international.

It is well known that arable farmers generally grow crops in rotation [3]. A crop rotation is the production of different plants in succession in a particular field or

group of fields. Crop rotations have been an important part of arable farming for a very long period of time [19] – evidence of crop rotations date back to at least the Middle Ages, where a two-field rotation was used. One field was planted with a crop, while the other was left fallow. The following year the opposite occurred. Over the following centuries, three and four field rotations were developed in Europe. In modern times, a typical rotation might be successive plantings of wheat, followed by a different crop such as peas [37].

Crop rotations are an important agricultural practice for a variety of reasons [3]. Planting the same crop year after year can reduce the amount and quality of the crop produced, as nutrients are depleted from the soil. Since different crops have different nutrient requirements, a rotation that involves a variety of crops can help the soil maintain a good balance. This is less important in modern farming systems as fertilisers can supply the required nutrients, although organic systems still place an importance on balancing nutrient requirements from the soil. Furthermore recently, the sustainability of systems relying on fertilisers is being questioned, due to the negative impact it can have on the structure of the soil [84]. Thus, in future, the ability of some crops (for example, legumes and grass) to increase soil fertility may be valued once again.

A more important reason for rotating crops in modern farming is that it helps prevent the build up of weeds, pests and diseases, which can occur if similar crops are planted in succession [3]. A rotation that includes a diverse choice of crops can help stop the build up of pests and diseases. It can also improve weed control options, since different weeding tools can be used on different crop types, which also helps prevent soil erosion. Planning a multi-year rotation also enables farmers to use land, labour and machinery more efficiently [37].

Since farming is primarily a business [23], we can assume that farmers are interested in maximising their yields and profit margins over the course of a rotation. The profit margin M_i (pounds per hectare) for crop i, at a particular time, can be defined as [65]:

$$M_i = P_i L_i + G_i - V_i \tag{2.1}$$

where:

- P_i is the selling price (pounds per tonne) of crop i.
- L_i is the yield (tonnes per hectare) of crop i, which is dependant on factors such as soil type, the skill level of the farmer, and the tolerance of the crop to the local climate.
- G_i are the grants or subsidies payable (pounds per hectare) for crop i. These are payments from the European Union or a government for growing crop i.
- V_i are the variable costs (pounds per hectare) for crop i. This includes costs such as fertiliser, seed purchases, and so on.

So, while profit margin may be the main factor the farmer considers, it is clearly broken down into four components, all of which can change. This can have an impact on the profit margin, and therefore the decision a farmer makes. In Europe, the European Union offers subsidies to farmers through the Common Agricultural Policy [35]. Currently, this subsidy is the same for all crops, but this was not the case in the past (including the time period we will study later on in the thesis). This could however change again in the future if there was a desire to encourage certain crops to grow. For example, a previously unappealing crop to grow may become favourable with a certain level of subsidy, although there would have to be environmental reasons for this change to avoid contravening international treaties aimed at preventing unfair competition.

Variable costs also impact the crop choice. Fertiliser price is an important component of the margin, and so crops which require large amounts of fertiliser may become less appealing in times of high fertiliser prices. As nitrogen is a key ingredient in many fertilisers [34], these prices can be driven by factors such as global nitrogen price. Situations related to subsidy levels and nitrogen prices will be explored in chapters 6 and 7.

The local climate, and how it interacts with soil type, plays a large role in the farmer's decision too [3]. Soil texture determines how free draining the soil is and how much water it holds, and therefore how easy it is for the roots to penetrate. If there is a large amount of rain in the months before sowing then the field may be sodden, and thus it may be difficult to cultivate and drill the field. As such, some

crop types may not be suitable in certain climatic situations on certain soil types. Furthermore, if it is very wet in the autumn (the traditional drilling period for many crops), a farmer may choose to wait until spring to plant crops, as it is likely to be drier. Spring crops are harvested slightly later than winter crops, but they still have a reduced growing time which will have an impact on the yield.

There are also many other factors which influence the decision, some of which are harder to quantify. For example, the farm type and risk attitude of the farmer influence the decision. A crop with a large profit margin may be more likely to fail and so return very few crops suitable for sale. As such, small-scale farmers or farmers with low risk tolerance may reject growing these crops in favour of safer crops. A larger farm with many fields can perhaps spread the risk by growing many varieties of crops over the entire farm. Similarly, the ownership of the farm can have an impact on crop choices. Farmers who rent their farm may have a more short-term need to make a profit than those who own their farm. There may also be a consideration of local competition, especially if the farmer is interested in selling produce locally.

2.2 The importance of land use modelling

Being able to model and predict crop rotations is an important research area. This is because future crop rotations may change, due to changes in those variables that influence a farmers choice. Firstly, future crop rotations may be influenced by climate change, either by affecting the biology of the crops themselves or the pests and diseases that attack them. If changes in climatic conditions favour one crop over another, this may lead to changes in profitability of different crops, and hence the balance of the rotation [1].

Secondly, government policy can change in an attempt to drive production of certain crop types. For example due to increasing meat demand, particularly from developing countries [108], there is a large increase in demand for protein crops such as soya. Alongside this, there is a decline in the use of European legumes such as peas and beans [73]. At the moment, the United Kingdom imports most of its protein

from abroad; however, these prices are going up due to the growing global demand for soya [41]. At the same time, growing more protein can improve diversity, and thereby increase resistance against disease and climate change, as well as improve supply security [53]. For these reasons, reforms of the Common Agricultural Policy that are now being implemented includes two measures specifically aimed at increasing the amount of protein crops grown [71].

Changes to crop rotations on a large scale may have implications for food security, as well as landscape and environmental impacts. It is therefore of interest to examine the impact that future changes in climate, government policy, or other relevant factors may have on future crop distributions.

2.3 Difficulties

There are a number of difficulties and complexities that arise in land use modelling that must be accounted for, and we will discuss some of these in this section.

As mentioned in section 2.1, there are a very large number of factors that influence a farmer's decision process. We cannot build a statistical model that includes all of these variables as the model would be too complex. In addition, data is not available for many of the variables. Thus, we must make some assumptions as to which factors are most important to a farmer when they make their decision about which crop to grow. Some are rather obvious to include – factors such as soil type, rotation history, or profit margin – but others are perhaps less obvious in their importance. We are, after all, trying to model a farmer's thought process.

In terms of data availability, a land use model can either be built at a field level, or at a regional level. A model built at field level is desirable, since it is more likely to capture small scale driving processes. However, data is not always available at that scale, as it is harder to collect. In the United Kingdom, field level data is available from 1993 until 2004 from the Integrated Administration and Control System (IACS) database [87]. This provides enough data to build a model, but in order to make predictions from 2015 onwards, we must make the the assumption that the farming situation in 2015 is the same as in 1993–2004. This is perhaps not

a reasonable assumption to make.

There are also issues with missing data. The IACS database was compiled through administrative returns – farmers had to submit forms detailing which crops they had grown, in order to receive subsidies (which at the time were different for different crops). If a farmer did not return the survey in a particular year, for example they may have deemed it not worthwhile claiming the subsidy if the payment would have been very small, then there is no data as to which crop was grown in that field at that time. Therefore we must make some assumptions about the missing data, and consider how we will deal with it.

Rare crop types also pose a difficulty. The overall size of the IACS dataset is very large – we have observations for every field (or, at least, every field for which the subsidy was claimed) in the arable regions of England for 12 years. However, there are very few observations for certain rare crop types, but we still want to be able to make some inferences about these crops. For example, legumes are rather rare, but if we are wanting to analyse the effect a policy change may have on the numbers of legumes being grown as discussed earlier, then we need a model which can still make reasonable inferences even when faced with low data counts.

One mathematical technique we can use in order to help us make inferences about rarer crops is to include prior expert opinion, in the form of a prior distribution which augments the data. However, obtaining this distribution can be tricky, particularly in this setting, as it is perhaps unreasonable to expect farmers to specify a prior distribution without training. This forms part of the process known as *elicitation* [38], which can be used to obtain prior information from an expert, but this can be difficult and time consuming. A different approach, and the one we will take, is to use sets of prior distributions. This is known as a robust Bayesian analysis, but performing this analysis brings complications with respect to making inferences from the model.

2.4 Comparison with other land use models

This work is not the first attempt at modelling agricultural land use (although it does present a novel approach to the problem). Such models have been developed since the 1970s [57], so there are a large number of them in existence. Because of the complex nature of land use, different models have very different inputs and outputs. In this section, we will discuss a few existing land use models.

One such model is the EURURALIS model [33] developed at Wageningen University in the Netherlands, which incorporates variables for climate, income and production levels in well-defined scenarios to make land use predictions. The model has two distinct parts; a *demand* model for the factors which may trigger a change in land use, and a *spatial* model which uses a logit model [2, Chapter 5] to calculate the probability of a specific location being converting from one land use to another. These predictions can be made from a global scale, down to 1 kilometre squares i.e. not at field level. The EURURALIS model has been used to make predictions about the impact of policy changes in rural Europe [83], and in particular was used to aid policy makers analyse the potential impact of 2013 Common Agricultural Policy (CAP) reforms [48]. While the methodology is different to the model we will propose, the concept of using well-defined scenarios is one that we will use in chapter 6.

The Land Use Allocation Model model [55], developed by the University of Reading, is a partial equilibrium economic model [51], aimed at modelling the impact of changes in agricultural policy on England and Wales, and in particular the impact of the 2003 CAP reforms [95]. However, it assumes only four crop types – arable, temporary grass, permanent grass, and rough grazing. It treats each crop type as a homogeneous unit, and the area of study as the farm. Thus, it assumes all agricultural land use across England and Wales is taking place in one farming system, which is not the case. However, we will analyse the impact of changes in policy in chapter 7.

Perhaps more similar to the model we will build is the LandsFACTS model [21], developed at Rothamsted Research. This is a stochastic model that models crop distributions at a field level, and takes into account crop rotations by using transition

matrices. However, the impact of economy, climate and soil are not considered directly, and there is no account of the uncertainty in the model. The LandsFACTS model has been used in a variety of studies relating to the impact of land use changes, including the impact on ecosystems [20], the impact on watersheds [86], and the impact on water quality [28].

This is just a small selection of existing land use models. We aim to improve on those discussed by modelling at a field level, taking into account the uncertainty, and directly looking at the impacts of soil type, the economy and climate.

The starting point for our model is previous work at the Food and Environment Research Agency by Luo [65]. He uses a Markov chain (see section 3.1) to represent crop choices at a field level, following earlier work by Castellazzi et al [22]. Luo uses data to estimate the non-parametric transition probabilities as follows:

$$\pi_{ij}(k) = \frac{F_{ij}(k)}{\sum_{h=1}^{J} F_{ih}(k)} \tag{2.2}$$

where

- $\pi_{ij}(k)$ is the transition probability for moving from crop i to crop j at time k.
- $F_{ij}(k)$ is the number of observations of transitions from crop i to crop j at time k.
- J is the number of crop types in the model.

Luo uses these transition probabilities to simulate future crop distributions. Separately, he assumes a linear regression model to analyse the impact of external influencing factors. In order to analyse the impact of rainfall, $\pi_{ij}(k)$ is assumed to be of the form

$$\pi_{ij}(k) = \beta_{ij0} + \sum_{p} \beta_{ijp} r_p$$

where r_p represents rainfall in a particular time period p. Luo uses standard statistical tests to find which β_{ijp} are significant, giving an indication of which periods of rainfall have a large impact on the transition probabilities. Similarly,

$$\pi_{ij}(k) = \beta_{ij0} + \sum_{c=1}^{J} \beta_{ijc} M_c$$

is used to analyse the impact of profit margin, and once again Luo finds which parameters β_{ijc} are statistically significant, to give an indication of the impact of profit margin on crop choice. Briefly, Luo found that rainfall in the months of August and September, preceding the winter drilling period, is very significant, while crop profit margin is significant for most transitions. An obvious weakness of this model is that he is modelling probabilities, which are naturally bounded between 0 and 1. However, a linear model does not account for this and Luo does not make it clear how he deals with this issue.

We aim to improve Luo's model as follows. We too will assume a stochastic process for crop rotations, but will assume a multinomial logistic regression model for the transition probabilities of this process. This will allow us to analyse the impact of external influencing variables directly within the process, as opposed to in a separate analysis as Luo did. Then, a robust Bayesian analysis will be carried out, in order to capture the uncertainty in the model. This model was developed in [77, 93], and will be discussed in detail in the next two chapters. Chapter 3 will explain how we use a stochastic process alongside a multinomial logit model in order to model crop rotations. Chapter 4 will explore how the robust Bayesian analysis is performed.

2.5 Chapter summary

In this chapter, we have explored some of the concepts associated with land use and land use modelling. In section 2.1, we discussed what drives a farmers crop choice, which includes geographical, climatic and economic factors. We assumed that farmers are mostly interested in maximising profits, although other factors play a role too. Section 2.2 gave an overview of why modelling this system is important. This is primarily because being able to predict future crop distributions, and how these may change with respect to changes in economic or climatic factors, can be very helpful to governments and policy makers. Section 2.3 described some of the difficulties in modelling this process, which mostly relate to the complexity of the system, lack of observations for rare crop types, and the difficulty in obtaining prior

information in the form of a probability distribution. Section 2.4 looked at some other work in this domain, and included a discussion of how we hope to improve on existing models. We particularly focused on Luo's model [65], which is the starting point for our land use model. We will start building this model in the next chapter.

Chapter 3

The model

We have discussed in chapter 2 why we wish to model agricultural land use. In this chapter, we will start building our novel land use model. In section 3.1, we will discuss stochastic processes and the Markov property. This is the starting point for Luo's model [65], and will be the starting point for ours. Section 3.2 will explore one way we can model the transition probabilities in our stochastic process, that is, by using a multinomial logit model. We wish to include prior information in the form of a prior distribution. A special class of prior, which easily updates to the posterior, is known as conjugate. In section 3.3, building on the work of Chen and Ibrahim [24], we will propose a conjugate prior distribution for the parameters of the multinomial logit model, and discuss some of its properties. The work in this chapter was primarily first presented in [77, 93].

3.1 Stochastic processes

We wish to model and predict crop choices in a particular field. That is, given a series of crop choices $y_1, y_2, \ldots, y_k$, we want to be able to predict $y_{k+1}, y_{k+2} \ldots$, and so on. Assume a farmer plants one crop in each field per year, and denote the choice at time k by y_k. Mathematically, we can treat y_k as the realisation of a random variable Y_k. So, over the history of the field, we have a sequence of these random variables and their realisations. This will be the starting point for building our model.

Definition 3.1 [13, Definition 1.1, p. 1] Given an index set I, a *stochastic process indexed by* I is a collection of random variables $\{Y_i : i \in I\}$ on some probability space, taking values in some set.

Thus, the sequence of crop choices in a particular field can be represented mathematically by a *discrete time* stochastic process $\{Y_k : k \in \mathbb{N}\}$, where we assume Y_k can take values in some finite set $\mathcal{Y}_k$, called the *state space*. The elements of $\mathcal{Y}_k$ are called *states*, and represent the farmer's crop choice. We assume that $\mathcal{Y}_k = \mathcal{Y}$ for every time step k; it is reasonable to expect farmers to be able to choose from the same set of crops at any given time. Therefore, we will generally drop the subscript k unless we need to be specific about which time step we are considering. We assume there are a finite number of states J, corresponding to J crop choices i.e. $\mathcal{Y} = \{1, \ldots, J\}$.

As mentioned in chapter 2, farmers often grow crops in rotations. This means that there is a dependency between crops from year to year. One simple mathematical model which captures this is a Markov chain.

Definition 3.2 [13, Definition 1.1, p. 109] A (finite-horizon) discrete time *Markov chain* is a stochastic process with the property that

$$P(Y_{k+1} = y_{k+1} | Y_1 = y_1, \ldots, Y_k = y_k) = P(Y_{k+1} = y_{k+1} | Y_k = y_k) \tag{3.1}$$

Equation (3.1) is known as the *Markov property*, named after the Russian mathematician Andrey Markov. It means that, given the present state, the future is conditionally independent of the past. The assumption of the Markov property simplifies the process considerably, making calculation much easier, as it means the stochastic process can be described simply by the one-step transition probabilities.

Definition 3.3 [13, Definition 2.1, p. 110] The one-step *transition probability* is the probability of moving from one state to another in a single step i.e.

$$\pi_{ij} = P(Y_{k+1} = j | Y_k = i) \tag{3.2}$$

Of course, a Markov chain as described above, where the state space $\mathcal{Y}$ is simply the selection of crops a farmer can choose from, will only capture dependency on

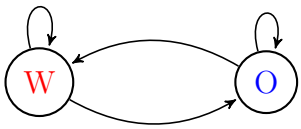
the previous crop grown but no further back in time, and hence will not capture the full influence of, say, a 3 or 4 year crop rotation. However, if we specify our "crop types" as multi-year sequences of crops, we can in fact model multi-year rotations.

For example, say we wish to capture a two year dependence, and, for simplicity, we assume a farmer is faced with a choice of two crops: either wheat (W), or any other crop (O). If we set $\mathcal{Y} = \{\mathrm{W}, \mathrm{O}\}$, then we will only model a one year dependence, as shown in figure 3.1a. Alternatively, we could specify $\mathcal{Y} = \{(\mathrm{W,W}), (\mathrm{W,O}), (\mathrm{O,W}), (\mathrm{O,O})\}$, where (W,W) represents a sequence of wheat in one year followed by wheat in the next year. Meanwhile (W,O) represents wheat in one year followed by other in the subsequent year, and so on. This specification of $\mathcal{Y}$ will capture a 2 year dependence, since our states are 2 year sequences of crops. As can be seen in figure 3.1b, this specification increases the number of states and transitions in our model, but some transitions do not occur in this model specification. For example, we can not go from the state wheat followed by wheat to the state other followed by other. While we generally will specify states as one particular crop, we will use the modified form of the model to look at repeated transitions of wheat – see section 5.2.3 for more detail.

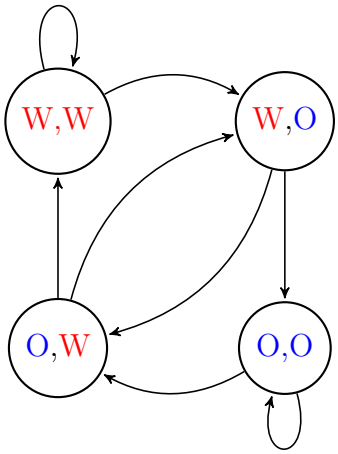
So, equation (3.2) represents a simple Markov chain model for crop rotations. This is the model that Luo [65] assumes, and he estimates π_{ij} from the data, as described in equation (2.2). However, we know that there are also external variables which impact on the transition probability, and we wish to include these in the model too.

As discussed in section 2.1, a farmer's decision is influenced by a large number of factors. Therefore, in order to build a model which more accurately captures the drivers of crop choice, we want to take these factors, or at least some of them, into account. That is, we want our transition probabilities π_{ij} to be a function of these factors.

Let us assume that we have M influencing variables, indexed by time. These variables can be anything which could influence the farmer's crop choice and for which we have data. Let $X_k = (X_{k0}, X_{k1}, \ldots, X_{kM})$ be the random variable which represents these factors. As is usual in a regression analysis (something we will



(a) One year dependence



(b) Two year dependence

Figure 3.1: State transition diagrams for two Markov chains, representing two different specifications of the state space $\mathcal{Y}$. Figure 3.1a show the model specification for a one year dependence: W represents wheat, while O represents other. Meanwhile figure 3.1b shows the specification for a two year dependence. Here (W,W) represents wheat followed by wheat, and so on.

introduce in section 3.2) we will assume that $X_{k0} = 1$. We can include these variables in our analysis by letting our one-step transition probabilities depend on them;

$$\pi_{ij}(x_k) := P(Y_{k+1} = j | Y_k = i, X_k = x_k) \tag{3.3}$$

This means that we do not have a Markov chain in the traditional sense, as described in equation (3.2). Rather, we have a stochastic process where the transition probabilities are influenced by the non-stationary random variable X_k. Note that X_k is not assumed to be part of the state space of the stochastic process. Our stochastic process is non-stationary due to the fact that X_k is clearly non-stationary.

Furthermore, there is perhaps a slight abuse of notation. The realisation $X_k = x_k$ does not occur at time k; rather it occurs somewhere between time k and time $k+1$, and influences the crop choice at time $k+1$. For example, we may choose to take the total rainfall in, say, the two months preceding the planting of a crop at time $k+1$ as one of our regressors. This information would be encoded in x_k. In order to ease notation, we will generally drop the subscript k on the x_k.

Given our current state is i at time k we have a probability mass function on

$\mathcal{Y}$, say $\pi_{i*}(x)$, which gives the probability distribution over the states at time $k+1$. Forward inferences in stochastic processes are generally carried out using using these probability mass functions, via the *transition matrix*. However, we can also use an expectation based approach, as favoured by some authors [30,104]. We are interested in the expectation of being in state j at time $k+1$, given that we are in state i at time k, and that $X_k = x$. We can use a so called transition operator T_x to calculate this expectation.

Definition 3.4 [30, Equation 3] Consider $\mathcal{L}(\mathcal{Y}_{k+1})$, the linear space of all random variables g on $\mathcal{Y}_{k+1}$. Given $X_k = x$, the *one-step transition operator* at time k is a map $T_x : \mathcal{L}(\mathcal{Y}_{k+1}) \mapsto \mathcal{L}(\mathcal{Y}_k)$ defined by:

$$\begin{aligned} T_x(g)(i) &:= E(g|Y_k = i, X_k = x) \\ &= \sum_{j \in \mathcal{Y}} \pi_{ij}(x) g(j) \end{aligned} \tag{3.4}$$

for all $g \in \mathcal{L}(\mathcal{Y}_{k+1})$.

Using the law of iterated expectation, that is, (see for example [104, Theorem 5.3.2]) for random variables X, Y and Z:

$$E(X|Z) = E(E(X|Y, Z)|Z)$$

we can easily calculate the expectation of a random variable $g \in \mathcal{L}(\mathcal{Y}_{k+1})$. For a stochastic process with initial expectation E_0 and transition operators $T_{x_1}, T_{x_2}, \ldots, T_{x_k}$ as defined in equation (3.4), the expectation of a random variable $g \in \mathcal{L}(\mathcal{Y}_{k+1})$ is given by the backwards recursion formula

$$E(g|X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = E_0(T_{x_1} T_{x_2} \ldots T_{x_k} g) \tag{3.5}$$

as defined in [30, Equation 9] (or for a full proof see [29, Section 8]).

Calculation of the one-step transition operator T_x is computationally efficient, as it is simply matrix multiplication. Thus, calculation of equation (3.5) is also straightforward. If we were considering a stationary stochastic process (i.e. $T := T_{x_1} = T_{x_2} = \cdots = T_{x_k}$), as is often considered in the literature, then equation (3.5) simply becomes

$$E(g) = E_0(T^n g)$$

and calculation is very efficient. In our non-stationary case, things are slightly more difficult, but still reasonably simple. We first calculate $T_{x_k}g$, and then $T_{x_{k-1}}(T_{x_k}g)$, and so on.

In the above formulae, we generally use the indicator random variable on $\mathcal{Y}_{k+1}$ of being in state j at time $k+1$. That is, we take our random variable to be

$$g(\mathcal{Y}_{k+1}) = I_j = \begin{cases} 1 \text{ if } Y_{k+1} = j \\ 0 \text{ if } Y_{k+1} \neq j \end{cases} \tag{3.6}$$

as we simply return back to the one-step transition probability since [30, Section 1.2]

$$\pi_{ij}(x) = T_x I_j(i)$$

One may wonder as to why we introduce the one-step transition operators when they simply return the one-step transition probabilities, and it is these probabilities which are usually used for inference in Markov chain theory. However, in chapter 4 we will use sets of probability mass functions to describe a non-stationary *imprecise* stochastic process. The expectation approach defined in this section is the approach commonly taken in imprecise Markov chain theory, as it has computational benefits in that setting. Therefore, we introduced the expectation based approach here to allow for a smoother generalisation to the imprecise case.

3.2 A multinomial logit model

We now have a model where $\pi_{ij}(x)$ represents the probability of moving from state i to state j, given $X_k = x$. In order to model the impact of X_k on this probability, we can assume a parametric form for $\pi_{ij}(x)$.

We have a nominal response variable Y_k with J categories. When $J = 2$ (i.e. we have binary responses), an obvious choice would be to use a logistic regression model (see, for example [2, Chapter 5]), if we were assume the true conditional probabilities are a logistic function of the regressors, and thus assume this probability is monotonically increasing or decreasing. However, farmers can choose from more than 2 crop types, so we require a model which can handle this more complex setting.

Multinomial logistic regression is a generalisation of standard binary logistic regression to J categories. We will follow the work of Agresti [2, Section 7.1], and use the *baseline-category* logit model. This model was developed by a variety of authors [15, 46, 66, 90] in the 1960s and 1970s.

Definition 3.5 [2, Equation 7.1] The *baseline category logit model* pairs each response category with a baseline category j^*, and assumes that, for each $j \neq j^*$:

$$\log \frac{\pi_{ij}(x)}{\pi_{ij^*}(x)} = \beta_{ij} x \tag{3.7}$$

where

$$\beta_{ij} x := \sum_{m=0}^{M} \beta_{ijm} x_m \tag{3.8}$$

For each category i we have a matrix β_i of model parameters. $\beta_i = (\beta_{i1}, \ldots, \beta_{iJ}) \in \mathbb{R}^{J \times (M+1)}$. β_{ij} is the vector of model parameters that describe the model for moving from category i to category j.

Definition 3.6 [2, Equation. 7.2] The *response probabilities* of the baseline-category logit model are:

$$\pi_{ij}^{\beta}(x) := \frac{\exp(\beta_{ij} x)}{\sum_{h=1}^{J} \exp(\beta_{ih} x)} \tag{3.9}$$

Logit models, both binary and multinomial, are commonly used to model probabilities. We assume that the response variable is the outcome of a multinomial trial, that is we are assuming that the response distribution is multinomial. We can see from equation (3.7) that we are also imposing a linear model on the relative probabilities. While this is a specific shape of dependence between any 2 options, interestingly this requirement of monotonicity does not carry through to the response probabilities in equation (3.9) when $J > 2$ (see [2, Page 272]). If $J = 2$, equation (3.9) returns us to the standard binomial logit model, and we do indeed have such a monotonicity restriction on the response probabilities.

When the multinomial logit model is used to model choices, we must also assume *independence from irrelevant alternatives (IIA)* [94, Page 228]. That is, the relative probabilities between two categories does not depend on the presence, or absence, of other irrelevant alternatives. This assumption allows us to model the J choices

as $J-1$ independent binary choices. One category is chosen to be the baseline, and the other $J-1$ are compared to it one at a time.

IIA fails if a perfect substitute exists. The classic example of this is when we assume we are faced with two travel choices – either taking the car or taking a blue bus – and we assume indifference between these methods of transport. If we then add the choice of taking a red bus, IIA implies that the relative probability of taking the car or a blue bus should not change, but this is counter-intuitive since it is reasonable to expect indifference between taking a blue bus and a red bus. However, each crop has unique properties, therefore no alternative is a perfect substitute. We will in fact go further than this. In chapter 5 we will group like crops together. That is, for example, all different types of wheat will be considered purely as one category. Therefore, it seems plausible that IIA holds in this case.

We can easily see that the above model is invariant under translations of the form $(\beta_{i1}, \ldots, \beta_{iJ}) \mapsto (\beta_{i1} - b, \ldots, \beta_{iJ} - b)$:

$$\begin{aligned}\frac{\exp(\beta_{ij}x)}{\sum_{h=1}^{J}\exp(\beta_{ih}x)} &= \frac{\exp(\beta_{ij}x)/\exp(bx)}{\sum_{h=1}^{J}\exp(\beta_{ih}x)/\exp(bx)} \\ &= \frac{\exp((\beta_{ij}-b)x)}{\sum_{h=1}^{J}\exp((\beta_{ih}-b)x)}\end{aligned}$$

This means that, without loss of generality, we can assume that for instance $\beta_{iJ} = 0$, for every i. The category for which the model parameters are zero is the baseline category j^*. As this choice is arbitrary, we choose category J as the baseline category.

There are many other models which can be used to model multinomial categorical responses. For example, we could use the *proportional odds* [2, Section. 7.2] form of the baseline category logit model. Here, we would assume that our β does not depend on j. That is, in equation (3.7), we would assume that $\beta_{i1} = \beta_{i2} = \cdots = \beta_{iJ}$, for each i. Therefore, this is a much more parsimonious model. However, since there are such differences between the different crop types, we would want a model which took this into account. The baseline category logit model does this, whereas the proportional odds form does not, and as such is probably not applicable in this setting. However, we may be able to impose some restrictions on individual β_{ij} in the baseline category logit model if we knew, for example, we never transitioned from rapeseed to legumes. This would reduce the number of parameters we would

need to estimate.

An alternative model used in modelling categorical data is the *adjacent category logit model* [2, Section 7.4.1]. Here, instead of considering the probability of each category versus a particular category as in the baseline category logit model, we compare category 1 to category 2, category 2 to category 3, and so on. However, this model requires the assumption of an underlying ordering of the J categories, so would not be appropriate in this case, as our data is clearly nominal.

3.3 A conjugate prior for the multinomial logit model

Currently, we have a stochastic process representing crop choice, and we assume a multinomial logit model for the transition probabilities. We now wish to include some expert opinion in the model by specifying a prior distribution for our model parameter β. We could choose any prior distribution for β that might reflect the expert's beliefs. Combining this prior with our data will yield a posterior distribution, meaning our inferences contain both observations and prior opinion – this of course is Bayes' theorem.

If our prior distribution and subsequent posterior distribution are from the same family of distributions, then they are known as *conjugate* (see [11, Section 5.2.1] for a formal definition). Finding a conjugate prior is appealing as, due to the fact they come from the same family of distributions, updating a conjugate prior to its posterior is straightforward. Therefore in this section we will define a conjugate prior distribution for the multinomial logit model.

Our starting point is the work of Chen and Ibrahim [24] who, building on the work of Diaconis and Ylvisaker [32], proposed a class of conjugate prior distributions for the family of generalised linear models. The multinomial logit model we have assumed for our transition probabilities is in fact a generalised linear model ([2, Section 7.1.5] shows the formulation of the model in terms of the link function if desired, although it is not necessary for our derivation of a conjugate prior distribution). Specifically, they proposed a prior distribution for the model parameter β

i	x_1	x_2	$n_i(x)$	$k_{i1}(x)$	$k_{i2}(x)$	$k_{i3}(x)$	$k_{i4}(x)$
1	68.1	100	2	0	0	1	1
2	41	148.5	1	1	0	0	0
1	5.6	121.5	1	1	0	0	0
4	56.6	129.1	1	0	1	0	0
4	15.7	133.1	3	1	0	0	2
3	68.9	129.1	1	0	1	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 3.1: An extract of crop rotation data. Here, we have $J = 4$ crop types, and $M = 2$ regressors. i is the previous crop grown. x_1 is the observed rainfall, x_2 is the nitrogen price, $n_i(x)$ is the current crop total for i and x, where $x = (1, x_1, x_2)$, and $k_{ij}(x)$ is the number of crop j being grown.

of a binary logistic regression model of the form [24, Equation 6.1]:

$$f_0(\beta|s_0, t_0, x_1, \ldots, x_n) \propto \exp\left(\sum_{i=1}^{n} s_0(t_{0i}x_i^T\beta - \log(1 + \exp(x_i^T\beta)))\right) \tag{3.10}$$

where $x_1, \ldots, x_n$ are vectors of covariates, and s_0 and t_0 are the hyperparameters. We can see that equation (3.10) is dependent explicitly on the covariates x_i. Chen and Ibrahim view the covariates as fixed a priori, stating that this removes the dependence of the prior on the data. However, even this approach still means we are reliant on knowing the exact locations of our observations before specifying our prior beliefs. However, by simply changing the way the data is structured, we can remove this dependence.

Let $n_i(x)$ represent the number of observations where the previous crop is i, and $X_k = x$. Obviously, $n_i(x)$ will be zero at all but a finite number of $x \in \mathcal{X}$, where $\mathcal{X} := \{1\} \times \mathbb{R}^M$. Thus, $\sum_{x\in\mathcal{X}} n_i(x)$ is a finite sum, for all i. Of these $n_i(x)$ observations, let the crop choice be j in $k_{ij}(x)$ cases, where $\sum_{j=1}^{J} k_{ij}(x) = n_i(x)$ for each i. An extract of the data in this form is shown in table 3.1. As we will see later, this simple transformation of the data structure removes the dependence of the prior on the data.

In order to define a conjugate prior distribution we need to find a prior distribution that, when combined with the likelihood, produces a posterior distribution from the same family. Therefore, the starting point for finding a conjugate prior is to calculate the likelihood of the multinomial logit model.

Definition 3.7 [77, Equation 4] The *likelihood* of the multinomial logit model is

$$\begin{aligned}
l(\beta) &\propto p(k|x,n,\beta) \\
&\propto \prod_{i=1}^{J}\prod_{x\in\mathcal{X}} \binom{n_i(x)}{k_{i1}(x),\ldots,k_{iJ}(x)} \prod_{j=1}^{J} \pi_{ij}^{\beta}(x)^{k_{ij}(x)} \\
&\propto \prod_{i=1}^{J}\prod_{x\in\mathcal{X}}\prod_{j=1}^{J} \pi_{ij}^{\beta}(x)^{k_{ij}(x)} \\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\sum_{j=1}^{J} k_{ij}(x)\log \pi_{ij}^{\beta}(x)\right) \\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\sum_{j=1}^{J} k_{ij}(x)\log \frac{\exp(\beta_{ij}x)}{\sum_{h=1}^{J}\exp(\beta_{ih}x)}\right) \\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\sum_{j=1}^{J} k_{ij}(x)\left((\beta_{ij}x) - \log\sum_{h=1}^{J}\exp(\beta_{ij}x)\right)\right) \\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\left[\left(\sum_{j=1}^{J} k_{ij}(x)\beta_{ij}x\right) - n_i(x)\log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right]\right) \qquad (3.11)
\end{aligned}$$

Note that we are omitting the probability of the initial state from the likelihood as is common in the literature (see for example [14]). We have written the likelihood directly as a function of the parameters. This allows us, using the methodology presented in [11, Page 272], to very simply define a conjugate prior distribution for the parameters of the multinomial logit model.

Proposition 3.1 [77, Equation 5] A conjugate family of prior distributions for the multinomial logit model is:

$$f_0(\beta|s_0,t_0) \propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}} s_{0i}(x)\left[\sum_{j=1}^{J} t_{0ij}(x)(\beta_{ij}x) - \log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right]\right) \qquad (3.12)$$

where s_{0i} and t_{0i} are non-negative functions such that $s_{0i}(x) = t_{0i}(x) = 0$ for all but a finite number of $x \in \mathcal{X}$, with $0 < t_{0ij}(x) < 1$ and $\sum_{j=1}^{J} t_{0ij}(x) = 1$ on those points of x where $s_{0i}(x) > 0$, for each i.

Proof: Using Bayes' theorem we know that

$$f_n(\beta|k,n,s_0,t_0) \propto f_0(\beta|s_0,t_0)p(k|x,n,\beta)$$

which, using equations (3.11) and (3.12), is:

$$\begin{aligned}
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\left[s_{0i}(x)\left(\sum_{j=1}^{J}t_{0ij}(x)(\beta_{ij}x)\right)+\sum_{j=1}^{J}k_{ij}(x)\beta_{ij}x\right.\right.\\
&\qquad\left.\left.-\left(s_{0i}(x)\log\sum_{j=1}^{J}\exp(\beta_{ij}x)+n_i(x)\log\sum_{j=1}^{J}\exp(\beta_{ij}(x))\right)\right]\right)\\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}\left[\left(\sum_{j=1}^{J}(s_{0i}(x)t_{0ij}(x)+k_{ij}(x))(\beta_{ij}x)\right)\right.\right.\\
&\qquad\left.\left.-\left((s_{0i}(x)+n_i(x))\log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right)\right]\right)\\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}(s_{0i}(x)+n_i(x))\left[\sum_{j=1}^{J}\frac{(s_{0i}(x)t_{0ij}(x)+k_{ij}(x))}{s_{0i}(x)+n_i(x)}(\beta_{ij}x)\right.\right.\\
&\qquad\left.\left.-\left(\log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right)\right]\right)\\
&\propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}}s_{ni}(x)\left[\sum_{j=1}^{J}t_{nij}(x)(\beta_{ij}x)-\log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right]\right) \qquad (3.13)
\end{aligned}$$

Thus, conjugacy is easy to see, as

$$f_n(\beta|k,n,s_0,t_0) = f_0(\beta|s_n,t_n)$$

where

$$s_{ni}(x) := s_{0i}(x) + n_i(x) \qquad (3.14)$$

$$t_{nij}(x) := \frac{s_{0i}(x)t_{0ij}(x)+k_{ij}(x)}{s_{0i}(x)+n_i(x)} \qquad (3.15)$$

Note that the subscript n here denotes the fact that we are considering the posterior parameters, as opposed to the hyperparameters, which are denoted by subscript 0. This is standard practice in conjugate analysis [11].

Equations (3.14) and (3.15) show the ease of updating the hyperparameters to the posterior parameters – they are simply functions of the observed data and the hyperparameters. This means that if we know the form of the prior distribution,

then calculation of the posterior distribution is very simple. This is an appealing property of conjugate distributions.

Our likelihood equation (3.11) comes from a multinomial distribution, which is a member of a family of distributions known as the *exponential family*. In general we can define, for observed data k and for some parameter β, a likelihood from the exponential family as follows.

Definition 3.8 [92, Definition 32] $k|\beta$ is from an *exponential family* if there exists functions g, ψ and t such that

$$p(k|\beta) = g(k) \exp\left(\psi(\beta)^T t(k) - b(\psi(\beta))\right) \tag{3.16}$$

for all β and k.

Equation (3.16) has three key components: the natural parameter $\psi(\beta)$; the natural statistic $t(k)$; and the normalising log-partition function $b(\psi(\beta))$, along with some function $g(k)$. It is clear to see from equations (3.11) and (3.16) that our natural statistic is the tensor

$$t(k) \coloneqq k = [k_{ij}(x)]_{i\in\mathcal{Y}, j\in\mathcal{Y}, x\in\mathcal{X}}$$

and that our natural parameter is the tensor

$$\psi(\beta) \coloneqq \left[\sum_{m=0}^{M} \beta_{ijm} x_m\right]_{i\in\mathcal{Y}, j\in\mathcal{Y}, x\in\mathcal{X}}$$

As $k_{ij}(x)$ is zero for all but a finite number of $x \in \mathcal{X}$, we can see that although the natural statistic technically lives in an infinite dimensional space, we can actually treat it as living in finite (but unbounded) space. However, in order to apply some of the common theorems for exponential families, we require that the natural parameter space is an open subset of the parameter space. It is obvious that this is not the case here – clearly, $\dim(\beta) < \dim(\psi(\beta))$. This means that our likelihood comes from an exponential family without full rank, that is, it is a *curved* exponential family.

A key problem we are faced with is the choice of hyperparameters. Equations (3.14) and (3.15) show that the hyperparameters can be interpreted as a prior virtual sample (see for example [107]). Here, for a given value of x, we would have

$s_{0i}(x)$ observations, of which $s_{0i}(x)t_{0ij}(x)$ are crop choice j, for each i. However, the implications of such a specification may not be entirely clear to an expert, and therefore it seems more appealing to relate the hyperparameters to the prior predictive. Furthermore, this method has been criticised for producing prior distributions with variances that are too small [38].

In the case of the regular exponential family, this is done through a famous result by Diaconis and Ylvisaker [32, Theorem. 2]. If this were the case, then $t_{0ij}(x)$ would be equal to $\hat{\pi}_{0ij}(x)$, the prior expectation of $\pi_{ij}(x)$. However, as we have a member of the curved exponential family, this result does not always hold. Inspired by [32, Theorem. 2], we can arrive at some sort of interpretation for t_0 as follows. Using the fundamental theorem of calculus, we know that, since f_0 is a density,

$$\begin{aligned}\int_{-\infty}^{+\infty} \frac{\partial}{\partial \beta_{ijm}} f_0(\beta_{ijm}, \beta_\bullet | s_0, t_0) \mathrm{d}\beta_{ijm} &= f_0(+\infty, \beta_\bullet | s_0, t_0) - f_0(-\infty, \beta_\bullet | s_0, t_0) \\ &= 0 - 0 \\ &= 0\end{aligned}$$

for a particular β_{ijm}, and where $\beta_\bullet$ represents all of our remaining model parameters, which are fixed. We can see that

$$\begin{aligned}&\int_{\mathbb{R}^{J^2\times(M+1)}} \frac{\partial}{\partial \beta_{ijm}} f_0(\beta | s_0, t_0) \mathrm{d}\beta \\ =&\int_{\mathbb{R}^{J^2\times(M+1)-1}} \left(\int_{-\infty}^{+\infty} \frac{\partial}{\partial \beta_{ijm}} f_0(\beta | s_0, t_0) \mathrm{d}\beta_{ijm} \right) \mathrm{d}\beta_\bullet \\ =&0\end{aligned} \tag{3.17}$$

Differentiating our prior distribution (equation (3.12)) with respect to β_{ijm} gives us

$$\begin{aligned}&\sum_{x\in\mathcal{X}} s_{0i}(x) \frac{\partial}{\partial \beta_{ijm}} \left[\sum_{j=1}^{J} t_{0ij}(x)(\beta_{ij}x) - \log \sum_{j=1}^{J} \exp(\beta_{ij}x) \right] f_0(\beta | s_0, t_0) \\ =&\sum_{x\in\mathcal{X}} s_{0i}(x) \left[x_m t_{0ij}(x) - x_m \pi_{ij}^{\beta}(x) \right] f_0(\beta | s_0, t_0)\end{aligned} \tag{3.18}$$

and when we combine equations (3.17) and (3.18), we obtain the following system of $(J-1)(M+1)$ equations:

$$\sum_{x\in\mathcal{X}} s_{0i}(x) x_m t_{0ij}(x) = \sum_{x\in\mathcal{X}} s_{0i}(x) x_m \hat{\pi}_{0ij}(x) \tag{3.19}$$

where

$$\begin{aligned}\hat{\pi}_{0ij}(x) &= P(Y_{k+1} = j | Y_k = i, X_k = x, s_0, t_0) \\ &= \int_{\mathbb{R}^{J^2 \times (M+1)}} \pi_{ij}^{\beta}(x) f_0(\beta | s_0, t_0) \mathrm{d}\beta\end{aligned}$$

These equations show that $t_{0ij}(x)$ in some sense 'matches' $\hat{\pi}_{0ij}(x)$. Of course, for any given prior specification of $\hat{\pi}_{0ij}$, there will be many different functions t_{0ij} that satisfy equation (3.19) when solving for t_{0ij}, so the choice of $t_{0ij}(x)$ will not be uniquely determined by our prior expectation about $\pi_{ij}(x)$.

There is, however, a special case where we do get a prior predictive interpretation of our hyperparameters. We can specify prior information at any finite number of x, and we can call this set of x values $\mathfrak{X} \subset \mathcal{X}$. We can see that, if we restrict the cardinality of $\mathfrak{X}$ to $(J-1)(M+1)$, that is $s_{0i}(x) \neq 0$ for only $(J-1)(M+1)$ occurrences, then we can see that we can solve equation (3.19) to find

$$t_{0ij}(x) = \hat{\pi}_{0ij}(x) \text{ for } x \in \mathfrak{X}$$

Therefore, if interpretability of the hyperparameters is of importance, we can obtain a direct prior predictive interpretation by specifying prior information at exactly $(J-1)(M+1)$ locations.

3.4 Chapter summary

In this chapter we have introduced the basic theory needed for our land use model. In section 3.1 we introduced stochastic processes and the Markov property, mathematical tools that allowed us to capture some of the behaviour of crop rotations, as the states of this process represent the crop choice in a particular year. A key component in a stochastic process are the transition probabilities, which describe the probability of moving from one state to another. We allowed our transition probabilities to be functions of external, non-stationary variables. In section 3.2 we then assumed that these transition probabilities can be modelled by a multinomial logit model, specifically the baseline-category logit model. In section 3.3 we discussed how expert opinion is included in the model. This involved specifying a

prior distribution, and we proposed using a conjugate prior, due to the simplicity of updating a conjugate prior to the corresponding posterior distribution. Building on the work of Chen and Ibrahim [24], we specified a conjugate prior distribution for the multinomial logit model, and by transforming the data, removed the dependence of their prior on the data. Some properties of the distribution were then discussed, specifically relating to the interpretability of the hyperparameters.

What has been presented in this chapter constitutes a novel approach to land use modelling. Next, we will discuss how we cope with the inherent uncertainty present in land use modelling. Specifically, including a prior distribution allows us to deal better with low data counts, a problem which occurs with rare crop types. This prior distribution represents prior beliefs, and as such is often specified by an expert. In this case, it is reasonable to say the farmer is the expert. However, asking a farmer to specify a unique statistical distribution is perhaps unreasonable. One solution to this problem is to perform a *robust Bayesian analysis*, which will be discussed in chapter 4.

Chapter 4

Robust Bayesian analysis

In the previous chapter we built a novel land use model, namely a Bayesian multinomial logit model on top of a non-stationary stochastic process. However, we made no mention of the difficulty we may be faced with in specifying a prior distribution. This chapter will focus on how we address this issue. We will perform a robust Bayesian analysis, which involves taking sets of prior distributions in order to reflect our uncertainty.

Section 4.1 will discuss why we want to perform a robust Bayesian analysis, while section 4.2 will discuss how we will actually perform this analysis by detailing how we will select our set of prior distributions. This material was first presented in [77, 93], and constitutes a novel approach to imprecise multinomial logistic regression, based on conjugate analysis. Performing this analysis turns our stochastic process into an imprecise stochastic process. Therefore we need to introduce new theory to make forwards inference – this will be discussed in section 4.4. We will use the work of de Cooman et al. [30] to take a first step at modelling imprecision in a non-stationary stochastic process which is influenced by non-stationary random variables.

Finally, in a robust Bayesian analysis, efficient computation is very important since we are dealing with a large number of prior distributions. Section 4.3 will discuss various approaches we could use, with specific focus on MCMC and MAP estimation.

4.1 Why perform a robust Bayesian analysis?

We wish to perform a Bayesian analysis for our model parameter β, which allows us to include expert opinion in our model. The expert formulates their beliefs in the form of a prior distribution for β. However, this can be problematic for a number of reasons. Firstly, as we are trying to model farmer behaviour, our experts are the farmers themselves. Therefore it is their prior knowledge about β which is of importance, but it is perhaps unreasonable to expect a farmer to be able to formalise their beliefs in the form of a specific prior distribution for β without training. This training constitutes part of the process of *elicitation* [25, 38], where we aim to translate an expert's prior beliefs into a mathematical distribution. After training an expert, one must elicit information from them, and then iteratively fit distributions until a distribution which has been deemed to 'adequately' represent their prior beliefs has been elicited [38, Section 1.2]. However, this is a non-trivial process.

Nevertheless, let us assume that we can obtain a prior distribution from a particular farmer. In a particular region there are a number of farmers, so we could ask multiple farmers to specify a prior distribution. How then do we use this collection of (presumably different) prior distributions? One solution would be to have a separate model for each farmer, but that would quickly become unmanageable. Another would be to use *opinion pools* [40, 88], where we take a weighted combination of the distributions, but this method is not without difficulties, such as how do we choose the weights?

Let us assume that our group of experts do produce a single prior distribution, either by all specifying the exact same distribution, or perhaps as the result of using an opinion pool. We can make very small changes to this distribution, which may still reflect the beliefs of our expert(s). This means that we have a number of distributions that could be considered for our analysis. A Bayesian analysis requires a single prior distribution to be specified, so which one of these distributions do we pick?

It is clear that we have uncertainty with respect to the specification of this subjective prior information. One way to represent this uncertainty is to use sets of

prior distributions - this is known as a *robust Bayesian analysis*. When each prior is combined with the likelihood function, we obtain a set of posterior distributions. We can then use this set of posteriors for inference, which results in bounds on inferences such as probabilities and expectations.

There is a large body of work on probability bounding, spanning back to early work by Boole [16] and Keynes [56]. Good [43–45] was the first Bayesian to explicitly use sets of priors. Further developments include Berger [8, 9], who developed robust Bayesian analysis, Williams [105, 106], Walley [98–101], Kuznetsov [58], Weichselberger [103], and many more. The robust Bayesian approach is just one method among many that can be included under the umbrella term *imprecise probability*.

We will assume that there is almost no prior information about our hyperparameters. As such, we will take a very large set of prior distributions to represent this ignorance. Rather than taking many distributions to model prior near-ignorance, one can also take a single *non-informative* prior distribution (see for example [11, Chapter 5]). Exponents of using a non-informative distribution include Jeffreys [52], Box and Tiao [17], and Bernardo [12]. Walley [99, Section 5.5] is a major critic of the approach, and believes that in order to truly represent prior near-ignorance one cannot use a single prior distribution. We will use the sets of priors approach here as it is highly intuitive – we have difficulty in specifying a single prior distribution, so we shall use all reasonable prior distributions.

Of course, this raises the question as to which prior distributions are 'reasonable' to include in our set of priors. If we were eliciting a prior distribution, then part of the process is to discuss the adequacy of the elicitation [38]. So, in some sense, a 'reasonable' prior distribution can be thought of as one that adequately represents the expert's prior beliefs. When considering a robust Bayesian analysis, Berger [10, Section 4.3.1] suggests that the size of the set of prior distributions should represent the prior knowledge of the expert. If the expert has large amounts of prior knowledge, then the set of priors should be chosen to be small in order to represent this. However, if we are assuming that we can obtain very little (if any) prior knowledge about the hyperparameters, then, following Berger's suggestion, our set of priors should contain all, or nearly all, priors. Therefore, for our purposes

in this prior near-ignorance model, the term 'reasonable prior distribution' actually has no real meaning, since we will simply take (nearly) all priors anyway. How we do this will be discussed in section 4.2.

The main criticism of the robust Bayesian approach is the increased complexity that arises – both Lindley [62] and Bernardo and Smith [11] express this concern. Clearly, this is a valid observation. However, it is important to have an understanding of the uncertainty in any inferences we make. In a situation such as this, where prior information is very difficult to obtain, the added computational complexity is justifiable as taking a set of prior distributions more accurately represents our knowledge of the system.

An alternative approach to obtaining a measure of uncertainty could be to obtain credible intervals on a single prior, through a method such as MCMC. However, it is non-obvious how this uncertainty would propagate through the stochastic process which underlines our model, whereas, if performing a robust Bayesian analysis, the resulting model aligns well with existing theory for imprecise Markov chains. Of course, this approach captures a different aspect of uncertainty – the robust Bayesian intervals acknowledge uncertainty with respect to specification of a prior distribution, whereas credible intervals provide uncertainty bounds for the parameters. Furthermore, a precise Bayesian model always gives an exact expectation, and so one would still worry about sensitivity against the prior. To analyse the impact of the choice of prior, we would end up using a set of prior distributions, thereby returning to a robust Bayesian analysis anyway. Therefore, a robust Bayesian analysis seems appropriate, particularly in this setting. The rest of this chapter details how we go about performing this analysis.

4.2 Sets of conjugate distributions

In order to perform a robust Bayesian analysis, we need to propose a set of prior distributions. Berger [10, Section 4.3.1] suggests various considerations which should be taken into account when defining this set of priors. This essentially involves making sure the set represents our prior uncertainty, by being both interpretable

and of a size that includes all reasonable prior distributions, whilst also maintaining computational efficiency.

We will follow a similar approach to Walley's *imprecise Dirichlet model* [100]. The Dirichlet prior distribution for a multinomial model with parameter θ can be written as [100, Section 2.2]

$$f_0(\theta|s,t) \propto \prod_{j=i}^{J} \theta_j^{st_j-1}$$

The imprecise Dirichlet model can then be defined as [100, Section 2.3] the set of all Dirichlet distributions such that $0 < t_j < 1$, $j = 1, \ldots, J$, and $\sum_{j=1}^{J} t_j = 1$, and for a fixed positive constant s. This set of prior distributions represents a situation of prior near-ignorance.

By taking sets of prior functions for our hyperparameters s_0 and t_0, we can perform a robust Bayesian analysis in a similar vein to the imprecise Dirichlet model. Sections 4.2.1 and 4.2.2 will explore how we can define our set of prior distributions.

4.2.1 Choice of s

As in the imprecise Dirichlet model, we will concentrate on studying inferences resulting from a fixed prior function for $s_{0i}(x)$, namely:

$$s_{0i}(x) := \begin{cases} s & \text{if } x \in \mathfrak{X}, \\ 0 & \text{otherwise,} \end{cases}$$

for some finite set $\mathfrak{X} \subset \mathcal{X}$. $\mathfrak{X}$ is the set of x values where we specify prior beliefs. For now, we will take an arbitrary set of prior functions $\mathfrak{T}$ for t_{0i}. The task we are now faced with is the choice of s. Walley presents various arguments [100, Section 2.5] to suggest that a choice of $s = 2$ is appropriate for the imprecise Dirichlet model, although of course these arguments do not apply to our model.

Diaconis and Ylvisaker [32, Equation 2.10] showed that, for conjugate analysis with a likelihood from a full rank exponential family, the predictive expectation is a convex combination of the natural statistic and its corresponding hyperparameter, where s controls the weight on the prior. In such cases, smaller values of s always produce tighter posterior predictive bounds, and hence stronger posterior inferences.

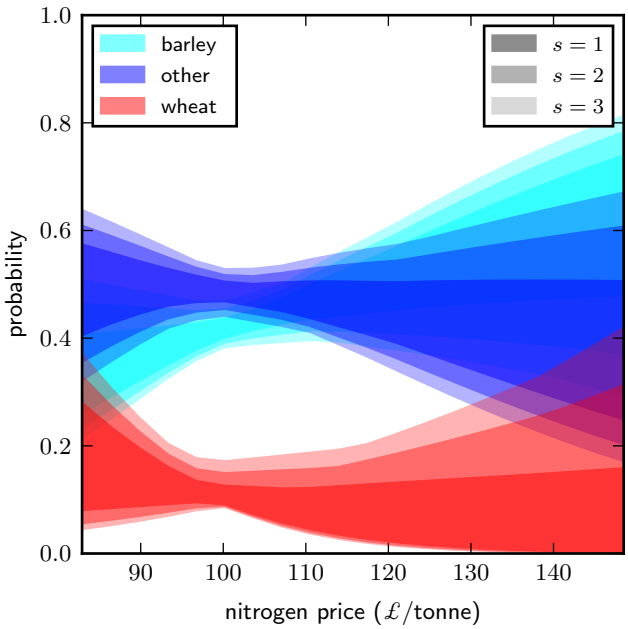


Figure 4.1: Estimated lower and upper posterior transition probabilities for transitions from barley, to either barley (cyan), wheat (red) or other (blue). There are three different values of the hyperparameter s; the darkest region is $s = 1$, the medium region is $s = 2$, and the lightest region is $s = 3$.

Larger values of s produce more cautious posterior inferences by generating wider posterior bounds.

However, as discussed in section 3.3, our likelihood comes from an exponential family which does not have full rank. Therefore, the predictive estimate can depend on the hyperparameters in non-linear or even non-monotone ways. As such, situations where smaller values of s produce wider posterior bounds can not be excluded.

We can see the effect of the choice of s visually, by comparing the results from three different models, each run with a different value of s. In order to do this we need to build the land use model. This will be explored in much greater detail in chapter 5, but, briefly, we use real data to fit our land use model, where here we assume farmers are faced with a choice of 3 crops – wheat, barley or other. We then estimate the lower and upper posterior transition probabilities, which are functions of observed regressors x, in this case rainfall and nitrogen price.

Figure 4.1 shows the bounds on the posterior transition probabilities, and also shows how the probabilities generally change as s changes for our land use model. It shows the lower and upper transition probabilities of moving from barley to each of the three crop choices, represented by the coloured regions, as nitrogen price varies. Furthermore, there are three different models shown in the figure, corresponding to three different values of the hyperparameter s. For the purpose of this illustration, we chose $s = 1, 2$ and 3. As mentioned earlier, $s = 2$ is commonly used in the imprecise Dirichlet model, and we simply varied this value by one either side. The darker the region in figure 4.1 the smaller the value of s. For example, the posterior transition probability bounds of moving from barley to wheat, given that $s = 1$, is represented by the darkest red region. The upper limit of this darkest red region corresponds to the upper posterior transition probability in this case, and the lower limit corresponds to the lower transition probability. The lightest red region shows the equivalent bounds on the probability of moving from barley to wheat, but for a choice of $s = 3$.

It is clear to see in this case that smaller values of s produce tighter bounds. While this is just one particular construction of the model, we note that the behaviour shown in figure 4.1 is typical – we have not witnessed a case where this is not the case. An intuitive explanation for this is that s still relates to the prior variance, and hence still weighs the influence of the prior on the posterior predictive. We will proceed with $s = 2$ for the rest of this thesis, simply because this is the choice commonly used in the imprecise Dirichlet model. However, the choice of s is entirely subjective, and, if having seen the posterior inferences the expert feels that they are too cautious or not cautious enough, the value of s can be changed accordingly.

While we will use a fixed constant value for s_0, there has been some work done on the case where s is not a constant. Walter and Augustin [102] approach the problem by letting both s_0 and t_0 vary freely across their spaces, primarily in order to deal with prior-data conflict. Benavoli and Zaffalon [7, Theorem 1] proposed, when defining a set of conjugate prior distributions for the one parameter exponential family, taking s_0 to be a positive function which can depend on t_0, while still letting

t_0 vary freely across the space it lives in. Our situation is somewhat different – we have a number of s_0, one for each $x \in \mathcal{X}$, while Benavoli and Zaffalon have a single s_0, and our exponential family has more than one parameter. If one has a near-ignorance model, as we will specify in section 4.2.2, there can be convergence issues with an unbounded t_0 and constant s. This is the situation addressed by Benavoli and Zaffalon, but as our t_0 is bounded between 0 and 1, we do not encounter this issue. As such, we will only look at a fixed constant value for s_0, as is common elsewhere in the literature.

4.2.2 Choice of $\mathfrak{T}$

We have a set $\mathfrak{X}$ where we specify our prior information, and at these locations we choose a fixed value s of s_{0i}. We have a set $\mathfrak{T}_i$ of t_{0i}, one for each i, and the choice of these sets affects the posterior inferences. We assume we have very little, if any, prior expert opinion. How can we sensibly choose $\mathfrak{T}_i$ to define a near-ignorance model to reflect this lack of knowledge, while still making reasonable posterior inferences?

As in the imprecise Dirichlet model [100] we are dealing with situations with very little prior expert information. A key feature of the robust Bayesian approach is that the set of prior distributions chosen reflects the prior knowledge available about the system. The obvious choice, if we were wanting to model complete ignorance, is the set of all distributions, which is known as the *vacuous model*. However, as Walley discussed [100, Section 1], a vacuous set of priors leads to a vacuous set of posteriors, and so no useful posterior inferences can be made.

We can, however, assume we are in a state of near-ignorance, and can identify a reasonably vacuous set of prior distributions to represent this lack of prior information. Since we are in a state of near-ignorance, we can take $\mathfrak{T}_i = \mathfrak{T}$ for all i. One way of choosing this near-vacuous set is:

$$\begin{aligned}\mathfrak{T}_v := \{t_{0i} : t_{0ij}(x) = 0 \text{ when } x \notin \mathfrak{X}, \\ 0 < t_{0ij}(x) < 1, \sum_{j=1}^{J} t_{0ij}(x) = 1 \text{ when } x \in \mathfrak{X}\}\end{aligned}$$

For high dimensional problems, calculating inferences from $\mathfrak{T}_v$ is somewhat computationally involved. We can restrict ourselves to the near-extreme points of $\mathfrak{T}_v$,

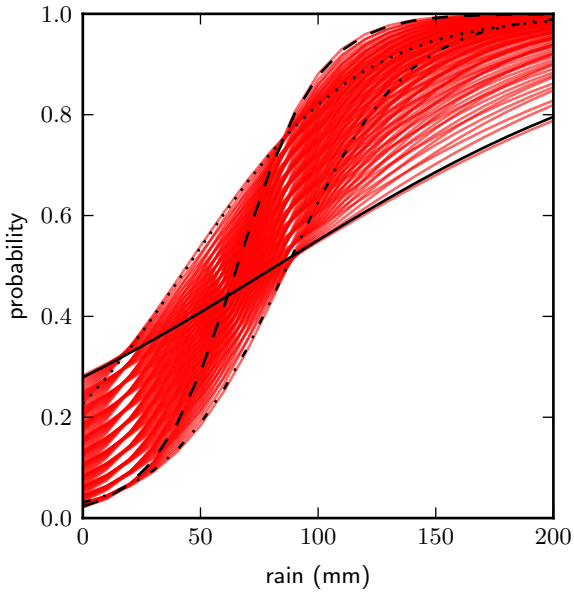


Figure 4.2: A comparison of the posterior transition probabilities for transitions from wheat to wheat in a simple binary model, as a function of rainfall. The convex hull of the red lines represents posterior transition probabilities under the near-vacuous set $\mathfrak{T}_v$, and the convex hull of the four black lines represents the posterior transition probabilities under $\mathfrak{T}'_v$, the extremal points of the near-vacuous set.

namely:

$$\mathfrak{T}'_v := \{t_{0i} : t_{0ij}(x) = 0 \text{ when } x \notin \mathfrak{X}, \\ t_{0ij}(x) \in \{\epsilon, 1-(J-1)\epsilon\}, \sum_{j=1}^{J} t_{0ij}(x) = 1 \text{ when } x \in \mathfrak{X}\} \tag{4.1}$$

for each i, for some $\epsilon > 0$. We have found that we need to bound the t_{0ij} parameters a suitable distance away from 0 and 1, at those points where $s_{0i}(x) > 0$, in order to maintain numerical stability in cases where we have very few observations. It is suspected that the prior blows up as we approach 0 and 1, and so we must bound away from these regions. By choosing ϵ small enough, it has no observable impact on the analysis – it was found through numerical experimentation that $\epsilon = 0.01$ is sufficiently small to maintain numerical stability.

If we wish to use $\mathfrak{T}'_v$ rather than $\mathfrak{T}_v$, then it is desirable that the inferences

generated by both are similar. To find out whether this is the case, we can compare the posterior transition probabilities generated by both sets. In order to do this, we will fit an example model. We will use the simplest possible land use model – a model with only $J = 2$ crop choices (wheat or any other crop) and $M = 1$ regressors (rainfall in the month prior to planting), and calculate the transition probabilities for a particular transition (wheat to wheat) under both sets of priors. We will specify prior information at $(J-1)(M+1) = 2$ locations, $\tilde{x}_1$ and $\tilde{x}_2$ (the reasons for this choice are discussed towards the end of section 3.3).

Figure 4.2 shows the posterior transition probabilities for moving from wheat to wheat generated by both $\mathfrak{T}_v$ and $\mathfrak{T}'_v$. Firstly, following equation (4.1) $\mathfrak{T}'_v$ consists of the following 4 vectors:

$$t_{0i}(\tilde{x}_1) = (0.01, 0.99)$$
$$t_{0i}(\tilde{x}_1) = (0.99, 0.01)$$
$$t_{0i}(\tilde{x}_2) = (0.01, 0.99)$$
$$t_{0i}(\tilde{x}_2) = (0.99, 0.01)$$

Each black line represents the posterior transition probability for moving from wheat to wheat, calculated using one of the vectors in $\mathfrak{T}'_v$. Our robust Bayesian estimate of the posterior transition probability based on $\mathfrak{T}'_v$ is the convex hull of these 4 lines.

The red region shows the transition probabilities under $\mathfrak{T}_v$, although we actually use a 20×20 grid over the parameter space. The estimate of the posterior transition probability under $\mathfrak{T}_v$ is the convex hull of the red lines. Clearly the posterior probabilities generated by $\mathfrak{T}'_v$ approximates the posterior estimates from $\mathfrak{T}_v$ well in this case. We note that this is purely an empirical result, but in the examples we have performed this behaviour is standard, and thus it seems acceptable to use $\mathfrak{T}'_v$ in order to save computation time. However, it is also easy to see that even the size of the set $\mathfrak{T}'_v$ can become large very quickly for higher dimensional problems, This makes the speed of computation very important, an issue which will be discussed in section 4.3.

There may of course be times where we can reduce the size of our set of prior

distributions if we were in a state of reduced ignorance. We saw in equation (3.19) that t_0, in some sense, matches the prior expectation of our transition probabilities, so we potentially could rule out some values of t_0. However, as we are trying to build a model that can still make reasonable inferences even when in a state of prior near-ignorance, we will not look at this case. Hence, from here on in, we will proceed with using $\mathfrak{T}'_v$.

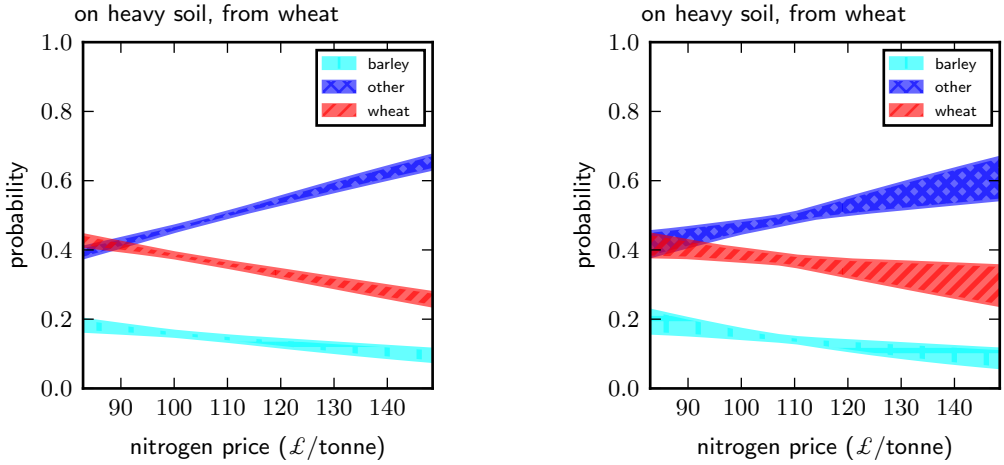
4.2.3 Choice of $\mathfrak{X}$

One final issue to address, albeit not one specific to the robust part of our Bayesian analysis, is the specification of $\mathfrak{X} \subset \mathcal{X}$. This is the set of x values where we specify prior information.

In specifying our conjugate prior distribution, we followed the work of Chen and Ibrahim [24]. They proposed a conjugate prior distribution for binomial logistic regression, which explicitly requires that the locations of x are known a priori. In section 3.3 we introduced the count functions $k_{ij}(x)$ and $n_i(x)$, and this released the constraint of knowing these locations. We have to specify some x for our prior information, but this choice does *not* have to be where our data is located. This allows our prior information to be specified before viewing the data, unlike Chen and Ibrahim's prior.

However, the choice of $\mathfrak{X}$ does influence the inferences and we will demonstrate this graphically by comparing the posterior transition probabilities generated by two different specifications of $\mathfrak{X}$. For the purposes of this example, we will choose the specifications of $\mathfrak{X}$ having actually looked at the values of x. While of course this should not be done in practice (since prior beliefs should represent exactly that), by doing this we can show the effect of having more accurate prior beliefs about the locations of the covariates (in this example, nitrogen price).

Figure 4.3 shows the predicted transition probability for transitions from wheat, as nitrogen price varies, under two different specifications of $\mathfrak{X}$. The nitrogen price in the observed data set ranges from £82.90 per tonne to £148.50 per tonne. Firstly, we chose $\mathfrak{X}_1$ to cover the range of the data set well, specifically $\mathfrak{X}_1 = \{(1, 85), (1, 150)\}$. Alternatively, $\mathfrak{X}_2$ was chosen to cover the data less well, and was chosen to be



(a) Inferences under $\mathfrak{X}_1$

(b) Inferences under $\mathfrak{X}_2$

Figure 4.3: Inferences under two different $\mathfrak{X}$, where $\mathfrak{X}_1$ can be thought of as being specified by an expert with strong prior knowledge of the locations of nitrogen price, while $\mathfrak{X}_2$ is chosen by an expert with less prior knowledge of these locations.

$\mathfrak{X}_2 = \{(1,0),(1,200)\}$. While these choices were driven by the data, an expert with very good prior knowledge of the location of historic nitrogen prices could end up specifying something similar to $\mathfrak{X}_1$, while $\mathfrak{X}_2$ perhaps represents a choice by an expert with less prior knowledge.

We can see that while the model predicts similar trends in the transition probabilities for both sets, the inferences under $\mathfrak{X}_1$ are more precise than those under $\mathfrak{X}_2$. This means our inferences are tighter if the $\mathfrak{X}$ chosen better reflects the actual data. This nicely reflects the idea of a robust Bayesian analysis. If the person specifying the prior information has some idea of where the data is likely to lie, then the posterior bounds are tighter. For example, values of nitrogen price are guaranteed to be positive, and unlikely to exceed a certain upper threshold, so that can be taken into account when specifying $\mathfrak{X}$.

4.3 Posterior inference

In order to make inferences regarding future crop distributions, we need to calculate the posterior transition probabilities:

$$\begin{aligned}\hat{\pi}_{ij}(x) &:= P(Y_{k+1} = j | Y_k = i, X_k = x, k, n, s_0, t_0) \\ &= \int_{\mathbb{R}^{J^2 \times (M+1)}} \pi_{ij}^{\beta}(x) f(\beta | k, n, s_0, t_0) \, \mathrm{d}\beta \\ &= \int_{\mathbb{R}^{J \times (M+1)}} \pi_{ij}^{\beta}(x) f(\beta_i | k_i, n_i, s_{0i}, t_{0i}) \, \mathrm{d}\beta_i \end{aligned} \tag{4.2}$$

where it is worth recalling that $\pi_{ij}^{\beta}(x)$ is a non-linear function of our model parameters β, namely (equation (3.9)):

$$\pi_{ij}^{\beta}(x) = \frac{\exp(\beta_{ij} x)}{\sum_{h=1}^{J} \exp(\beta_{ih} x)}$$

It is equation (4.2) we must evaluate. This is a highly multidimensional integral, over a posterior distribution with no closed form, and thus challenging to evaluate. When we also consider that we are performing a robust Bayesian analysis, so we must evaluate equation (4.2) for each posterior distribution generated by our set of priors, then our method of calculation becomes highly important. The following subsections discuss some possible approaches we can take to calculation.

4.3.1 Markov chain Monte Carlo

A very well known group of methods for evaluating complex posterior distributions, are *Markov chain Monte Carlo* (MCMC) methods. These algorithms provide an approach for sampling from a probability distribution f, by constructing a Markov chain $\{Y_k\}$ that converges to f. That is, we construct a Markov chain where if $Y_k \sim f$ then $Y_{k+1} \sim f$. Samples of this Markov chain are then simulations from f, and can be used to approximate the distribution, or various standard properties of the distribution, such as expectations. MCMC algorithms deal with how to construct a Markov chain that will converge to the distribution of interest. Therefore, MCMC methods present one option for evaluating equation (4.2).

With the advent of modern computing, MCMC methods rose to prominence in the late 1980s and early 1990s, and there is now a huge body of work in the field. The

first MCMC algorithm is widely acknowledged to be the Metropolis algorithm [70] from 1953, which was generalised in 1970 by Hastings [47], into what is known as the *Metropolis-Hastings algorithm*.

For a target distribution f, the Metropolois-Hastings algorithm states [85, Chapter 6] that we can take an arbitrary conditional distribution $q(z|y)$. Then, choosing any reasonable starting point y_0:

1. At iteration k, generate a candidate state $Z_k \sim q(z|y_k)$.

2. Take
$$Y_{k+1} = \begin{cases} Z_k & \text{with probability } \alpha(y_k, Z_k) \\ y_k & \text{with probability } 1 - \alpha(y_k, Z_k) \end{cases}$$
where
$$\alpha(y, z) = \min\left\{1, \frac{f(z)q(y|z)}{f(y)q(z|y)}\right\}$$

3. Repeats steps 1 and 2 n times, where n is the number of iterations specified.

We can use this algorithm to simulate from our posterior distribution. However, in our robust Bayesian setting, we actually have a set of posterior distributions. This means that we need to repeat this process for each of the distributions in our set of posteriors, which is computationally expensive and thus very time consuming. All the calculations for our model we performed in `python` [97]. In order to perform an MCMC analysis, we used the `PyMC` module [36], which uses the Metropolis-Hastings algorithm. By default, this uses a random walk with a standard multivariate Gaussian proposal distribution. Tuning the parameters of this distribution can be performed in order to speed convergence up, although this was not performed here. Furthermore, there are more advanced MCMC methods than Metropolis-Hastings which aim to optimise the convergence of the chain, thereby speeding up the calculations. However, even if we used these more advanced methods, or performed tuning of the current method, we would still be faced with the problem of having to repeat the simulation for each separate distribution. Thus, we ideally want a quicker way to calculate, or at least approximate, the posterior distribution(s).

4.3.2 Maximum a posteriori estimation

A crude but potentially very fast approximation is to use the maximum a posteriori (MAP) estimate for β.

Definition 4.1 The MAP estimate is the mode of the posterior distribution, and is the solution to the system of equations:

$$\frac{\partial}{\partial \beta_{ijm}} \log f(\beta|k, n, s_0, t_0) = 0 \tag{4.3}$$

for $i = \{1, \ldots, J\}, j = \{1, \ldots, J\}$ and $m = \{0, \ldots, M\}$.

That is, we assume all of the mass of the posterior distribution is at the mode. During our MCMC simulation we are estimating the mean of each posterior distribution generated by our set of priors, while when using the MAP estimate we are taking the mode of each posterior. If the posterior is symmetric then the posterior mode will be a good estimate of the posterior mean, but if it is not then using the MAP could be a very crude estimate. However, the major advantage that this approximation has over an MCMC simulation is that it is simply the solution of a system of equations, and therefore we might expect the speed of calculation to be much quicker. Recall from equations (3.8), (3.9) and (3.13) that

$$f(\beta|k, n, s_0, t_0) \propto \exp\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}} s_{ni}(x)\left[\sum_{j=1}^{J} t_{nij}(x)(\beta_{ij}x) - \log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right]\right)$$

$$\beta_{ij}x = \sum_{m=0}^{M}\beta_{ijm}x_m$$

and

$$\pi_{ij}^{\beta}(x) = \frac{\exp(\beta_{ij}x)}{\sum_{h=1}^{J}\exp(\beta_{ih}x)}$$

Using these, equation (4.3) becomes:

$$\begin{aligned}
&\Leftrightarrow \frac{\partial}{\partial \beta_{ijm}}\left(\sum_{i=1}^{J}\sum_{x\in\mathcal{X}} s_{ni}(x)\left[\sum_{j=1}^{J} t_{nij}(x)(\beta_{ij}x) - \log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right]\right) = 0 \\
&\Leftrightarrow \sum_{x\in\mathcal{X}} s_{ni}(x)\frac{\partial}{\partial \beta_{ijm}}\left[\sum_{j=1}^{J} t_{nij}(x)(\beta_{ij}x) - \log\sum_{j=1}^{J}\exp(\beta_{ij}x)\right] = 0 \\
&\Leftrightarrow \sum_{x\in\mathcal{X}} s_{ni}(x)\left[x_m t_{nij}(x) - x_m \pi_{ij}^{\beta}(x)\right] = 0 \qquad (4.4)
\end{aligned}$$

for $j = \{1, \ldots, J\}$ and $m = \{0, \ldots, M\}$. Note that equation (4.4) is very similar to equation (3.19). However, we solve equation (4.4) for $\pi^{\beta}_{ij}(x)$, and can easily use a method for finding roots of a system of equations, such as a Newton-Raphson method.

In our `python` [97], we used the `scipy` [54] function `optimize.root` to solve equation (4.4). By default, this function uses *Powell's hybrid method* [80], an algorithm, which, given an initial value, iteratively searches along conjugate directions until convergence. Powell shows that theoretical convergence is guaranteed for any initial value as long as the functions in equation (4.4) have bounded continuous first derivatives. In practice we have observed no issues with convergence other than, as mentioned in section 4.2.2, when the t_{0ij} parameters are chosen to to be very close to 0 and 1. It is suspected that the prior blows up as we approach these extremal values, and we note that we did indeed observe some instances of convergence failure if we did not bound sufficiently far from 0 and 1.

Whatever method we choose to solve equation (4.4), we obtain an estimate β^* of β, and then our estimate of the posterior transition probabilities is simply:

$$\hat{\pi}^{\beta}_{ij}(x) := \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J}\exp(\beta^*_{ih}x)} \tag{4.5}$$

This is obviously a very simple approximation, but we note that it corresponds to the maximum likelihood estimate, where the data has been augmented with pseudo counts. Hence, it reflects standard practice quite well.

If we can find an estimate of β for all $t_0 \in \mathfrak{T}$ (whether this is the posterior mode using the MAP or the posterior mean from an MCMC simulation) we obtain a set B^* of solutions β^*, one for each $t_0 \in \mathfrak{T}$. Each member of B^* corresponds to an estimate of the posterior transition probability, as in equation (4.5). Therefore,

$$\underline{\hat{\pi}}^{\beta}_{ij}(x) \approx \inf_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J}\exp(\beta^*_{ih}x)} \tag{4.6}$$

is the desired lower approximation of the posterior transition probability, and

$$\overline{\hat{\pi}}^{\beta}_{ij}(x) \approx \sup_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J}\exp(\beta^*_{ih}x)} \tag{4.7}$$

is the desired upper approximation of the posterior transition probability.

The MAP estimation is appealing because of the speed of calculation. Generally, we expect it converge to the solution quicker than an MCMC simulation. Of course, we used a relatively simple MCMC algorithm, and there are algorithms (such as Langevin MCMC [89] or Stochastic-Newton algorithms [67]) designed to speed up convergence, particularly for very high dimensional posterior distributions. In these high dimension situations, there is no guarantee that the MAP calculation will remain computationally quicker. However, for this particular problem the dimensionality is not the computational issue. Rather, it is the number of posterior distributions we obtain due to our set of prior distributions, and hence either the number of MAP calculations or the number of MCMC simulations which must be performed. All of the numerical examples we have performed indicate that, in this setting, estimating β using the MAP estimate is quicker than estimating β using an MCMC simulation, although we note that this may not necessarily always be the case.

If we wish to use the MAP estimate for computational reasons, we do need to check whether it is actually a reasonable estimate of the true posterior β. In order to check this, we can compare the MCMC and MAP estimates of our interval of posterior transition probabilities $[\hat{\underline{\pi}}^{\beta}_{ij}(x), \hat{\overline{\pi}}^{\beta}_{ij}(x)]$, assuming that we run the MCMC simulation long enough to provide accurate inferences.

Before doing so, we must make a note about what we will actually be comparing here. When using MCMC, we simulate samples from the posterior distribution, and use these to estimate the posterior mean. We perform one simulation for each posterior, and hence obtain a set of estimates of the mean. We calculate the lower and upper posterior transition probabilities using this set of estimates, as described in equations (4.6) and (4.7). Similarly, we will obtain a set of estimates of β using the MAP approximation, and again we use equations (4.6) and (4.7) to calculate the lower and upper posterior transition probabilities. The MAP is not equal to the posterior mean, but if the inferences generated by the MAP estimate are similar to the inferences generated by the MCMC simulation, then, as discussed above, it may be desirable to use the MAP estimate for computational ease.

Furthermore, in section 4.2.2 we showed that the inferences generated by the set

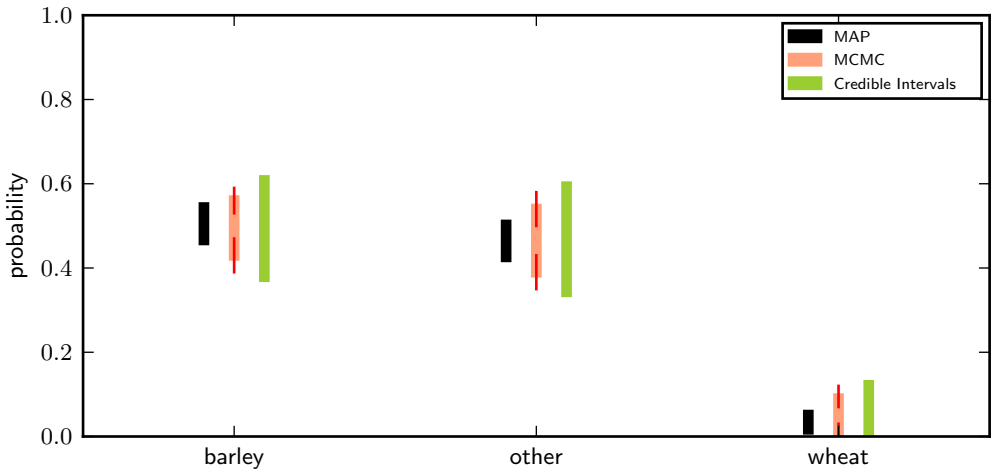
of extremal priors $\mathfrak{T}'_v$ provided a very good approximation to the inferences from the full set of priors $\mathfrak{T}_v$. This comparison was in fact calculated using the MAP estimate of β. For the MCMC estimates we will once again use $\mathfrak{T}'_v$, since it eases computational demands by reducing the number of simulations we must perform. This raises the question as to whether the inferences from $\mathfrak{T}'_v$ will still provide a good approximation of the inferences from $\mathfrak{T}_v$ when we use MCMC, since we are not simply producing a point estimate but rather a full distribution. The answer to this question is that without further analysis we do not know whether using $\mathfrak{T}'_v$ is still appropriate, and performing the analysis to show the validity of this approximation would be incredibly computationally intensive. As such, we will proceed under the assumption that the result we found in section 4.2.2 still holds when we are using MCMC to calculate β.

Figure 4.4 shows the results of our comparison. We used a 3 state model (wheat, barley and other), with 2 regressors (rainfall and nitrogen price), and calculated the lower and upper posterior transition probabilities for transitions from barley to each of the three crops generated by both the MAP estimates and MCMC estimates of β. Recall from equations (4.6) and (4.7) that our estimates of the lower and upper posterior transition probabilities are calculated by

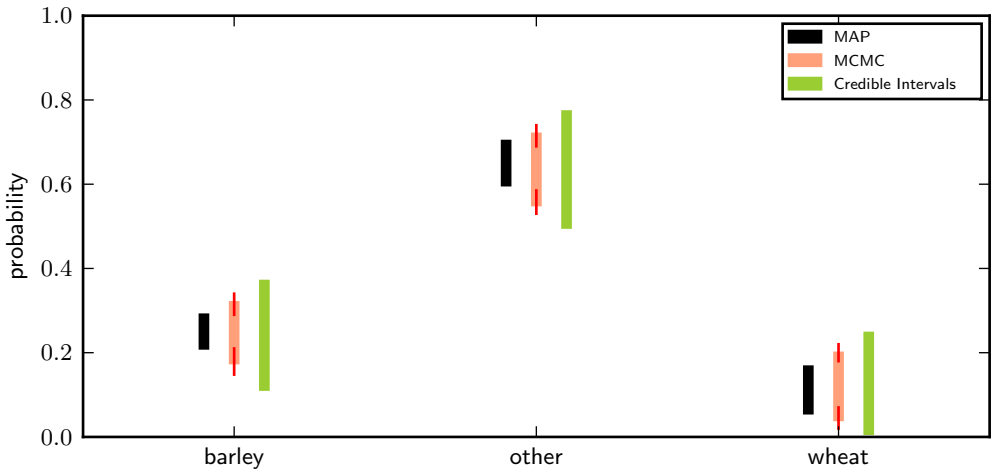
$$\underline{\hat{\pi}}^{\beta}_{ij}(x) \approx \inf_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J} \exp(\beta^*_{ih}x)}$$
$$\overline{\hat{\pi}}^{\beta}_{ij}(x) \approx \sup_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J} \exp(\beta^*_{ih}x)}$$

where B^* is our set of solutions β^*, one for each posterior generated by our set of priors. In the case of the MAP estimate, each β^* is calculated by taking the MAP of a posterior distribution, while in the case of the MCMC estimate, each β^* corresponds to an estimate of the mean of a posterior distribution. The black lines show the interval of the transition probabilities for transitions to barley, other and wheat generated by the MAP estimates of β, and the pink lines shows the corresponding estimates generated by the MCMC estimates of β. We present results for 2 different values of x.

It is clear to see, that for this particular model, the posterior transition probabilities $[\underline{\hat{\pi}}^{\beta}_{ij}(x), \overline{\hat{\pi}}^{\beta}_{ij}(x)]$ generated by the MAP estimate of β (the black lines) are



(a) $x = (1, 10, 50)$



(b) $x = (1, 100, 150)$

Figure 4.4: A comparison of the posterior transition probability bounds for transitions from barley to either barley, other or wheat, generated by the MAP estimate (black line) and an MCMC simulation (pink line), under 2 different values of x. This figure also shows the 90% credible intervals around the extremal predictions from the MCMC (the red lines), and the maximum and minimum of all 90% credible intervals generated by the MCMC simulations, represented by the green lines.

somewhat tighter than the corresponding bounds produced by the MCMC simulation (the pink lines). Nevertheless, the MAP transition probabilities do seem a reasonable approximation for the MCMC transition probabilities as they are located in the same region. Although we only present two figures here, it has been found that this is typical behaviour.

One possible reason why the MCMC bounds are wider than the MAP bounds is that there may be some sampling error present in the MCMC estimates. We can perform some post-simulation diagnostics (see for example [39, Section 11.6]) to investigate this issue further. A major source of sampling error is correlation in the samples we draw, since inference from correlated samples is generally less accurate than from the same number of independent samples [39, Page 294]. To investigate this, we can look at the *autocorrelation* plots from our samples, where we calculate the correlation between samples at different lags. We ran our simulations for 10000 iterations each. *Thinning* [39, Page 295] is often performed, whereby we use only every p^{th} sample for our posterior inferences. This is in order to ease computational demands (since we do not need to store every sample, meaning the simulation can be ran for longer), and also an attempt to remove the dependence between successive samples, although there is some debate as to whether this is justifiable (see [63]). We took every 10th sample, so therefore our inferences are based on 1000 samples.

Figure 4.5 shows a selection of the autocorrelation plots for our model, produced by `PyMC` [36]. Since we are running many MCMC iterations, one for each distribution in our set of priors, we obtain many autocorrelation plots, but we will only show a handful of them here (for space reasons it would not be feasible to show them all). We can see that there is some autocorrelation present, and that this behaviour appears to be typical across all chains. Of course, some autocorrelation is expected, due to the Markov chain structure of the simulation. However, plot **E** in figure 4.5 shows quite large amounts of sequential dependence present for that particular chain even at large lags. We can plot the trace of this simulation to investigate this chain further – this is shown in figure 4.6.

It appears that this particular chain has not converged by the end of our simulation, and we can also clearly see the autocorrelation present in the trace plot. There

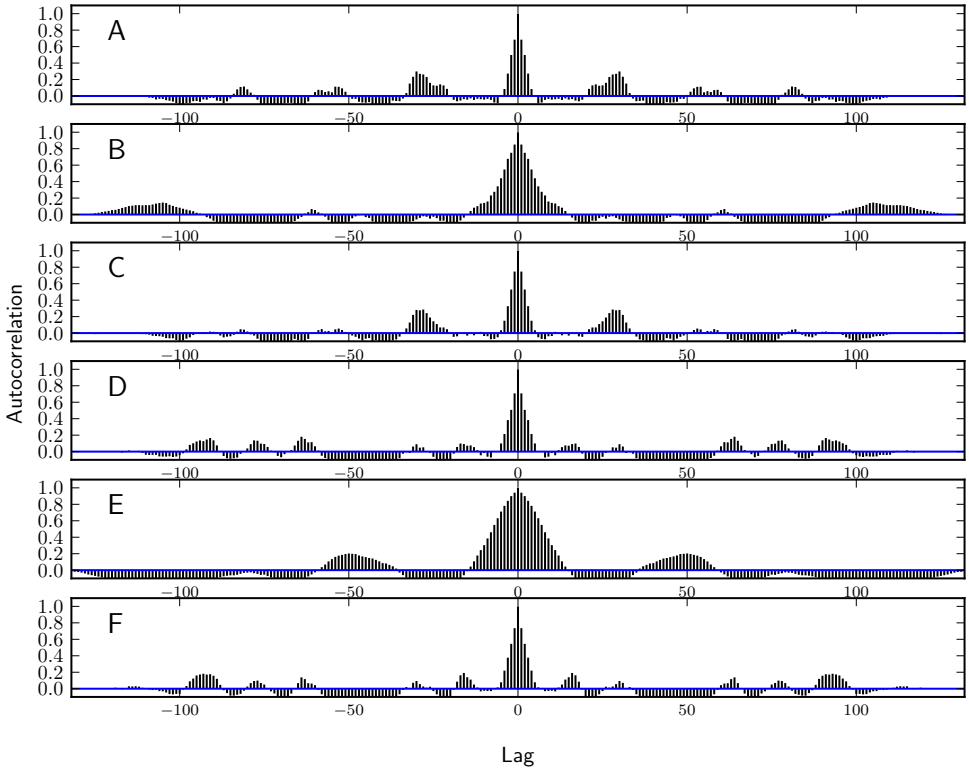


Figure 4.5: A selection of autocorrelation plots from the MCMC simulation.

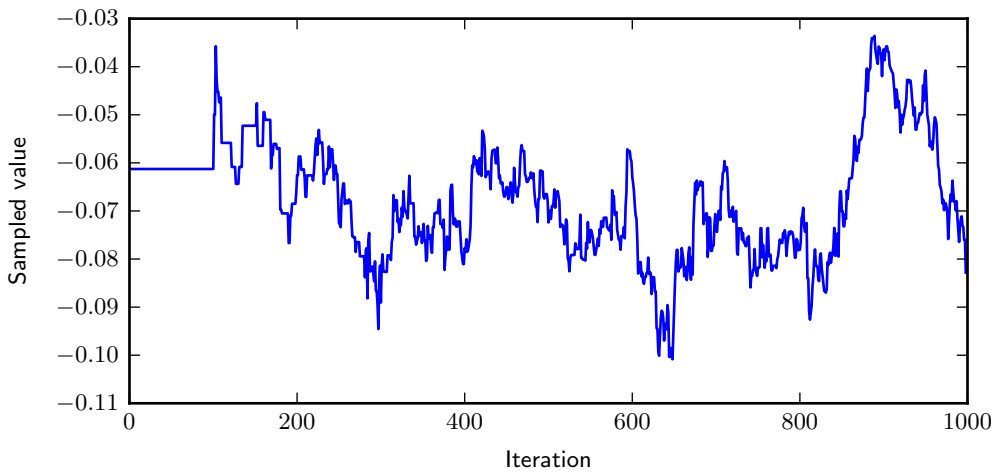


Figure 4.6: Trace plot of the simulation shown in plot **E** of figure 4.5

is a concept known as the *effective number of independent draws* [39, Page 298], where we estimate the equivalent number of independent samples we may have from a correlated sample. For a chain with no autocorrelation, the effective number of independent draws would be equal to the number of samples. Clearly in this case, the effective number of independent draws will be far lower than the 1000 samples we drew because of the sequential dependence present. This will effect the accuracy of the inferences we make. We could run the simulation for longer so that we were effectively obtaining more independent draws. However, even the simple simulation we performed took a very long time to calculate due to our set of priors, as we have to perform the simulation for each individual distribution.

We can also calculate credible intervals [11, Definition 5.4] from our MCMC output, one for each posterior distribution generated by our set of priors. This will provide a measure of the posterior parameter uncertainty present in our system. If we take the MAP estimate of β then we are ignoring this additional uncertainty. For each estimate of the bounds of the posterior transition probabilities generated by the MCMC simulations (the pink lines in figure 4.4), there is one estimate of β which corresponds to the the upper limit of this line, and one estimate of β which corresponds to the the lower limit. We calculated a 90% credible interval around these parameters, and these are shown as the red lines in figure 4.4. We can see that there is some uncertainty around the posterior parameter which we ignore if we use the MAP approximation. However, we can clearly see that there is also uncertainty with regards to specification of the prior in our posterior inferences.

In addition, as we obtain a set of posterior parameter estimates, we can calculate a 90% credible interval for each parameter in this set. The green lines in figure 4.4 show the maximum and minimum of all of these intervals, and we can compare these with the corresponding red lines. As can be seen, these two intervals are generally quite similar. That is, the lowest point of each red line corresponds quite well with the lowest point of the corresponding green line, and similar for the upper limits. As they do not coincide exactly there must be some values of β which do have large amounts of parameter uncertainty, although they do not give us an extreme value of posterior transition probability. Nevertheless, since the intervals are reasonably

similar, we can conclude that by only looking at the extremal values of the posterior transition probability estimates, we are not missing too much extra uncertainty from those values of β that are ignored i.e those that do not generate either the upper or lower posterior transition probabilities. Of course, some of this extra uncertainty may be due to noise as a result of the low effective number of independent draws, as discussed above.

Overall, figure 4.4 suggests that using the MAP approximation for β rather than performing a full MCMC simulation is a crude approximation, but that it produces inferences with the correct location. We will proceed with the MAP approximation due to the expected computational benefits. However, we must acknowledge that by simply taking a point estimate of β we are ignoring the posterior parameter uncertainty. In addition, by using the MAP our posterior inferences appear to be tighter than when using the posterior mean, so this should also be taken into consideration when considering the output of the model.

4.3.3 Other possible approaches

There are other approaches that could be used, although we have not investigated them in great detail. As such, in this section we will present only a brief overview of these potential methods.

Chen and Ibrahim [24, Theorem 2.3] showed that asymptotically, their conjugate prior distribution is reasonably well approximated by a multivariate Normal distribution. While we have a large number of observations, we have a different model for each previous crop. For popular crops, for example wheat, it may be a reasonable approximation. However, in the case where our previous crop is rare, such as legumes, an asymptotic result is highly unlikely to be applicable. Furthermore, we will still require some form of numerical integration. For these reasons we discount this method from our analysis. Using the Laplace approximation (see, for example, [11, Section 5.5.1]) would remove the requirement to perform a numerical integration, but the method is again based on asymptotic assumptions, so may not be appropriate for the reason discussed above.

In addition to the more advanced MCMC methods mentioned earlier, another

approach which may speed up calculation is a likelihood free approach such as approximate Bayesian computation, which provides approximations of the posterior distribution without explicit calculation of the likelihood. This is not something that has been looked at in the scope of this project, however it may provide a better approximation than the MAP estimate, and therefore it would be interesting to investigate in future work.

4.4 An imprecise stochastic process

Our robust Bayesian model means that we have sets of prior distributions distributions, and we use the MAP estimate for β. If we can find a MAP estimate for all $t_0 \in \mathfrak{T}$ then we obtain a set $B^\star$ of solutions $\beta^\star$, one for each $t_0 \in \mathfrak{T}$.

This means that, given $X_k = x$, we no longer have a single probability mass function $\pi_{i*}(x)$ over the states at time $k+1$ in our stochastic process. Rather, we have a set of these probability mass functions, say $\Pi_{i*}(x)$, generated by the set $B^\star$. Therefore, we are no longer considering the stochastic process that was defined in section 3.1. Instead, we have what can be described as a non-stationary *imprecise* stochastic process. We can use, with a slight adaptation, the theory of imprecise Markov chains, specifically the work of de Cooman et al. [30], to define our model and show how we perform inferences in this new setting.

Definition 4.2 [Sec. 2] [30] An *imprecise Markov chain* is a stochastic process that obeys the Markov property with, at every time k, a set Π_{i*} of candidate probability mass functions over the states at time $k+1$, given we are currently in state i.

As in section 3.1, we let each $\pi_{i*} \in \Pi_{i*}$ depend on the non-stationary external variable X_k, and, as such, we have a non-stationary imprecise stochastic process. In order to make inferences, we want to find lower and upper expectations for being in a particular state at a particular time. To do this we can generalise the definition of the transition operator T_x found in equation (3.4).

Definition 4.3 [30, Section 3] Given that $X_k = x$, the *lower one-step transition*

operator $\underline{T}_x : \mathcal{L}(\mathcal{Y}_{k+1}) \mapsto \mathcal{L}(\mathcal{Y}_k)$ is:

$$\begin{aligned}\underline{T}_x(g)(i) &:= \underline{E}(g|Y_k = i, X_k = x) \\ &= \inf_{\beta^\star \in B^\star} \sum_{j \in \mathcal{Y}} \pi_{ij}^{\beta}(x) g(j)\end{aligned}$$

Similarly, the *upper one-step transition operator* $\overline{T}_x : \mathcal{L}(\mathcal{Y}_{k+1}) \mapsto \mathcal{L}(\mathcal{Y}_k)$, given $X_k = x$, is:

$$\begin{aligned}\overline{T}_x(g)(i) &:= \overline{E}(g|Y_k = i, X_k = x) \\ &= \sup_{\beta^\star \in B^\star} \sum_{j \in \mathcal{Y}} \pi_{ij}^{\beta}(x) g(j)\end{aligned}$$

As in the precise case, we can concatenate these operators to get the lower and upper expectations of a gamble $g \in \mathcal{L}(\mathcal{Y}_{k+1})$, similarly to equation (3.5):

$$\underline{E}(g|X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = E_0(\underline{T}_{x_1}\underline{T}_{x_2} \ldots \underline{T}_{x_k} g) \tag{4.8}$$

$$\overline{E}(g|X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = E_0(\overline{T}_{x_1}\overline{T}_{x_2} \ldots \overline{T}_{x_k} g) \tag{4.9}$$

Note that we use a precise initial expectation E_0. We could easily use lower and upper initial expectations, if so desired. However, we have real crop data with which we build our model. Our initial distribution for our imprecise stochastic process can be taken directly from this data, and thus treated as a precise expectation.

Recall from section 3.1 that calculation of the precise transition operator T_x is computationally simple, and this is clearly still true in the imprecise case for $\underline{T}_x$ and $\overline{T}_x$. It is in this imprecise case where using transition operators have a real advantage over using the probability mass functions directly. If we wanted to calculate the expectation of being in a particular state at time k, we would have to look at every combination of possible probability mass function, at every time up to time k, which can quickly become a very large number. By using the backwards recursion formulae in equations (4.8) and (4.9), calculation is simply a sequential linear programming exercise.

Furthermore, it is well known (see for example [30, Page 6]) that, for a random variable g,

$$\underline{E}(g) = -\overline{E}(-g)$$

This conjugacy means that we calculate the lower expectation and the upper expectation follows simply. Thus, performing forward inferences in this non-stationary stochastic process is computationally very efficient.

Again, we can use an indicator function to calculate probability mass functions; in this imprecise setting, we are interested in lower and upper probabilities. The indicator function is of a similar form to the indicator function defined in equation (3.6). However, now we can be in a set of states $\mathcal{J} \subseteq \mathcal{Y}$, and thus our indicator function is of the form:

$$g(\mathcal{Y}_{k+1}) = I_{\mathcal{J}} = \begin{cases} 1 \text{ if } y_{k+1} \in \mathcal{J} \\ 0 \text{ if } y_{k+1} \notin \mathcal{J} \end{cases} \tag{4.10}$$

Primarily, we are interested in the lower and upper probabilities of moving from crop i to crop j, so we actually restrict $|\mathcal{J}| = 1$, and equation (4.10) reduces to the indicator function in the precise case (equation (3.6)), namely:

$$g(\mathcal{Y}_{k+1}) = I_j = \begin{cases} 1 \text{ if } Y_{k+1} = j \\ 0 \text{ if } Y_{k+1} \neq j \end{cases}$$

In chapter 6 we will use the methodology presented in this section to make predictions of future crop distributions.

4.5 Chapter summary

In this chapter, we have explored a robust Bayesian analysis. Section 4.1 explained why we favour a robust Bayesian approach, particularly in this setting. Section 4.2 detailed how we go about performing this analysis. Inspired by the imprecise Dirichlet model, we described a near-vacuous set of prior distributions, and explored some properties of this model.

We then explored how we perform computation in this framework. Section 4.3 detailed the main objects of interest – the posterior transition probabilities. We discussed the calculation difficulties that this poses. This is due to the large dimension of the required integral, and that we have a set of posterior distributions to consider. Section 4.3.1 discussed one possible approach to calculation, namely Markov chain

Monte Carlo simulation. We described one of the most well known MCMC schemes, the Metropolis-Hastings algorithm, that can be used to sample from the posterior distribution. However, as we have sets of prior distributions, it is computationally expensive to calculate posterior inferences using an MCMC method. Section 4.3.2 introduced the idea of using the MAP estimate, that is, assuming all of our posterior mass is at the posterior mode. This is clearly simplistic, but it is much quicker to calculate, and it was found that it gives reasonable approximations of the transition probabilities that were calculated from the MCMC method, and thus we decided to proceed with this approach. Section 4.3.3 briefly talks about other possible approaches to calculation that could be attempted, although we did not go into much detail.

Finally, section 4.4 introduced the concept of imprecise Markov chains. We explored how we use expectation operators to make forward inferences in our non-stationary imprecise stochastic process. These calculations will be used in chapter 6 in order to make predictions of future crop distributions. In chapter 5 our data sets will be introduced and we will explore the modelling of the posterior transition probabilities.

Chapter 5

Analysis of the National Character Area data

In chapters 3 and 4 we built a robust Bayesian land use model for crop rotations, using an imprecise stochastic process. A key part of this process are the transition probabilities, and in this chapter we will use real data to analyse these probabilities. Section 5.1 describes the data set, and we will perform some initial exploratory analysis. Section 5.2 presents some results of the model when using the above data. We will fit the model to a variety of different specifications of the model, and see the impact that different regressors have on the transition probabilities. Some of this analysis was presented in [77] and [78], while some of it is new here. Section 5.3 will describe three approaches to model validation, one of which is a novel non-parametric estimate of the transition probabilities, and was first presented in [78].

5.1 Initial analysis

In this section, we will introduce the data sets we have, and perform some initial exploratory analysis. A full description of the data and how to obtain it is available at [76].

5.1.1 Crop data

Land use models built at a field level are preferable, since they are more likely to capture small scale driving processes [57]. Therefore, we need field level crop data, which the Integrated Administration and Control System (IACS) database [76, 87] provides. The dataset contains year, field code, and crop code for National Character Areas (NCAs) in England for the years 1993-2004 inclusive.

NCAs [72] are regions of the country defined by shared geographical similarities (rather than by political boundaries), making them useful to study from a land use perspective. There are 159 of them in England, and we will use two distinct regions in our analysis – the East Anglian Chalk NCA and the Mease/Sence Lowlands NCA, which are shown in figure 5.1. It is clear from figure 5.1 that East Anglia is much larger than Mease/Sence. In fact, East Anglia is 83870 hectares in size, while Mease/Sence is 32353 hectares. Both of these regions are predominantly arable farming, making them ideal candidates for analysis. There are 4650 fields used over the 12 years in Anglia, providing 35700 observed transitions, and 2094 fields in Mease/Sence providing 13800 observed transitions.

For each field, and each year, we have a crop code detailing which crop was grown at that location at that time. The crop codes in the IACS database are very detailed – it contains over 160 unique crops. For example, there are five different types of wheat, depending on the breed or growing process. This level of detail is unnecessary for our interests in this thesis – here we will explore uses of our model where we do not need to distinguish between crops in such detail. Additionally, using all of the available crop codes would mean we would have a very large amount of parameters to estimate which is not desirable, particularly when we consider the computational challenges we have with our model, as discussed in chapter 4. Therefore, following Luo [65], we will group like crops together. The groupings for the crop types we will consider in our analysis are shown in table 5.1. Other crop types (for example potatoes) can very easily be analysed by the model, but for reasons that will be explored later in this chapter we will concentrate solely on these crop types.

The data was collected from administrative returns submitted by farmers in order to receive subsidy payments. As such there is missing data, presumably primarily

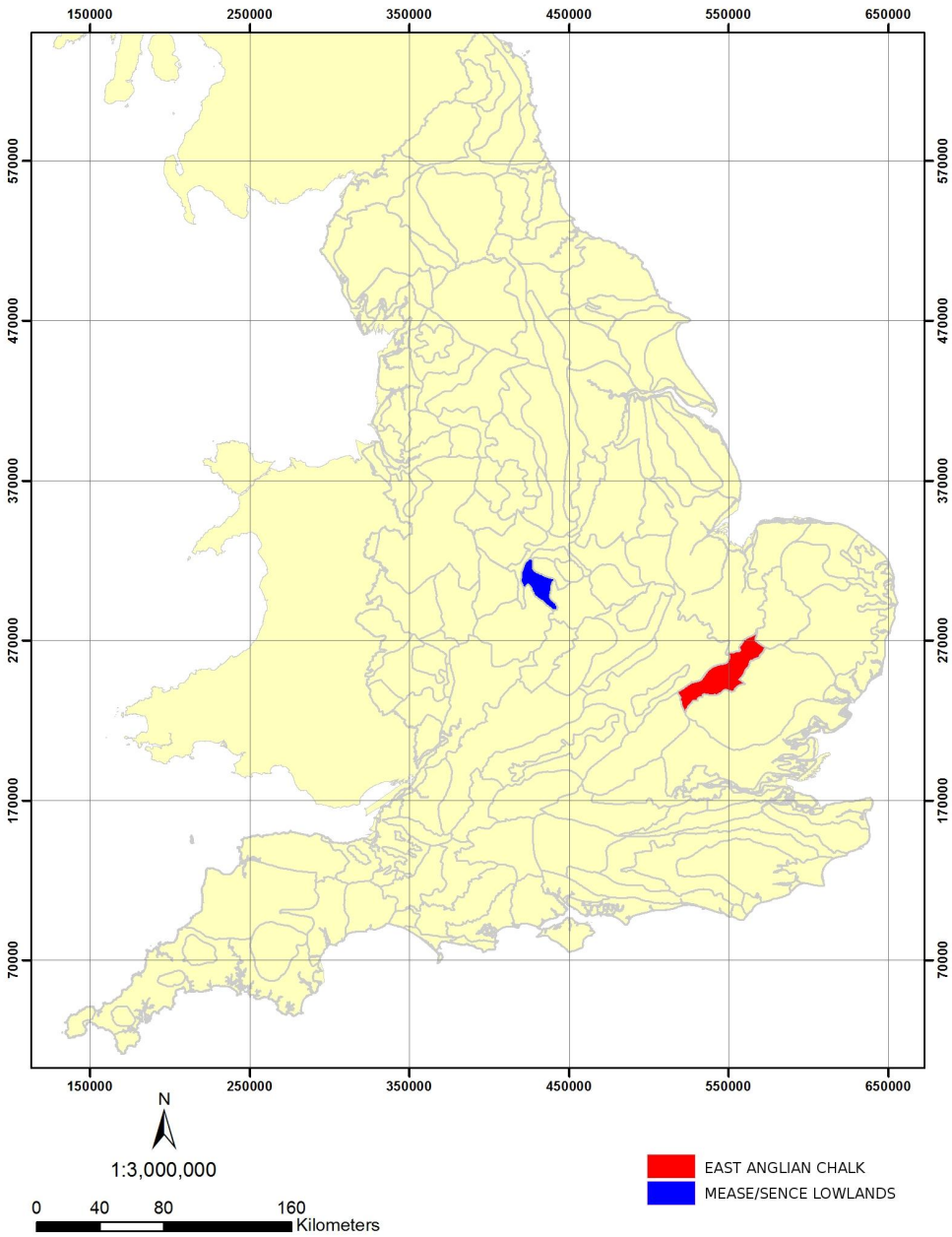


Figure 5.1: The location within England of the East Anglian Chalk NCA, shown in red, and the Mease/Sence NCA, shown in blue.

Crop group	Abbreviation	IACS Crop codes
Wheat	WH	WH1, WH2, WH3, SP1, BU1
Barley	BA	BA1
Rapeseed	RA	RA1, RA6, RA7, RA8
Beans	BE	BE3, BE7, BE9, BE12,
Peas	PE	PE1, PE5, PE9

Table 5.1: Main crop code classification for the NCAs. All other crop types were grouped into the category "Other".

due to farmers not returning the forms. However, this missing data is not particularly troublesome for us. Our model is built on modelling and predicting transitions between crop types, so we can simply eliminate from our database any transitions which are not complete. Out of 49500 observed transitions, there are 1629 instances of incomplete transitions, so our total data set contains 47871 transitions.

After 2004, the way cropping data was collected was changed. Field level data was not collected any more, and instead was collected at a more regional level. To date 2004 is the last field level data collected in England [18]. This of course poses limitations when predicting future crop distributions or analysing hypothetical policy decisions, as we will do in section 5.2 and chapters 6 and 7. Our model is built on data which is over a decade old, so interpretations of these analyses must take that into account. There is potential for field level data to become available again in the near future (section 8.2 includes some discussion about this), thus these analyses are still worth performing, even if just as a proof of concept.

5.1.2 Soil data

Soil type is a significant driver of crop choice [3]. Different crops are more suited to certain soil types, due to nutrient requirements and drainage capabilities. Therefore, it is essential that we include soil type in our analysis.

The national soilscapes survey and the National Soil Map of England and Wales [60, 76] provides a database of soil type in our regions of interest. There are 27

soilscapes in England, which describe the main properties of the soil, and 16 of these are present in our NCAs. The database is, like crop type, too high a resolution than is needed, or indeed feasible, for this study. There are similarities between some soilscapes, meaning we can group them together. The drainage characteristics of the soil are a major factor in the crop choice, as they determine the water holding capacity of the soil, which determines the amount of water available to the crop [3]. As such, grouping together soils with similar drainage capabilities seems very logical. A land-use expert at the Food and Environmental Research Agency proposed that we classify the soilscapes into the following three categories:

- Light soils – free drainage
- Medium soils – slightly impeded drainage
- Heavy soils – impeded drainage

Table 5.2 presents the full classification, and we will use these three categories in our analyses. The soil classification for our regions can be seen in figure 5.2. Clearly Anglia is dominated by light soils, whereas Mease/Sence is dominated by heavy soils. One of the reasons these two particular NCAs were chosen was to give a good spread of all three soil types across both regions. We can see from figure 5.2b that there is very little light soil in Mease/Sence. Due to the small size of this region, this means that there are very few observations on this soil. As such, we generally ignore this soil type in this region – it often leads to numerical problems when considering transitions to or from rarer crop types, where the number of observed transitions can be as low as 0.

Soil type can be included in our model in a number of different ways. We will include it by treating each soil type as a different stochastic process. That is, instead of our transition probabilities being (equation (3.3))

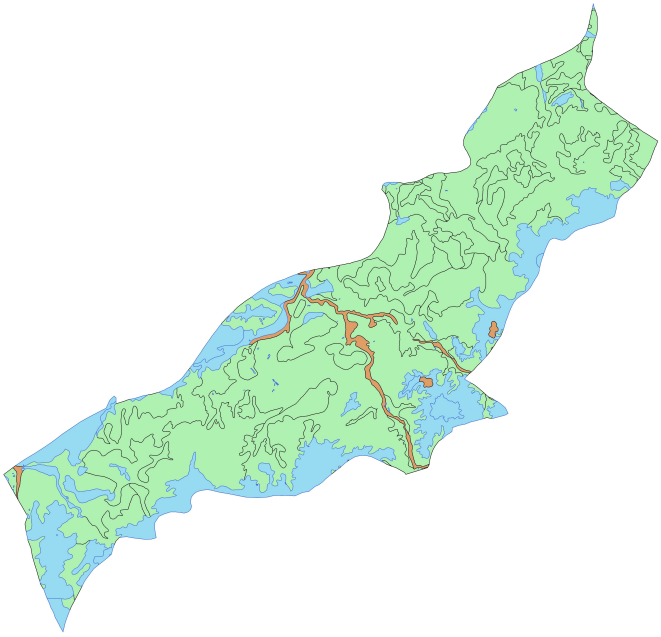
$$\pi_{ij}(x) = P(Y_{k+1} = j | Y_k = i, X_k = x)$$

we actually have

$$\pi_{\gamma ij}(x) = P(Y_{k+1} = j | Y_k = i, X_k = x, \gamma)$$

Soilscape ID	Soilscape	Classification
3	Shallow lime-rich soils over chalk or limestone	Light
5	Freely draining lime-rich loamy soils	Light
6	Freely draining slightly acid loamy soils	Light
7	Freely draining slightly acid but base-rich soils	Light
8	Slightly acid loamy and clayey soils with impeded drainage	Medium
9	Lime-rich loamy and clayey soils with impeded drainage	Medium
10	Freely draining slightly acid sandy soils	Light
11	Freely draining sandy Breckland soils	Light
13	Freely draining acid loamy soils over rock	Light
17	Slowly permeable seasonally wet acid loamy and clayey soils	Heavy
18	Slowly permeable seasonally wet slightly acid but base-rich loamy and clayey soils	Heavy
20	Loamy and clayey floodplain soils with naturally high groundwater	Heavy
22	Loamy soils with naturally high groundwater	Medium
23	Loamy and sandy soils with naturally high groundwater and a peaty surface	Medium
24	Restored soils mostly from quarry and opencast soil	Medium
27	Fen peat soils	Medium

Table 5.2: The soil classification of the different soilscapes present in the Anglia and Mease/Sence NCAs



(a) East Anglia

(b) Mease/Sence

Figure 5.2: Soil distributions for our two NCAs of interest. Green represents light soil, blue medium soil, and orange heavy soil.

where $\gamma \in \{L,M,H\}$ represents the soil type. For notational convenience we will always drop γ. However, it is important to remember that all analyses are repeated three times, once for each soil type. We could of course use a nominal covariate to measure the effect of soil type. This would mean we would have more data for estimation, but it would also involve further parameters to estimate and stronger modelling assumptions. Furthermore, Luo [65] showed that when considering each soil type as a separate stochastic process, the empirical transition probability matrices that he uses are statistically significantly different. For these reasons, we will proceed with using separate models for soil type.

We can easily see that soil type is important by plotting the historic crop distributions on each soil type. Figure 5.3 shows the historic crop distributions for Anglia (figure 5.3a) and Mease (figure 5.3b), separated by soil type, for the crops detailed in table 5.1 (where legumes are beans and peas). Firstly, it is clear that the combined category of other crops is generally dominant (there is one period of exception to this on medium soil in Anglia, where wheat is the most prevalent crop). This is not a surprise, as the category "other" contains many crop codes, including grass which is very common. It also includes land that was not planted on in that year. Wheat is generally the most grown unique crop type, followed by barley, and rapeseed is generally preferred to legumes.

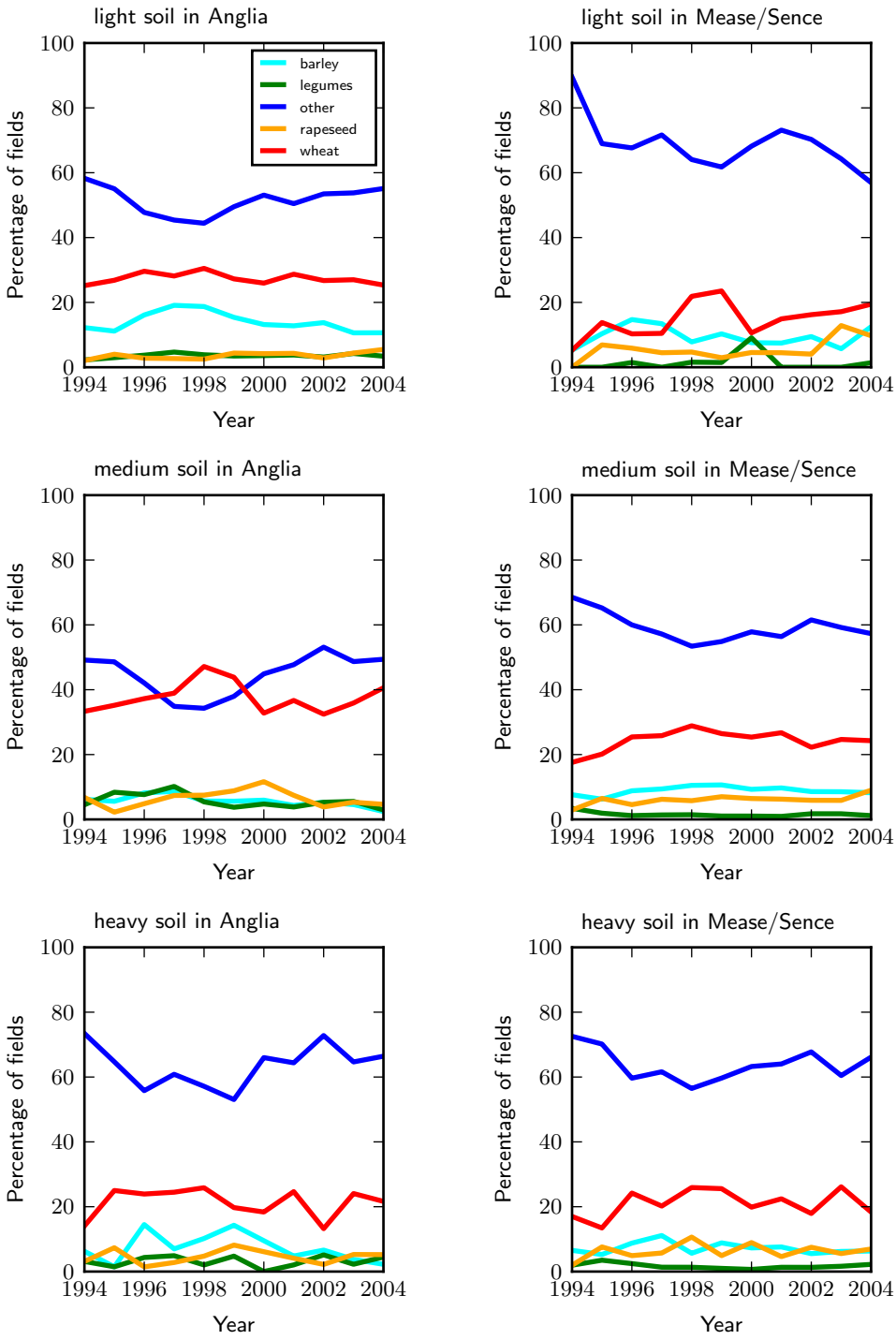
5.1.3 Economic data

Economic factors clearly play a large role in crop choice – farms are businesses [23]. Therefore, we want to include an economic variable in our model, and there are a variety that we could include.

Recall from section 2.1 that each crop has an associated profit margin for each year, defined as (equation (2.1))

$$M_i = P_i L_i + G_i - V_i$$

where, for crop i, P_i is the selling price, L_i is the yield, G_i is the grants or subsidies available, and V_i are the variable costs of growing the crop. The data source we will use for these margins is the John Nix Farm Management Pocketbook [74, 76]. This



(a) East Anglia (b) Mease/Sence

Figure 5.3: Historic crop distributions for the main crops in our NCAs, split by soil type.

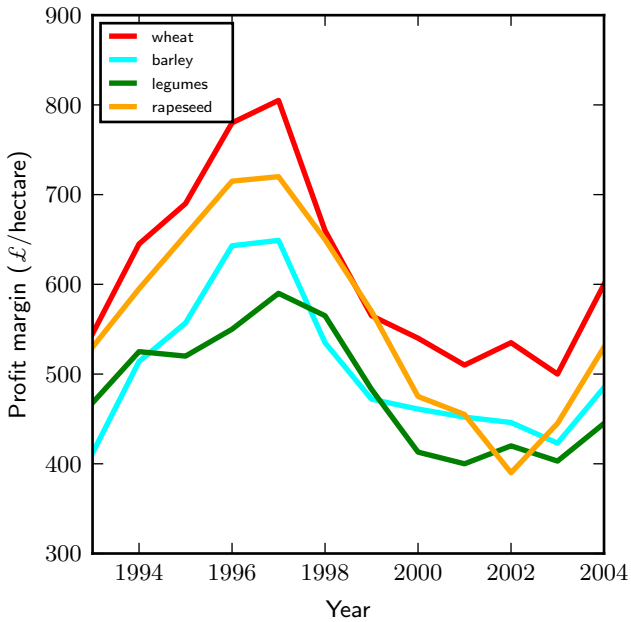


Figure 5.4: Historic profit margins for the main crops of interest. Data reproduced from the John Nix Farm Management Pocketbook [74].

is a yearly series of predictions for various variables of agricultural economy. Each edition gives estimates of gross margin per hectare for each crop for the following year. We could of course use actual observed selling prices for crops. However, we want to model a farmers decision making process - when they choose what crop to grow they do not know what the selling price will be in the coming year. It is for this reason we use this predicted data.

Figure 5.4 shows the historic profit margins for wheat, barley, legumes and rapeseed. We can see that wheat is always the most profitable crop to grow, at least from the crops we are considering. Some other crop types may actually be more profitable, but they are not as widely grown, perhaps due to a need for specialised machinery or a very specific type of soil. The profitable nature of wheat is perhaps one of the reasons why it is the most dominant crop type in our regions of study. The value for legumes shown in figure 5.4 is actually an estimated value. Legumes is not a specific crop type; rather it is a group containing beans and peas. Figure 5.5 shows the historic profit margins of beans and peas, alongside the estimated profit margin for legumes. We take the mean of beans and peas – we could take a weighted

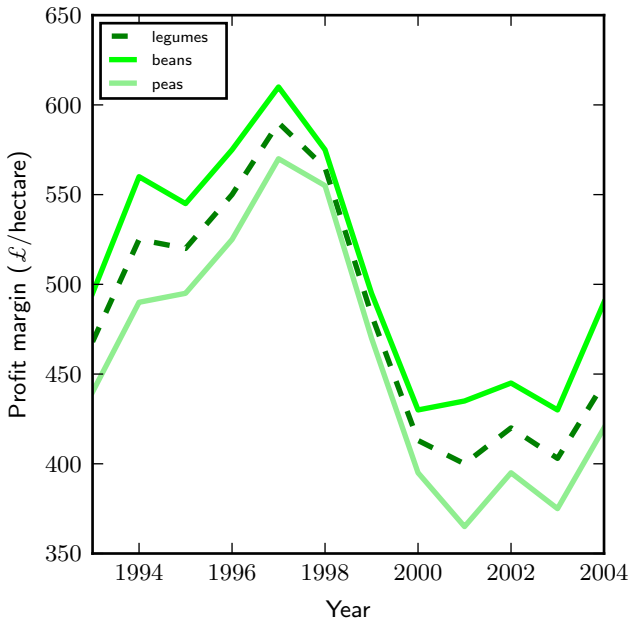


Figure 5.5: Historic profit margins of beans and peas, and the assumed profit margin of legumes. Data reproduced from the John Nix Farm Management Pocketbook [74].

average of the two margins, but as they are grown in reasonably equal proportions it seems logical to simply take the mean.

Yield affects the profit margin too, and this depends on a number of factors, including the skill level of the farmer; climatic factors; soil productivity, and so on. The John Nix books reflect this by including predicted profit margins for three different levels of yield: low, medium and high, which reflect, for each crop, historic average levels of productions [74, Section 1]. However, we have no information as to what level of production actually occurred in our data. We will assume that the production level was always medium, with one exception: farmers sometimes plant wheat repeatedly, which can reduce the yield (due to the build up of the soil-borne disease "take-all" [5]). As such, in this specific case, we will assume the production level is low. Figure 5.6 shows the historic profit margins for wheat under average and low yields (we will discuss repeated wheat in more detail in section 5.2.3).

Variable costs also affect the profit margin. A large part of this component is fertiliser cost, and nitrogen is the nutrient in fertilisers that is most directly related to crop yield [34]. As such, we could use global nitrogen price in our model, and see

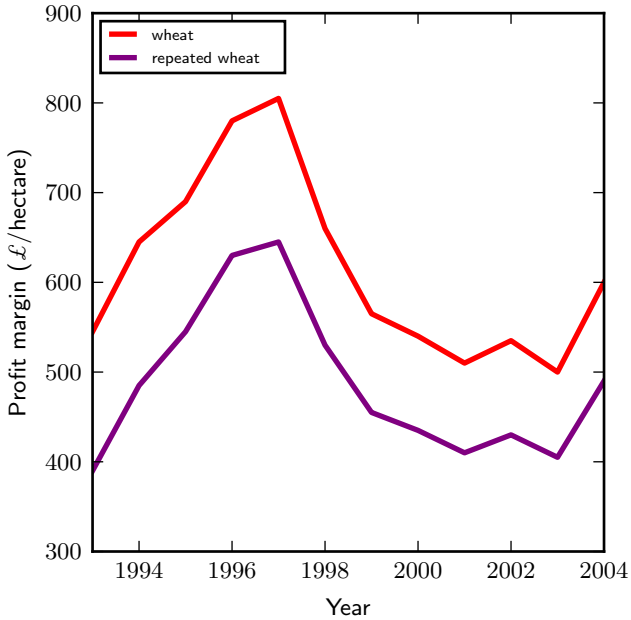


Figure 5.6: Historic profit margins of wheat and repeated wheat, when we assume the yield for wheat is "average" and for repeated wheat is "low". Data reproduced from the John Nix Farm Management Pocketbook [74].

the effect that this has on crop distributions. We will use the mean yearly indexed nitrogen price [31, 76], where the year 2010 is set to be £100/tonne – see figure 5.7.

5.1.4 Rainfall data

Climate plays a role in determining a farmers crop choice [49]. The Met Office provides [68, 76] free data from 37 weather stations around the United Kingdom, and they have many more stations which provide data at a cost. These weather stations provide a number of climatic variables, including monthly rainfall. We will use this data in this study.

We have to decide which month(s) of rainfall to include in our model. The dominant crop in our database is wheat, and it has two main planting periods: winter and spring [96]. We have no information as to which type of wheat the farmers grew, although winter wheat is preferable as it allows the crop to grow for longer, increasing yields [74]. The drilling period for winter wheat is generally September [96]. As such, after discussion with a former wheat farmer, this seems

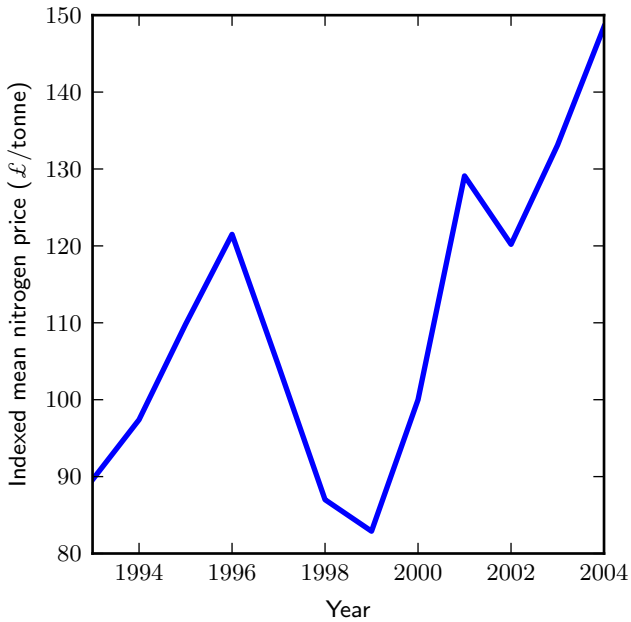


Figure 5.7: The historic mean yearly indexed nitrogen price. The price in the year 2000 is set to be £100/tonne.

like a sensible month to include in our analysis, and coincides well with the results of Luo [65]. Although this is a rather simplistic measure of climate, we will proceed with this for now.

There are a number of weather stations near each of our regions. Ideally, we want to know the rainfall at each field, but this figure is unknown. So, what we can do is to interpolate, which will give us a *prediction* of the rainfall in each individual field at a given time, based on what we have observed at the weather stations. A common method to do this is known as *kriging*.

Kriging

We will follow Cressie [27] to describe the basic theory of kriging. We are interested in predicting the rainfall at each field d in a particular region D at a particular time. Let

$$\varphi(\cdot) = \{\varphi(d) : d \in D\}$$

be a Gaussian random field. That is, a stochastic process with a multidimensional Gaussian parameter. We assume this is the true, unobservable rainfall value at

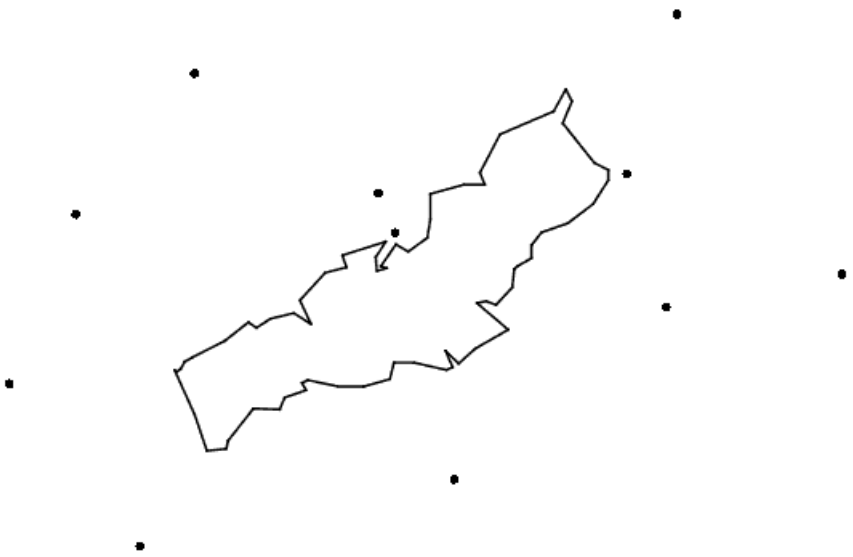


Figure 5.8: The observed rainfall locations for the East Anglian NCA. Each black circle represents one weather station, each of which provides measurements of monthly rainfall.

a given time. We have observations of rainfall at n weather stations, at known locations $\boldsymbol{d} = (d_1, \ldots, d_n)$. These observations are noisy versions of the true rainfall. Therefore, we actually observe the process $\phi(\cdot)$, where

$$\phi(d_i) = \varphi(d_i) + \varepsilon(d_i)$$

and we assume $\varepsilon(d_i) \sim N(0, \sigma_\varepsilon^2)$. Figure 5.8 shows the East Anglian NCA (in this example, our region D), and the observed rainfall locations $\boldsymbol{d}$ for September 1994. We wish to predict $\varphi(d)$, the true rainfall value, for each $d \in D$, from the observed $\phi(d_i)$. For a particular field, located at known but unobserved location d^*, the *simple kriging predictor* of the true rainfall $\varphi(d^*)$ is given by [27, Equation 4.25]:

$$E\left[\varphi(d^*)|\phi(\boldsymbol{d})\right] = E\left[\varphi(d^*)\right] + \mathrm{Cov}\left[\varphi(d^*), \phi(\boldsymbol{d})\right] \mathrm{Var}\left[\phi(\boldsymbol{d})\right]^\dagger \left(\phi(\boldsymbol{d}) - E\left[\phi(\boldsymbol{d})\right]\right)$$

while the *simple kriging variance* of this prediction is given by [27, Equation 4.26]:

$$\mathrm{Var}\left[\varphi(d^*)|\phi(\boldsymbol{d})\right] = \mathrm{Var}\left[\varphi(d^*)\right] - \mathrm{Cov}\left[\varphi(d^*), \phi(\boldsymbol{d})\right] \mathrm{Var}\left[\phi(\boldsymbol{d})\right]^\dagger \mathrm{Cov}\left[\varphi(d^*), \phi(\boldsymbol{d})\right]$$

$\mathrm{Var}(\cdot)^\dagger$ represents the *Moore-Penrose generalised inverse.* When $\mathrm{Var}(\cdot)$ is non-singular, $\mathrm{Var}(\cdot)^\dagger \equiv \mathrm{Var}(\cdot)^{-1}$. For the case when $\mathrm{Var}(\cdot)$ is not non-singular, we refer

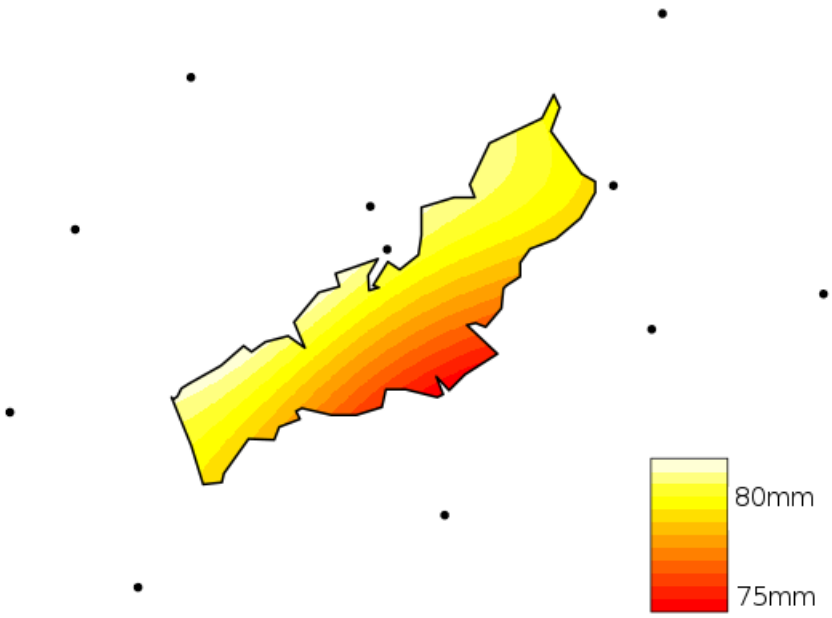


Figure 5.9: The bounded rainfall prediction for September 1994 in the East Anglian NCA, based on kriging.

to [42, Section 11.3.2]. Note that these equations simply come from well known theory about multivariate normal theory for jointly normally distributed vectors; see, for example [6, Page 35], or alternatively the concept of adjusted expectation and variance; see [42, Section 3.1].

We now must assume a form for the covariance structure. There are a variety of methods for this (see Cressie [27, Section 4.1]) but we will not discuss these here, as they involve making numerous assumptions and the mathematics involved is rather cumbersome. However, it involves estimating the spatial similarity between any 2 points in D. This is used to make predictions, one of which is shown in figure 5.9. The figure shows the rainfall predictions for September 1994 in the East Anglian NCA, for a particular assumed form of the covariance. Calculations were performed using the `R` [81] As can be seen, package `geoR` [82]. there is not a huge range in the rainfall predictions over the region. This is due to the reasonably small size of the region. As such, the same happens for any month of rainfall we choose to interpolate on, since rainfall does not vary that much over small changes in space. As Mease/Sence is even smaller, the interpolation seems even less worthwhile. As

such, for computational ease, we will assume the rainfall across an entire region is constant for any given month. We choose the weather station closest to the NCA region as the source of the rainfall data. This is Cambridge for the East Anglian NCA, and Sutton Bonnington for the Mease/Sence NCA. However, if we were to expand the region of study (which could easily be done, as the IACS database contains crop codes for all of England) the methodology presented in this section could be applied when it is not feasible to assume constant rainfall, say if considering multiple neighbouring NCAs.

5.2 Results

We will now use the robust Bayesian land use model developed in the chapters 3 and 4 to model the data described in section 5.1. This section will concentrate on modelling the transition probabilities of the stochastic process, that is, modelling the probability of moving from crop i to crop j in successive years.

5.2.1 A simple model

To begin with we will shall perform a very simple analysis, in order to introduce the model. We will consider a model which involves only three crop types: wheat and barley (the two most popular crop types), and then all other crops grouped together as "other". For now, we will model the transition probabilities as a function of rainfall only, and we will fit the model only on the East Anglian NCA. Figure 5.10 shows the state transition diagram for this simple model. Figure 5.11 shows the upper and lower posterior transitions probabilities from wheat, as a function of rainfall, for all soil types. Figure 5.12 shows the probabilities for transitions from barley. and figure 5.13 shows the same from other.

This is a very simple analysis, however we can still take something from the results. For example, we can see that our model behaves as we would want a robust Bayesian model to – the width of the posterior bounds reflects the amount of information available to us. In this case, transitions from barley (the rarest crop type in this model) on heavy soil (the rarest soil type in Anglia) is the combination

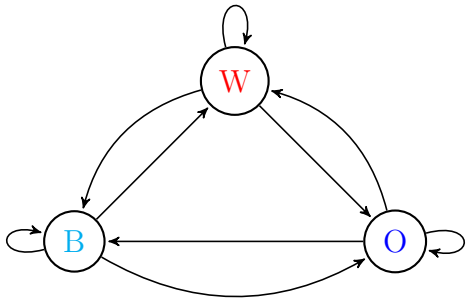
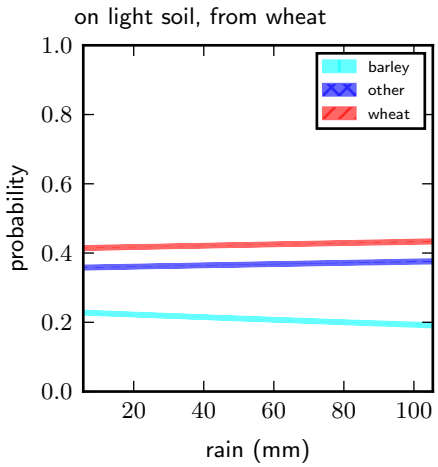


Figure 5.10: The state transition diagram for the simple model, where W represents wheat, B represents barley, and O represents other.

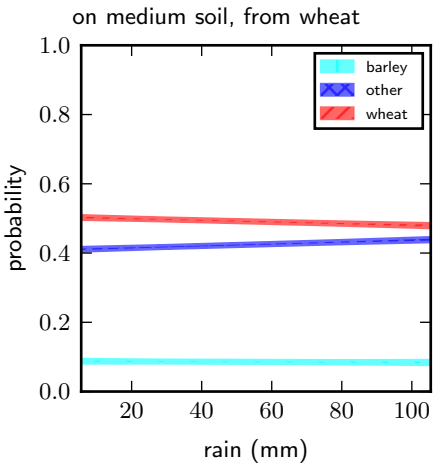
where we have least observations. This is reflected by figure 5.12c having the widest posterior bounds in all of our figures for this model specification (figures 5.11 to 5.13).

By comparing each sub-figure within a figure, we can see there are differences between the soil types. In figure 5.11, we can see that the probability of moving from wheat to barley decreases slightly as rainfall increases on light soil, but on heavy soil we see that this particular transition becomes much more likely as rainfall increases. A likely explanation for this is that on heavy soil a lot of rain in the autumn will increase the probability of not being able to plant wheat, as heavy soil takes a long time to dry out. If a farmer is unable to drill in the autumn, spring sown barley is a common replacement [4], and it appears the model is picking up some of this behaviour. On medium soil when it is wet, the probability of a repeated transition from wheat to wheat decreases, but the probability of transitions to other increases slightly, although of course with the set up of this model we can make no claims as to which crop this is, other than it is not barley or wheat. If we wanted to see which crops were actually increasing, we would have to specify a new model with a greater number of crop types.

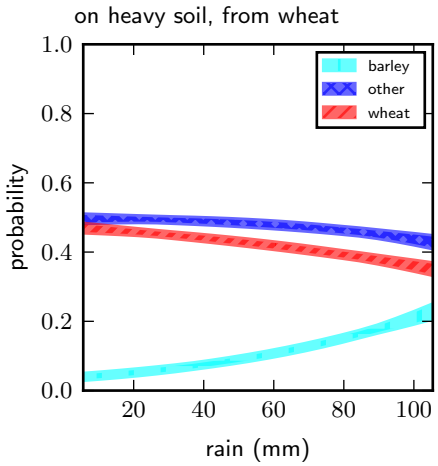
Figure 5.12 shows the transition probabilities from barley. We can quickly see that transitions from barley to wheat are rather rare, for all soil types. Barley and wheat are quite similar crops, as they are both cereals. As such, they are both susceptible to the same diseases, particularly the take-all disease, and so transitions from one to the other are not that common. Wheat to barley is more common as



(a) Light soil

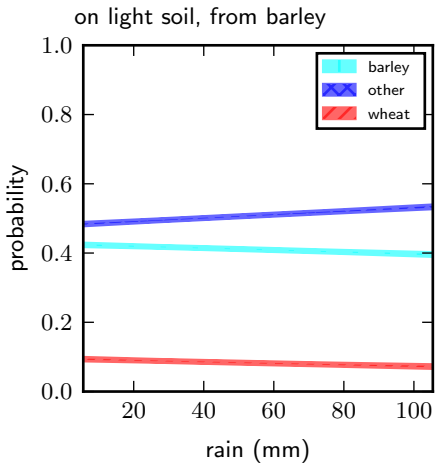


(b) Medium soil

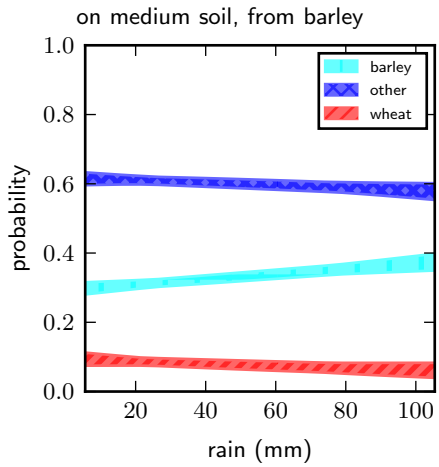


(c) Heavy soil

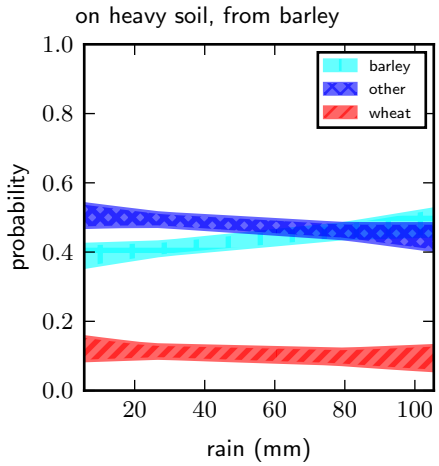
Figure 5.11: Transition probabilities from wheat in Anglia on all soil types, as monthly rainfall varies, for the simple model



(a) Light soil

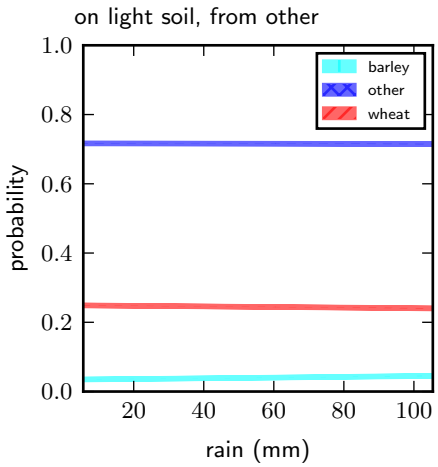


(b) Medium soil

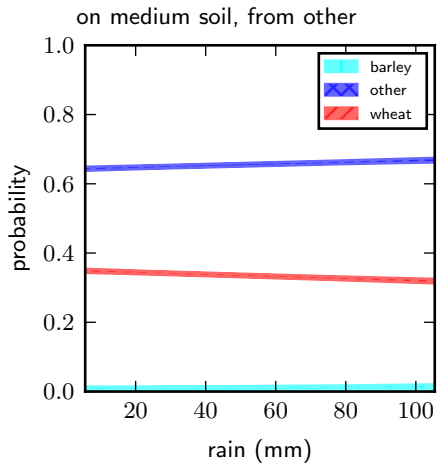


(c) Heavy soil

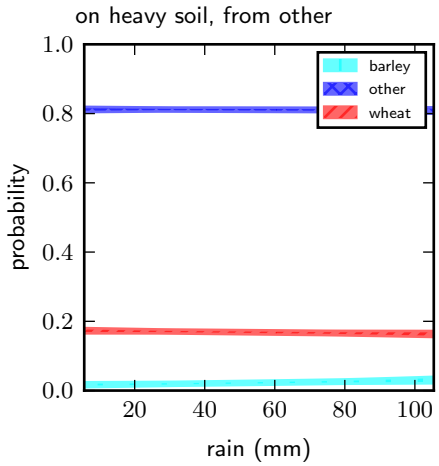
Figure 5.12: Transition probabilities from barley in Anglia on all soil types, as monthly rainfall varies, for the simple model



(a) Light soil



(b) Medium soil



(c) Heavy soil

Figure 5.13: Transition probabilities from other in Anglia on all soil types, as monthly rainfall varies, for the simple model

barley is less susceptible to take-all. There are once again differences between the soil types. On light soil, as rainfall increases, the probability of transitions to other increases slightly, while on both medium and heavy soil, the probability of repeated transitions to barley increases as rainfall increases.

Figure 5.13 shows the transition probabilities from other. Generally we have a high probability of moving from other to other, but apart from that not much else happens. Other is a very general category, so we cannot really draw any conclusions from this. Therefore, we will exclude figures showing transition probabilities from other from the presentation of future results.

For this simple model, in one region, we already have 9 figures to show. When including further crop types, regressors, and performing the analysis on both regions, it would clearly become a very unmanageable number of figures very quickly. Therefore, from here on in, we will present just the particularly interesting graphs.

5.2.2 Break crop model

The previous analysis was included mainly as an introduction to the model and the presentation of results. We can perform far more interesting and relevant analyses. Of particular interest to agricultural scientists is the break crop, which is the crop grown after a main crop in order to "break" the cycle of pests and diseases [3]. In our regions of study wheat is the main crop, and the break crop is often rapeseed or legumes (i.e. peas and beans), since they are not cereal crops. Rapeseed is generally more profitable, but it requires more fertiliser than legumes, since legumes can "fix" their own nitrogen from the air [64]. Therefore, we will include these crops in the model. Barley is of less interest here since it is a cereal crop and hence is not a break crop from wheat. Therefore barley, and all other crop types, will be collected together as "other".

One thing to note is that transitions between legumes and rapeseed occur very infrequently – across both regions and all soil types (47871 observed transitions), there are 5 instances of transitions from legumes to rapeseed, and only 2 observations of rapeseed to legumes. Farmers often aim to grow wheat since it is the most profitable, so a successive growing of break crops is very rare. As such, we

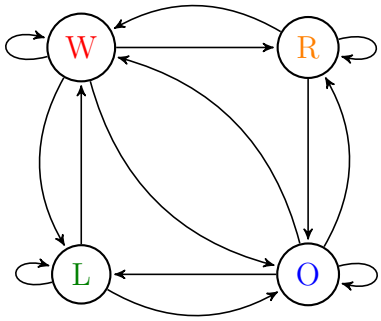


Figure 5.14: The state transition diagram for the break crop model, where W is wheat, R is rapeseed, L is legumes, and O is other. Transitions between R and L do not occur in practice so have been excluded from the model

will exclude these transitions from the model. They are not particularly of interest to agriculturalists, and they can lead to convergence issues within the model. Figure 5.14 shows the state transition diagram of this particular model, where we can see that transitions between legumes and rapeseed are not allowed in this model.

Next, we must decide on which regressors to include. We will once again use rainfall, and for now we will use the difference between the profit margins of rapeseed and legumes, specifically $M_R - M_L$ for M as defined in equation (2.1). We use this as, if we are interested in the break crop, it could be the relative profitability that drives the choice of break crop. As legumes become relatively more profitable, we would expect to see a higher probability of transitions from wheat to legumes.

Figure 5.15a shows the transition probabilities for transitions from wheat in Anglia, as rainfall varies, for profit margin difference fixed at the historic mean (£79.20 per hectare). We can see that rainfall has little impact on light and medium soils. On heavy soils rainfall does have more of an effect – as rainfall increases, the transition probability of moving to rapeseed or other increases slightly, while the probability of repeated transitions to wheat decreases. If the autumn rainfall prevents the drilling of winter wheat, then the farmer may prefer to move to a spring sown break crop rather than spring wheat.

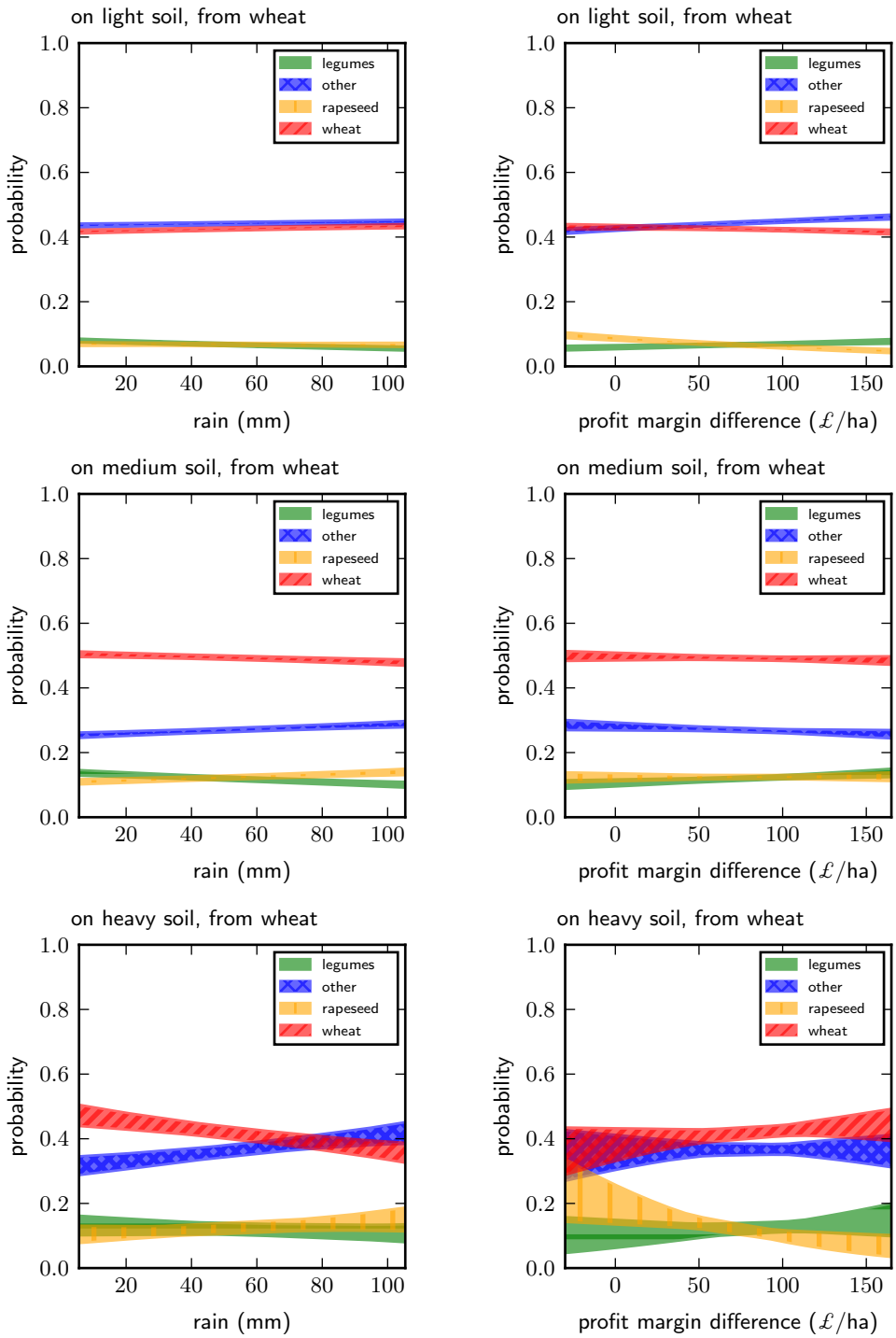
Of more interest is figure 5.15b, which shows the same region and transitions, but for varying profit margin difference, for rainfall fixed at the historic mean value

(53.4mm). A higher value of profit margin difference indicates that rapeseed is relatively more profitable than legumes. Thus, as profit margin difference increases, we might expect more rapeseed to be grown as the break crop from wheat. Figure 5.15b shows that this is not the case. Rather counterintuitively, we can see that on light and heavy soil, there is a definite decrease in rapeseed as profit margin difference increases, while on medium soil it seems to have little impact. Conversely, on all soil types, the probability for moving to legumes increases as legumes become relatively less profitable.

Figure 5.16a and figure 5.16b show the same analysis performed on the Mease region. Note that, as discussed in section 5.1.2, we omit transitions on light soil in Mease. This means that there are only 2 subfigures in each figure – one for medium soil, and one for heavy soil. The results are fairly similar as in Anglia, and still show the same counterintuitive behaviour with respect to the profit margin difference, although less dramatically.

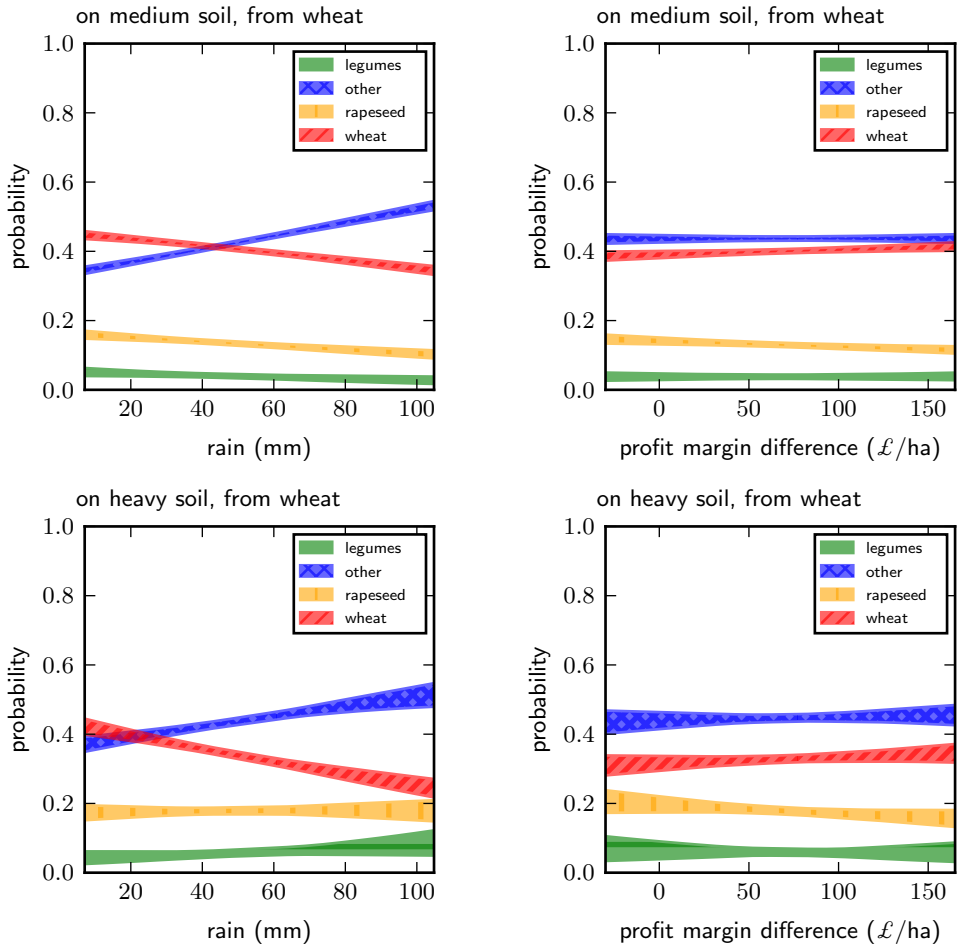
Clearly there is a something unusual happening in the modelling process. What this may be is open to debate. It is not simply the value of rainfall we fix at, as if we repeat the analysis with a different value of rainfall the same issue still arises. There may perhaps be some confounding issue with the model, or alternatively, it could be that farmers do not actually make their decision based on the profit margins of the break crops. If wheat is particularly profitable, then the profit margin of the break crop is of less importance. Alternatively, legumes are often linked into long-term contracts, as they need to be processed very quickly after harvesting. As such, farmers may be contractually obliged to grow them irrespective of the profit margin. This may be particularly important in East Anglia, which is the main region for pea production in the UK, and in fact the counter-intuitive behaviour is most dramatic in that region.

There may also be an issue with the profit data we used [74], which is actually the predicted profit for the next year. We used this as that is the information farmers have available when making their decision. However, these predictions may not be accurate. Also, the profit margin is subject to the cyclical nature of supply and demand. If the predicted profit margins are high, then more farmers may choose



(a) Varying rainfall (b) Varying profit margin difference

Figure 5.15: Transition probabilities from wheat for the break crop model with rainfall and profit margin difference as the regressors, for all soil types in Anglia. The left column fixes profit margin difference, and varies rainfall. The right column fixes rainfall and varies profit margin difference.



(a) Varying rainfall

(b) Varying profit margin difference

Figure 5.16: Transition probabilities from wheat for the break crop model with rainfall and profit margin difference as the regressors, for both soil types in Mease. The left column fixes profit margin difference, and varies rainfall. The right column fixes rainfall and varies profit margin difference.

to plant legumes, meaning the actual selling price will decrease. Clearly, our model will not pick this feature up, and this may cause, or perhaps simply contribute, to the unusual behaviour seen.

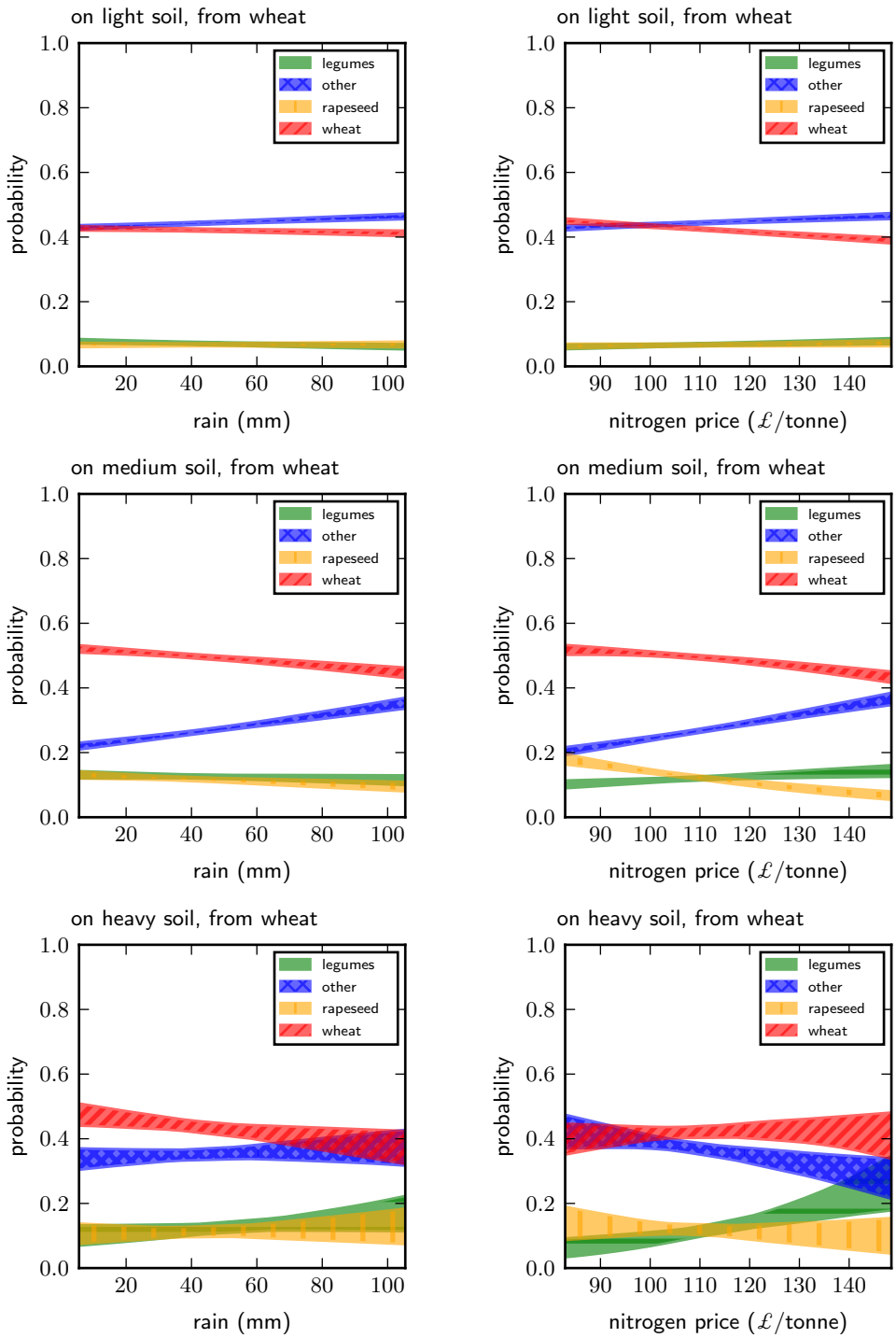
We do have another economic variable we could include – nitrogen price. Of course, the profit margin will include the price of nitrogen in some sense, since it is a major component of many fertilisers. However, analysing it separately may give more intuitive results. We will use the same group of crops, and analyse them with nitrogen price and rainfall as the regressors. Figure 5.17a and figure 5.17b show the results for transitions from wheat in Anglia. Results in Mease are reasonably similar, so we omit them for space saving purposes.

As nitrogen price increases, we would expect more legumes to be grown, as legumes fix their own nitrogen from the air and therefore require little or no nitrogen based fertiliser to be used. These results are far more intuitive. They show that as nitrogen price increases, we do see an increase in transitions to legumes, and a decrease in transitions to rapeseed, particularly on medium and heavy soils. Therefore, from here on in, analyses will be performed with nitrogen price as our economic variable.

5.2.3 Including repeated wheat

Farmers often grow different crops in successive years, primarily to help prevent the build up of pests and diseases (see section 2.1). However, one crop which farmers do frequently grow in succession is wheat. Planting the same crop can be easier for the farmer in terms of planning and equipment needed. While the yield and therefore the profit decreases with a second planting of a crop, even the reduced profit associated with a second growing of wheat can be appealing to a farmer. We can include repeated wheat in our model as a separate state. Transitions from a first planting of wheat to a second planting of wheat are fairly common, but second wheat to third wheat, third wheat to fourth wheat, and so on, are increasingly rare, since the yield decreases further with each repeated planting.

Due to this behaviour, we will include one state – repeated wheat – to capture all instances of wheat which are not a first planting. Only certain transitions are



(a) Varying rainfall

(b) Varying nitrogen price

Figure 5.17: Transition probabilities from wheat for the break crop model with rainfall and nitrogen price as the regressors, for all soil types in Anglia. The left column fixes nitrogen price, and varies rainfall. The right column fixes rainfall and varies nitrogen price.

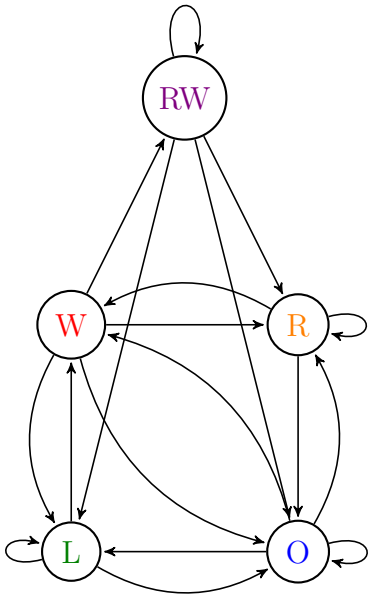
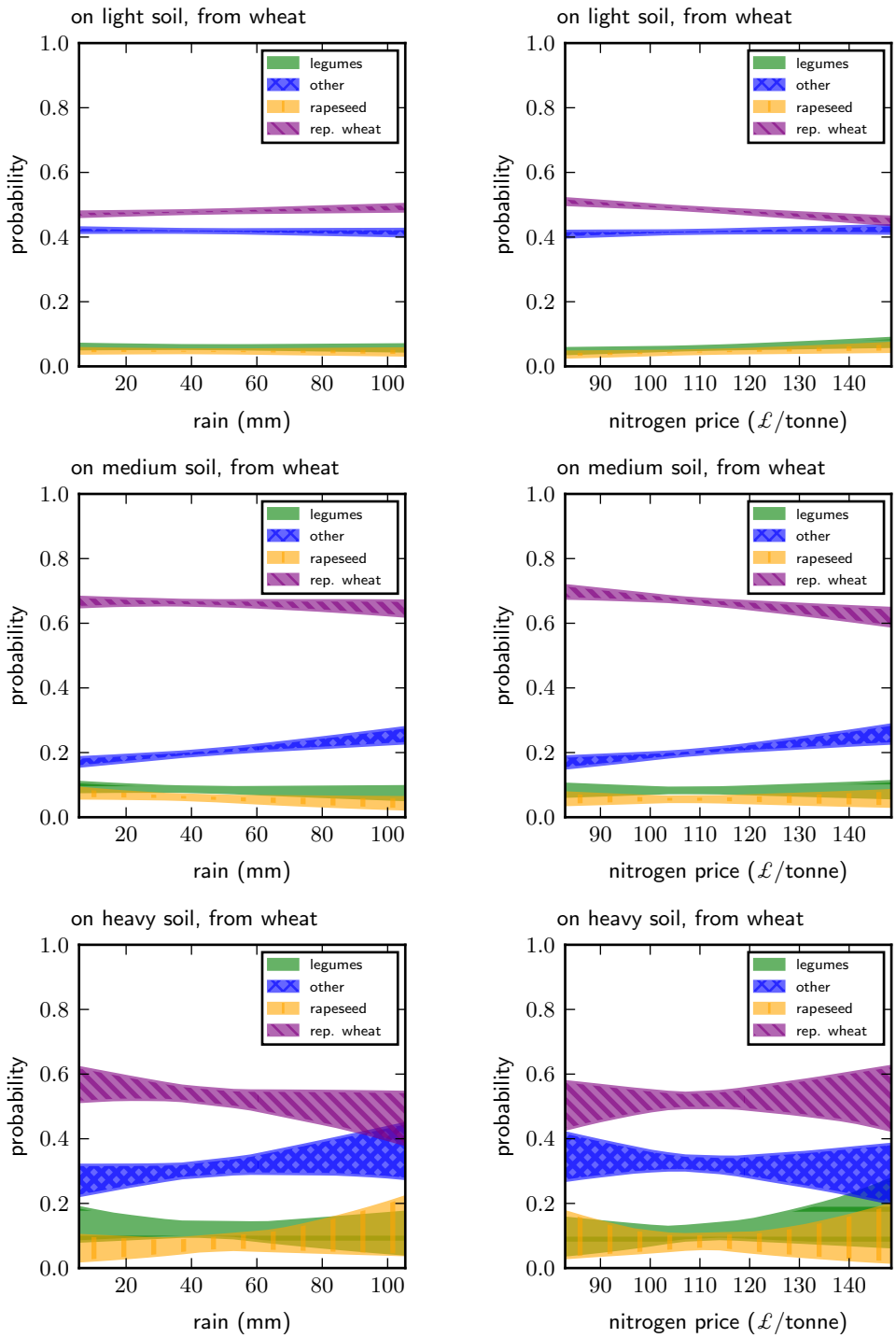


Figure 5.18: The state transition diagram for the break crop model when we include repeated wheat, where W is wheat, RW is repeated wheat, R is rapeseed, L is legumes, and O is other. Transitions between R and L do not occur in practice so have been excluded from the model. Transitions to RW only occur from W, and we can not move from RW to W. We also cannot move from W to W.

allowed in this model – we can not go from wheat to wheat, as this second growing of wheat is considered as a different crop. Furthermore, we can not move into the state repeated wheat from any state other than wheat. See figure 5.18 for the state transition diagram for this model.

As mentioned in section 5.1.3, when growing the same crop year after year, yields decrease. However, we will fit the model with nitrogen price, as it gives us more intuitive results with regards to the break crop, which is something of interest to land use modellers. By using nitrogen price as a regressor in our model, we do not directly investigate this information, but the model learns from the data that farmers do not often repeatedly grow wheat past a second successive planting.

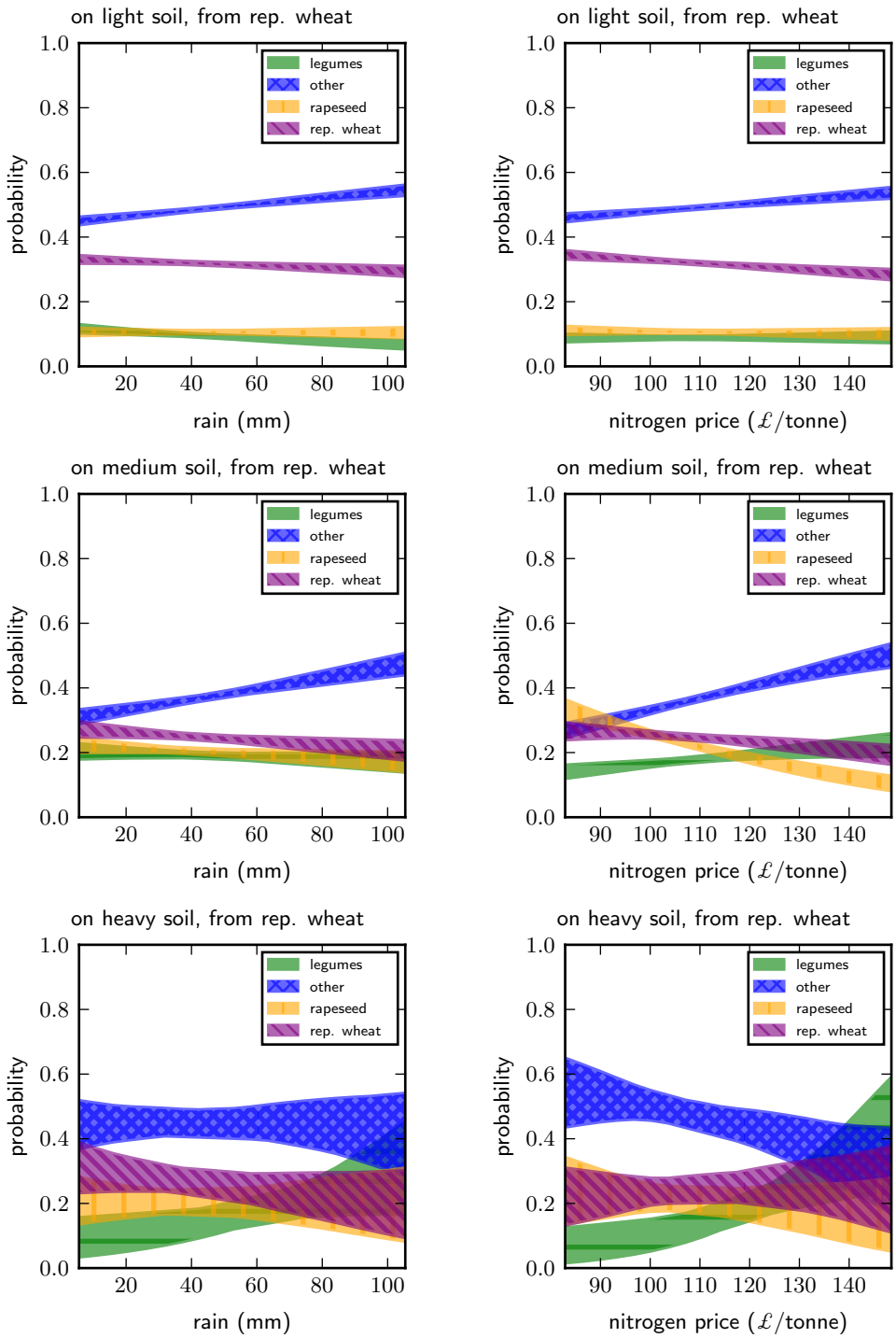
Figure 5.19 shows the results for transitions from wheat and figure 5.20 shows the results for transitions from repeated wheat in Anglia. The results from wheat, shown in figure 5.19, are fairly similar to the results from the standard break crop



(a) Varying rainfall

(b) Varying nitrogen price

Figure 5.19: Transition probabilities from wheat for the break crop model, with repeated wheat, with rainfall and nitrogen price as the regressors, for all soil types in Anglia. The left column fixes nitrogen price, and varies rainfall. The right column fixes rainfall and varies nitrogen price.



(a) Varying rainfall

(b) Varying nitrogen price

Figure 5.20: Transition probabilities from repeated wheat for the break crop model, with repeated wheat, with rainfall and nitrogen price as the regressors, for all soil types in Anglia. The left column fixes nitrogen price, and varies rainfall. The right column fixes rainfall and varies nitrogen price.

model (figure 5.17), which is to be expected, since they are plotting very similar events. Most repeated wheat transitions are from a first planting of the crop to a second planting. Subsequent repetitions are much rarer in comparison. The standard break crop model, shown in (figure 5.17), includes all repeated wheat transitions, whereas the results from wheat in the modified break crop model only capture the first consecutive growing. Any subsequent repetitions are counted as repeated wheat to repeated wheat, and are shown in figure 5.20.

From figure 5.20 we can also see that the probability of growing legumes after repeated wheat increases dramatically as both rainfall and nitrogen price increase, on medium and heavy soil. This behaviour is not seen for transitions from first wheat. Furthermore, we can see that, from repeated wheat, as nitrogen price increases, the probability of moving to rapeseed decreases too, particularly on medium soil. This, alongside the increased probability of growing legumes, implies that the break crop from repeated wheat is very dependant on nitrogen price. This could be because successive wheat plantings takes large amounts of nutrients out of the soil. While rapeseed needs a great deal of nitrogen based fertiliser, legumes do not (in fact, they return nitrogen to the soil), so in times of high nitrogen prices, legumes may be very appealing. Figures 5.19 and 5.20 show that we should separate first wheat from repeated wheat in our model, as the behaviour is rather different.

We must note here that all the inferences we have presented in section 5.2 must be made under the knowledge that our data stops in 2004. As such, care must be taken when drawing conclusions about current farming practices. However, the results show what the model can do, and that, on the whole, it seems to produce sensible results.

5.3 Model validation

When building a statistical model it is important to perform some checks to see if the assumed relationship is actually reasonably close to what the data suggests it is. The next three subsections present three different approaches to model validation: a visual approach (section 5.3.1), an approach based on cross validation (section 5.3.2),

and a comparison with a covariate free model (section 5.3.3). In the previous section we discussed that it was important to include repeated wheat in the model, however, for ease of presentation we will perform all three verifications on the break crop model without repeated wheat, with rainfall and nitrogen price as the regressors.

5.3.1 A non-parametric estimate

We can compare our parametric estimates of the transition probabilities $\pi_{ij}(x)$ with a non-parametric estimate. This will give us some idea of whether our parametric assumptions are reasonable. A simple non-parametric estimate of $\pi_{ij}(x)$ consists of taking a weighted average of the observations around x:

$$\tilde{\pi}_{ij}(x) := \frac{\sum_{x' \in \mathcal{X}} w(x - x') k_{ij}(x')}{\sum_{x' \in \mathcal{X}} w(x - x') n_i(x')} \tag{5.1}$$

where w is some suitably chosen kernel, that is, a non-negative symmetrical function centred around the origin. A key choice in this function is the so-called bandwidth, which quantifies the smoothness of the estimate. Here, we take a multivariate Gaussian kernel:

$$w(x) := |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right)$$

with

$$\Sigma^2 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 20^2 & 0 \\ 0 & 0 & 20^2 \end{bmatrix}$$

Note that the first component of x is always taken to be the constant 1, hence only the lower right 2×2 submatrix of Σ^2 is relevant. The choice of 20 for both components was done by trial and error to get sufficiently smooth estimates. Figures 5.21 and 5.22 depict $\tilde{\pi}_{ij}(x)$ as calculated from equation (5.1) and the parametric transition probability bounds $[\hat{\underline{\pi}}_{ij}(x), \hat{\overline{\pi}}_{ij}(x)]$ as calculated from equations (4.6) and (4.7), for all cases of previous crop i and soil type in Anglia. We plot the probabilities as functions of nitrogen price, while fixing rainfall in September at the mean value for Anglia (53.4mm). We can see that our model predictions and the non-parametric estimates coincide quite well. The most notable differences are located at the extremes of our observed nitrogen data. The worst fits are observed from wheat on

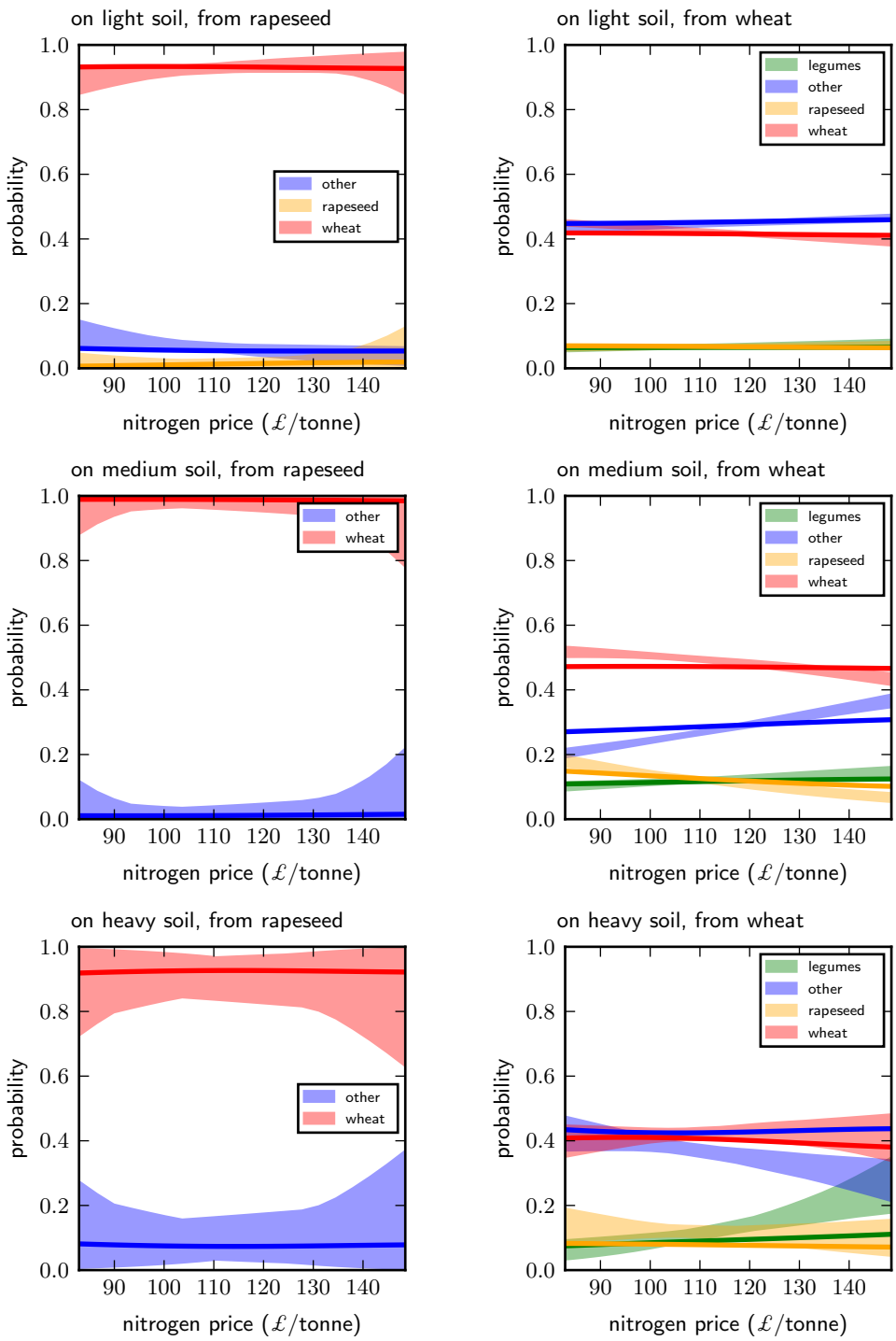


Figure 5.21: Parametric (shaded region) and non-parametric (solid line) estimates of the transition probabilities for all soil types in Anglia, from rapeseed and wheat, as nitrogen price varies, for rainfall in September fixed at the mean

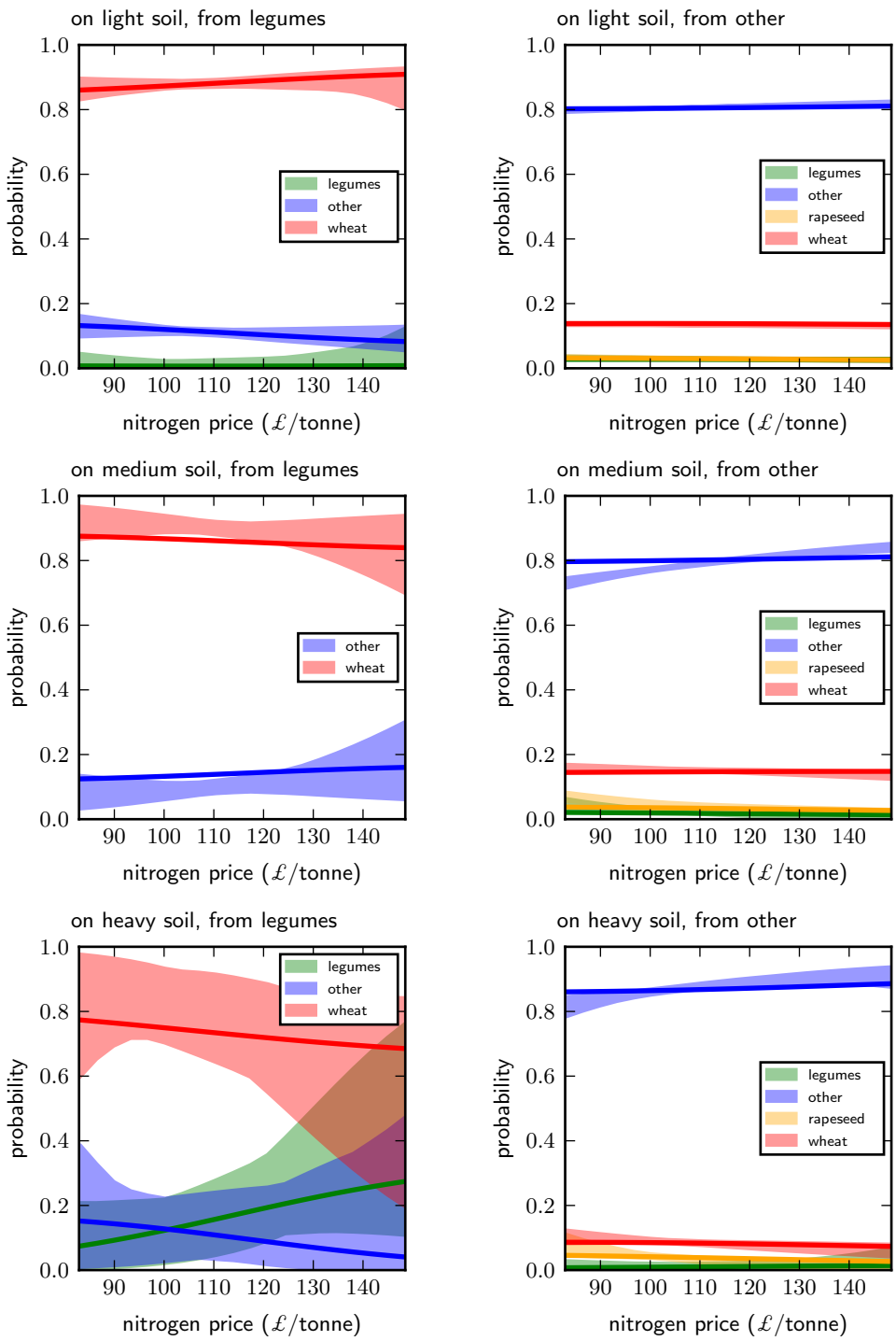


Figure 5.22: Parametric (shaded region) and non-parametric (solid line) estimates of the transition probabilities for all soil types in Anglia, from legumes and other, as nitrogen price varies, for rainfall in September fixed at the mean

medium and heavy soil (i.e. the bottom 2 plots in the right column of figure 5.21), where the robust Bayesian model seems to slightly overestimate the slopes of the curves. We attempted modifying the bandwidth specified in equation (5.1), but no better fit was found than that shown in figure 5.21. One potential reason for this is that there is a reasonably strong correlation between yearly global nitrogen price and rainfall in Anglia in September: -0.624. Since we are only looking at a particular slice of the model in these plots, we may be seeing an artefact of this result. An alternative hypothesis is that the MAP approximation for our posterior model parameter does not perform well in this specific situation.

Furthermore, we can also compare the transition probabilities from both the non-parametric model and our robust Bayesian model with the probabilities produced from a standard maximum likelihood estimate of β. The results of this analysis are shown in figures 5.23 and 5.24, where the maximum likelihood estimate is shown as a dotted line. We can see that the transition probabilities from the maximum lkelihood analysis quite closely resemble the probabilities from the MAP approximation. This is not particularly surprising since, as mentioned in section 4.3.2, the MAP estimate corresponds to the maximum likelihood estimate where the data has been augmented with pseudo counts.

Figure 5.25 shows a smoothed version of $n_i(x)$, that is:

$$\sum_{x' \in \mathcal{X}} w(x - x') n_i(x') / w(0)$$

These plots give an idea of the size of the denominator in equation (5.1) and thereby how much data is near each point x. The lowest data densities are observed from legumes on heavy soil type, where the average number of observations lies belows 10. The highest data density is observed from other on light soil type, where we see numbers between 1000 and 2200. This difference in data density is well reflected in the robust Bayesian estimates. The data density decreases substantially as nitrogen price increases, and interestingly our robust Bayesian intervals also become wider in this direction, as desired: we built a robust Bayesian model to capture exactly this sort of feature.

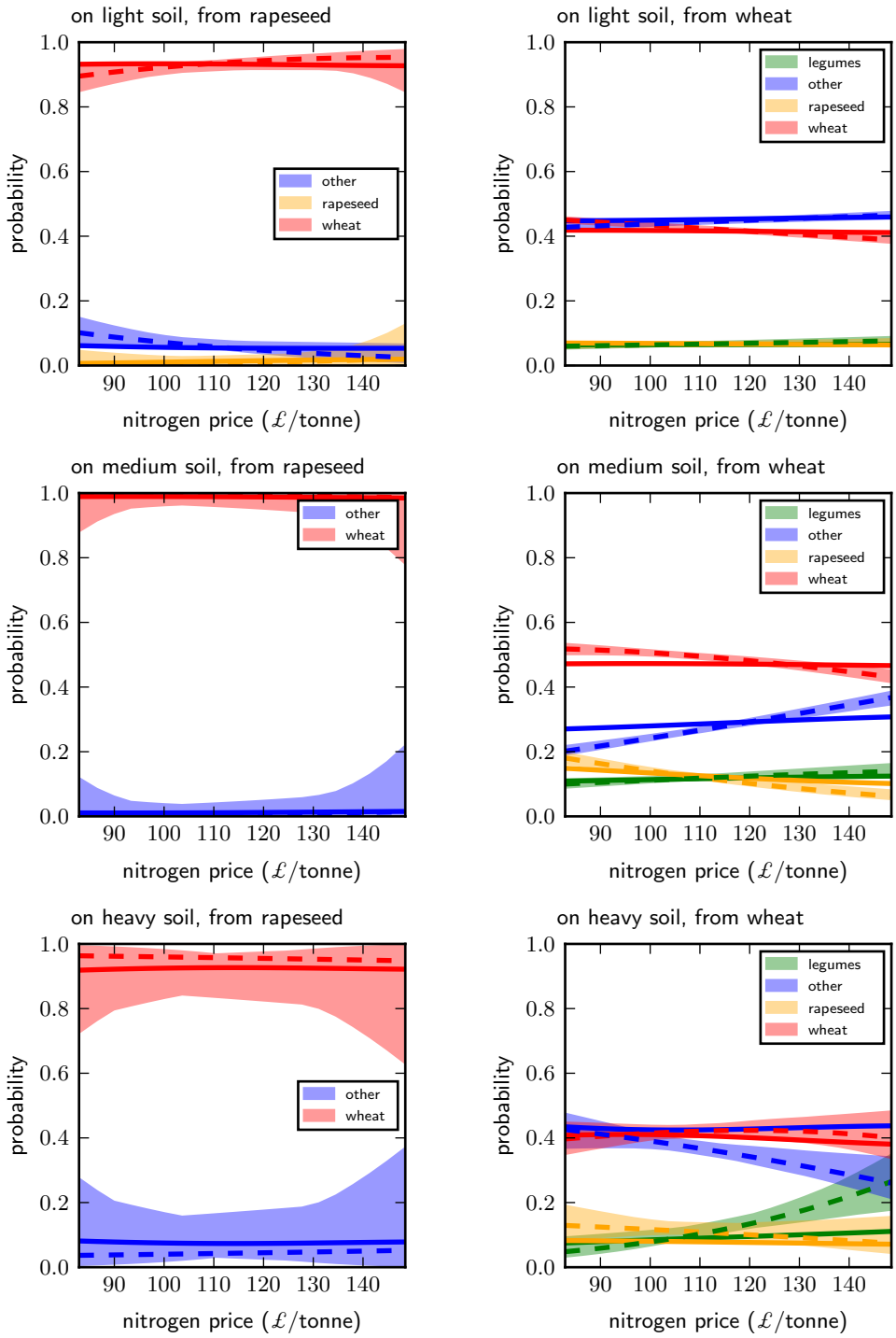


Figure 5.23: Parametric (shaded region), maximum likelihood (dashed line) and non-parametric (solid line) estimates of the transition probabilities for all soil types in Anglia, from rapeseed and wheat, as nitrogen price varies, for rainfall in September fixed at the mean

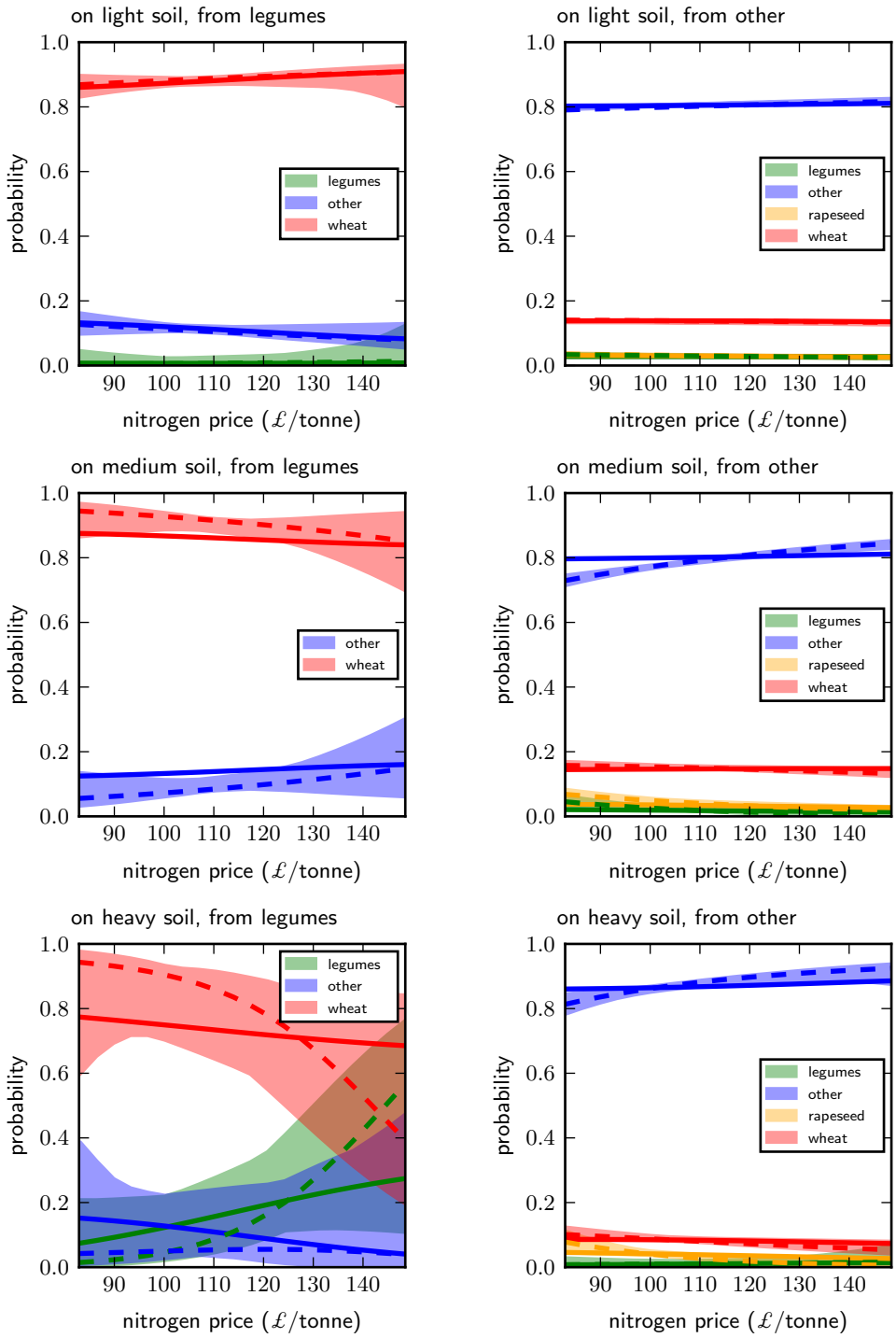


Figure 5.24: Parametric (shaded region), maximum likelihood (dashed line) and non-parametric (solid line) estimates of the transition probabilities for all soil types in Anglia, from legumes and other, as nitrogen price varies, for rainfall in September fixed at the mean

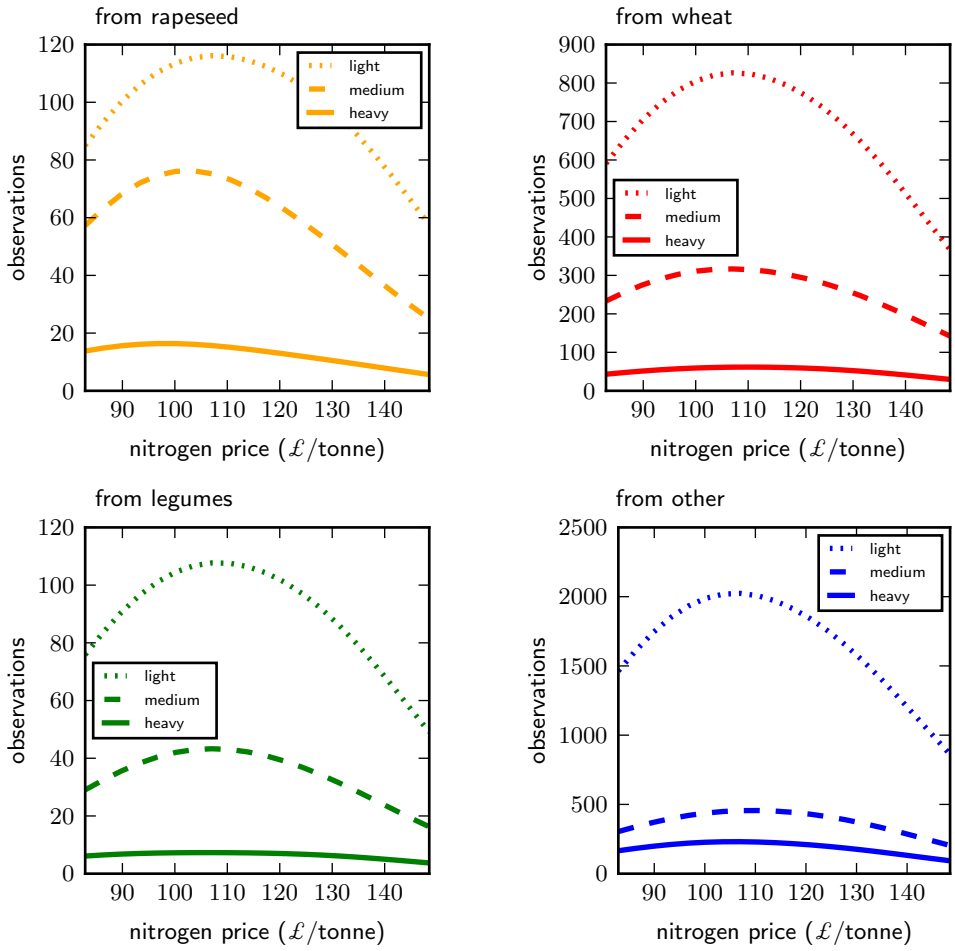


Figure 5.25: Smoothed $n_i(x)$ as a function of nitrogen price (x), in Anglia, for each soil type. This tells us how much data is near each point x, and gives an idea of the accuracy of the non-parametric estimates plotted in figures 5.21 and 5.22.

5.3.2 Cross validation

A typical method for validating classifiers is to split the data into training and test data, and then to compare the predicted class (or set of classes) from the model based on the training data, with the actual classes in the test data. We can consider our model as a classifier, in the following sense: we compare the farmer's actual choice with the most likely predicted crop. For example, for the predictions in figures 5.21 and 5.22, for that particular value of rainfall, the most likely crop from other is other, wheat from legumes and rapeseed, and either other or wheat from wheat, depending on nitrogen price. Of course, in the test data, rainfall will vary as well; figures 5.21 and 5.22 just show a particular slice of the model. Note that our model sometimes produces a set of most likely crops, as we do a sensitivity analysis over all $\beta^* \in B^*$.

For credal classification, there are a number of performance measures [26] we can use to see how a classifier is performing. The *determinacy* is the percentage of classifications where the output class is unique. The *single accuracy* is then the accuracy of those predictions. The *indeterminate output size* is the average number of classes when the output class is not unique. Finally, the *set-accuracy* is the percentage of times an indeterminate set contains the correct classification.

To ensure that all the data is used for testing the analysis is typically repeated, say 10 times, by splitting the original data set into 10 parts, and then repeatedly testing on each of these parts, based on training on the complement of the testing data. We can use a similar approach to validate our model. We have two distinct geographical regions in our dataset. We can perform cross validation within each region, and also combine the two regions together and perform cross validation on the entire data set.

The results are presented in table 5.3. Of most interest are the single accuracy and the set accuracy. The single accuracy is quite excellent: the model predicts the correct crop in 70–75% of the cases. The set accuracy is even better, around 80–90%. Finally, we note that the set accuracy is at its lowest for the full data. A logical explanation for this is that the regions are geographically quite distinct. Even despite these differences, the model copes well.

region	determinacy	single accuracy	indeterminate output size	set accuracy
Anglia	0.968	0.722	2.008	0.855
Mease	0.988	0.758	2.140	0.929
All	0.976	0.734	2	0.795

Table 5.3: Cross validation results for East Anglia, Mease/Sence, and the combined data set. Determinacy is the percentage of classifications where the output class is unique, and the single accuracy is the accuracy of those predictions. The indeterminate output size is the average number of classes when the output class is not unique, and the set-accuracy is the accuracy of these predictions.

We note that the determinacy is quite high as a result of the large data set used. This is mostly due to the fact that there is a clear dominant crop type for most combinations of soil and previous crop growing. In order to analyse this, we can split our analysis by previous crop grown. We would expect transitions from wheat to have lower single accuracy – there is much less dominance in what is grown following wheat, meaning it is harder to predict. Table 5.4 shows results split by previous crop type (and region), and we can see that our suspicions are correct. Note that, when determinacy is 1, our model never outputs a set of predictions, therefore we do not have an indeterminate output size. This happens only from rapeseed, and is probably due to wheat being such a dominant crop following rapeseed (see figure 5.21).

5.3.3 Comparison with the imprecise Dirichlet model

In order to assess the predictive power of the covariates we can compare our model with a more general model. We will use a multinomial model with an imprecise Dirichlet model [100] prior, where we use no covariates.

We know from equation (3.9) that a multinomial model would have probability

region	previous crop	determinacy	single accuracy	indeterminate output size	set accuracy
Anglia	wheat	0.919	0.454	2	0.882
	legumes	0.984	0.895	2.217	1.000
	rapeseed	1	0.951	-	-
Mease	wheat	0.990	0.448	2	0.995
	legumes	0.969	0.917	2.133	0.900
	rapeseed	1	0.954	-	-

Table 5.4: Cross validation results for East Anglia, Mease/Sence, and the combined data set, split by previous crop grown

of success for category j of:

$$\pi_{ij}^{\beta}(x) = \frac{\exp(\beta_{ij}x)}{\sum_{h=1}^{J}\exp(\beta_{ih}x)}$$

However, when we include no covariates (i.e $x = 1$) this simply reduces down to

$$\pi_{ij}^{\beta}(1) \equiv \theta_{ij} := \frac{\exp(\beta_{ij0})}{\sum_{h=1}^{J}\exp(\beta_{ih0})}$$

where $\sum_{j=1}^{J}\theta_{ij} = 1$. It is well known that the conjugate prior distribution of a multinomial model is a Dirichlet prior distribution. We can specify a Dirichlet prior distribution for θ_i [100, Section 2.2]:

$$f_0(\theta_i|s_0, t_0) \propto \prod_{j=1}^{J}\theta_{ij}^{st_j-1}$$

The imprecise Dirichlet model is then defined [100, Section 2.3] as the set of all these prior distributions for a fixed $s_0 = s$, and all $0 < t_j < 1$. In order to remain consistent with Walley, we will choose $s = 2$.

Due to the amount of observations in our data set, this multinomial model always predicts a single crop type. In regions where data is abundant, the logistic model also outputs a single predicted crop, and the models perform similarly (around 73% accuracy in both cases). However, in regions where the data is sparse and where therefore the logistic model produces a set of predictions, the logistic model has 84%

set accuracy, whereas the multinomial model has only 43% accuracy. This shows the benefits of our logistic model in regions of sparse data.

Note that this method for validation assesses only whether the farmer grows the most likely predicted crop. If this is what we are interested in then, in regions where there is abundant data, the multinomial model is preferable: it produces similar performance (as generally one crop dominates the others), but it is a much simpler model. However, we are interested in understanding the drivers behind farmer's crop choices, and obviously the multinomial model cannot capture this, unlike the logistic model. Consequently, it seems the traditional classification performance measures are not entirely suitable to assess model performance. This also raises an interesting question in how classification performance measures could be adapted to capture model performance not only related to the most likely predicted class.

5.4 Chapter summary

This chapter has focussed primarily on the application of the model to the data sets that we have access to. In section 5.1 we introduced the data, and performed some initial exploratory analysis. We have field level data for crop choice in two distinct regions in England for the years 1993–2004 inclusive. We also have the soil type of each of these fields. Furthermore, we have various data sets for external influencing variables on this crop choice, including two separate economic data sets; profit margin and nitrogen price.

We also introduced our rainfall data. Within this we introduced the basic theory of kriging. This is an interpolation method which can be used to predict the rainfall in each of our fields. However, due to the reasonably small size of our regions, we chose to keep rainfall fixed across each of our NCAs, although the theory presented could be used if we had larger regions of data.

In section 5.2 we used the data to fit our robust Bayesian land use model, and analysed the impact of the regressors on the posterior transition probabilities. In order to do this, we needed to specify which crop types and which regressors were to be included. We performed our analysis on four different model specifications.

Firstly, we carried out a simple proof of concept, involving 3 crop choices and one regressor. From then on, we fitted models of real-world interest, specifically relating to the break crop from wheat.

We included wheat, legumes, rapeseed, and other in our model. Legumes and rapeseed are traditional break crops from wheat, and there is currently an interest in increasing the amount of legumes being grown. To begin with, we used rainfall and the profit margin difference between legumes and rapeseed, with the expectation that as legumes became relatively more profitable in comparison to rapeseed, we would see a higher probability of growing legumes after growing wheat. However, the model actually showed the opposite happening. We discussed various hypotheses as to what may have caused this, including a confounding variable in the model, the format of the data, contractual situations of the farms, or that the profit margin of the break crops are simply not important to the farmer. However, we do not know for definite what caused this issue. If we had access to the observed profit margins, rather than the predicted margins, then we could investigate whether it is an issue with the structure of the data.

In an attempt to resolve this issue, we fitted a model containing the same crops, but with rainfall and nitrogen price as our regressors (nitrogen being a key ingredient of many fertilisers). Legumes require less fertiliser than rapeseed, so as nitrogen prices increase, we would expect to see a higher probability of legumes being grown. The model showed this to be the case.

Into this nitrogen based model, we then added another crop type – repeated wheat. Wheat is generally the only crop that farmers grow in successive years, so is an interesting special case to consider. The results showed that transitions from repeated wheat to repeated wheat were quite rare, as suspected. They also showed that the break crop from repeated wheat, on medium and heavy soil, was very sensitive to nitrogen price. This is possibly because successive plantings of wheat takes large amounts of nutrients out of the soil, meaning crops which are dependent on fertilisers, such as rapeseed, will require much more fertiliser.

Finally, in section 5.3 we performed three model validation procedures. The first was a comparison with a (novel) non-parametric estimate of the transition

probabilities. The second was a method based on cross-validation. The third was a comparison with a covariate free multinomial model with an imprecise Dirichlet prior. All three approaches indicated our model performed well, particularly in regions of sparse observations.

The results in this chapter focussed on the modelling the transition probabilities of the stochastic process that is the foundation of our land use model. The following chapter will discuss how we can use this process to make predictions about future crop distributions.

Chapter 6

Future crop distributions

We have an imprecise stochastic process built on precise historic data. In section 5.2 we analysed how the transition probabilities of this process depended on external factors. What we now want to do is predict future crop distributions, or more specifically, bounds on the future crop distributions. The mathematical theory for calculating these bounds was presented in section 4.4. In order to make predictions of crop distributions, we require predictions of the regressors we include in our land use model. In this chapter, we will discuss how we can obtain these predictions, and what we can do with them. In section 6.1 we will discuss a comparison with the actual observed crop distributions. Section 6.2 will explore a hypothetical future scenario, which is the type of analysis which could be useful to agricultural scientists and policy makers. In section 6.3 we will briefly discuss some existing models which predict future values of our regressors.

6.1 Comparison with observed crop distributions

In order to predict future crop distributions, we require future values of the regressors in our model. However, as discussed in section 5.1.1, our crop data stops in 2004. This means that we can use the actual observed values of rainfall and nitrogen price to make a "prediction" of the crop distributions for 2005 onwards. Figure 6.1 shows the results for such an analysis in Anglia, where we have predicted the crop distributions on medium soil for the next 5 years.

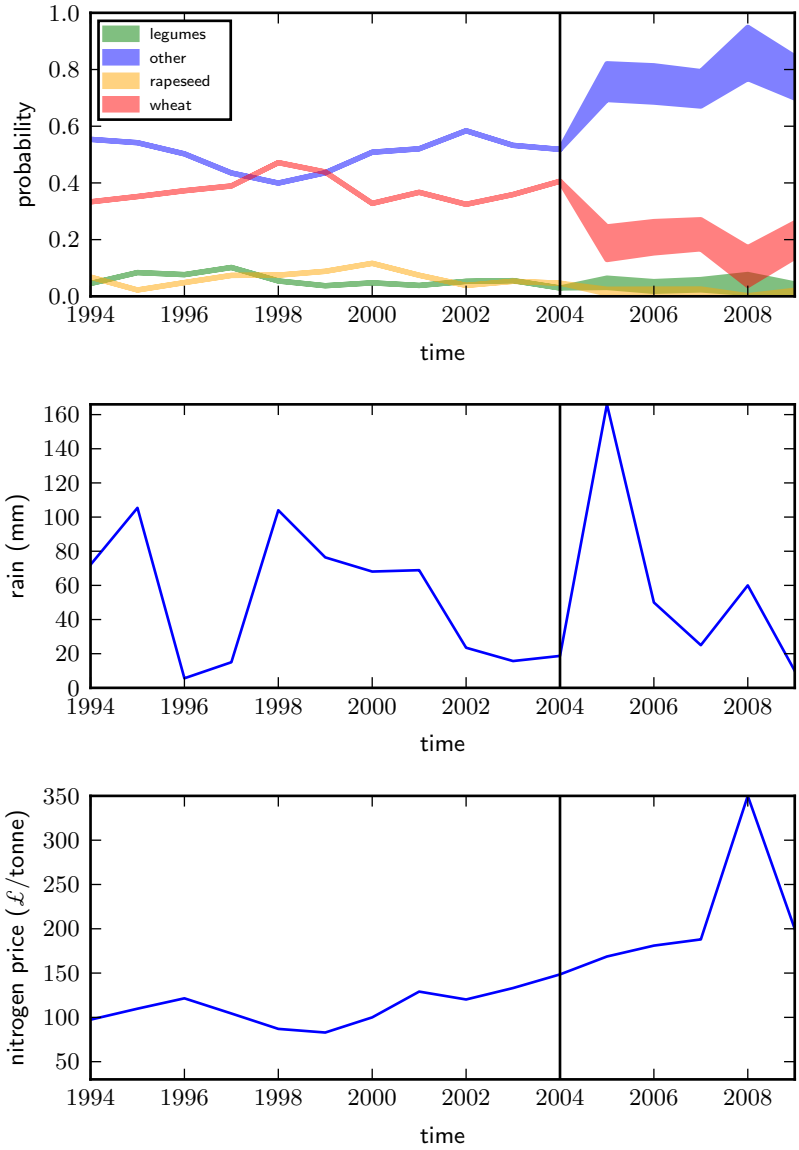


Figure 6.1: A prediction of crop distributions in 2005–2009, in Anglia on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–2004

Everything to the left of the vertical black line is the data that the model is built on. On the right hand side of the line is the "future", in this case the predicted crop distributions for 2005–2009, given the actual observed values of rainfall and nitrogen price for those years. We can see that in 2005 rainfall was dramatically higher than previously, but after that it was quite low. Nitrogen price showed a steady rise, until 2008 when it increased a lot, before returning to a more expected price in 2009. The effect this has on the predicted crop distributions is generally a decrease in wheat, courtesy of an increase of other. One thing to note here is that the future values of nitrogen price and also one future value of rainfall were outside the range of values that the model is built on, so we should be wary when drawing inferences from this.

Ideally, what we would now do is a comparison between these results and the actual observed crop distributions for 2005–2009. This would give us another indication as to how well our model is performing. Unfortunately, this data is not available, as collection of field level data stopped in 2004. However, what we can do is build our model on part of the data, say observations until 1999, and then predict the remainder of the data set, which would be the year 2000 onwards.

Figures 6.2 to 6.6 show the results for both of our regions of study, and all soil types. The dashed lines represent the actual crop distributions, while the shaded regions show the predicted bounds from our model. We can see that generally the model captures the farmer behaviour quite well, particularly when considering that we are not building the model on much data (we are only building the model on half of our data set), and that nitrogen price in the "predictions" goes outside the range of values our model is built on. Of course, the model is designed to capture uncertainty due to a lack of data, but we designed the model to handle a lack of crop data and a lack of prior information, rather than limited amounts of time dependent data such as nitrogen price and rainfall. The occasional failures of the model in capturing the "future" may be caused by a lack of this time dependent data, or it could actually be a fault in the model. Unfortunately, as mentioned above, the field level data we would need to test for this is no longer collected, so such an analysis is not possible.

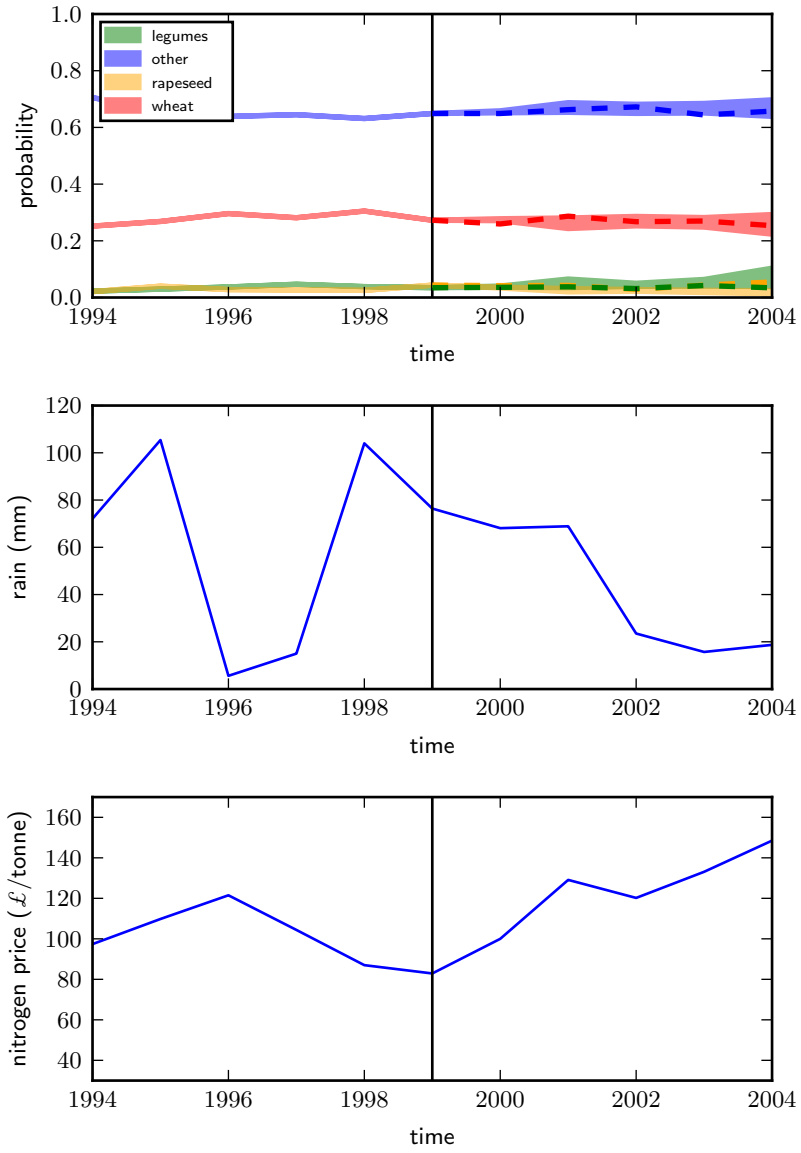


Figure 6.2: A prediction of crop distributions in 2000–2004, in Anglia on light soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999. This prediction is shown as the shaded region in the top plot, while the dashed lines show the actual crop distributions for 2000–2004.

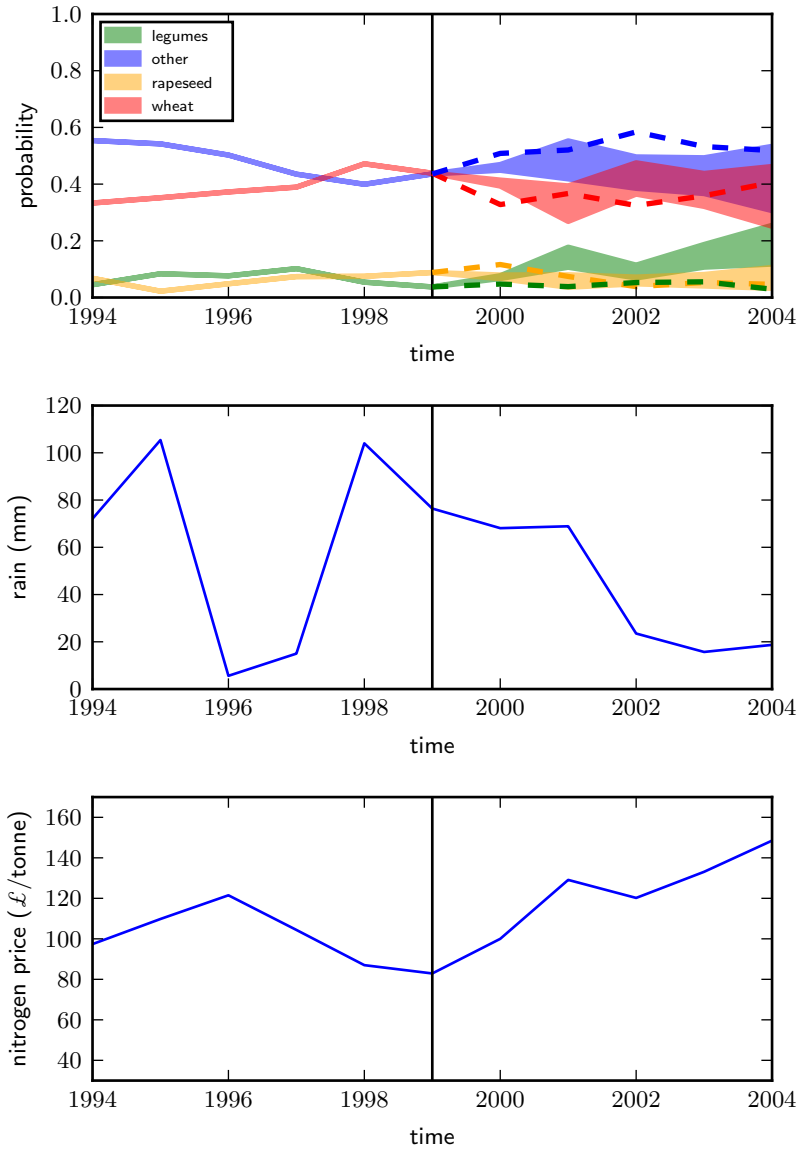


Figure 6.3: A prediction of crop distributions in 2000–2004, in Anglia on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999. This prediction is shown as the shaded region in the top plot, while the dashed lines show the actual crop distributions for 2000–2004.

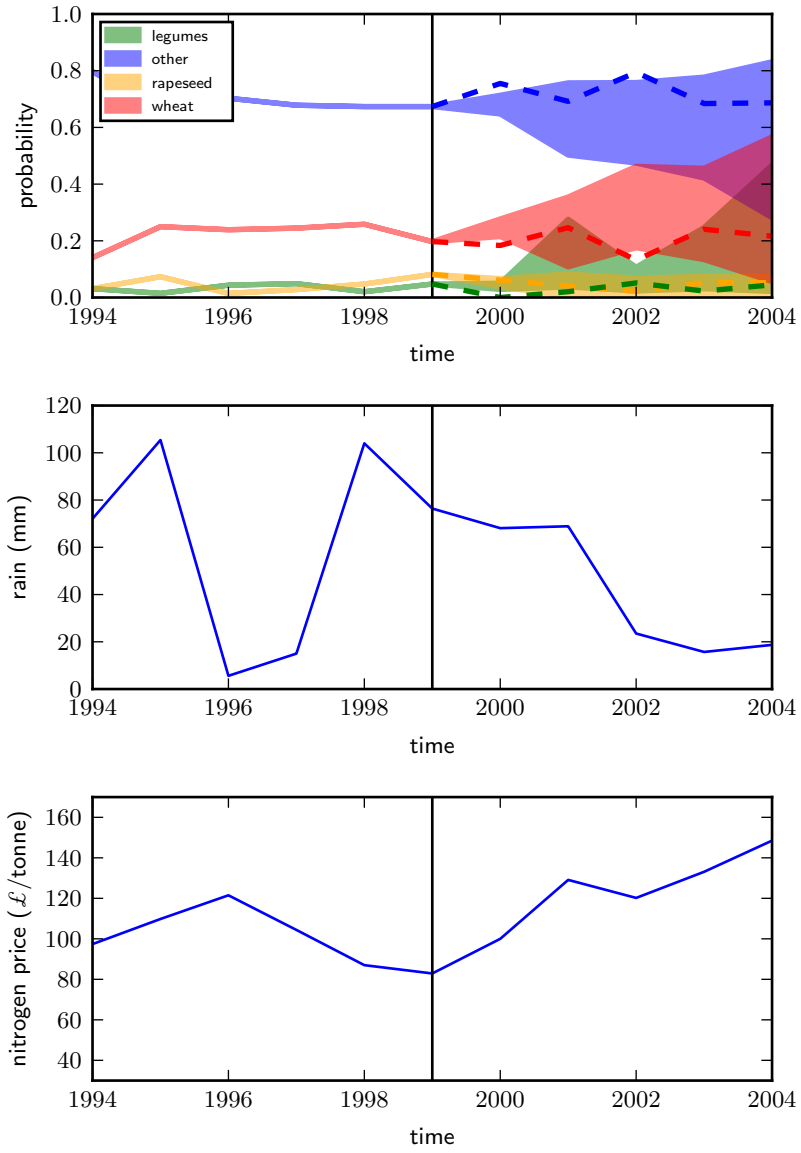


Figure 6.4: A prediction of crop distributions in 2000–2004, in Anglia on heavy soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999. This prediction is shown as the shaded region in the top plot, while the dashed lines show the actual crop distributions for 2000–2004.

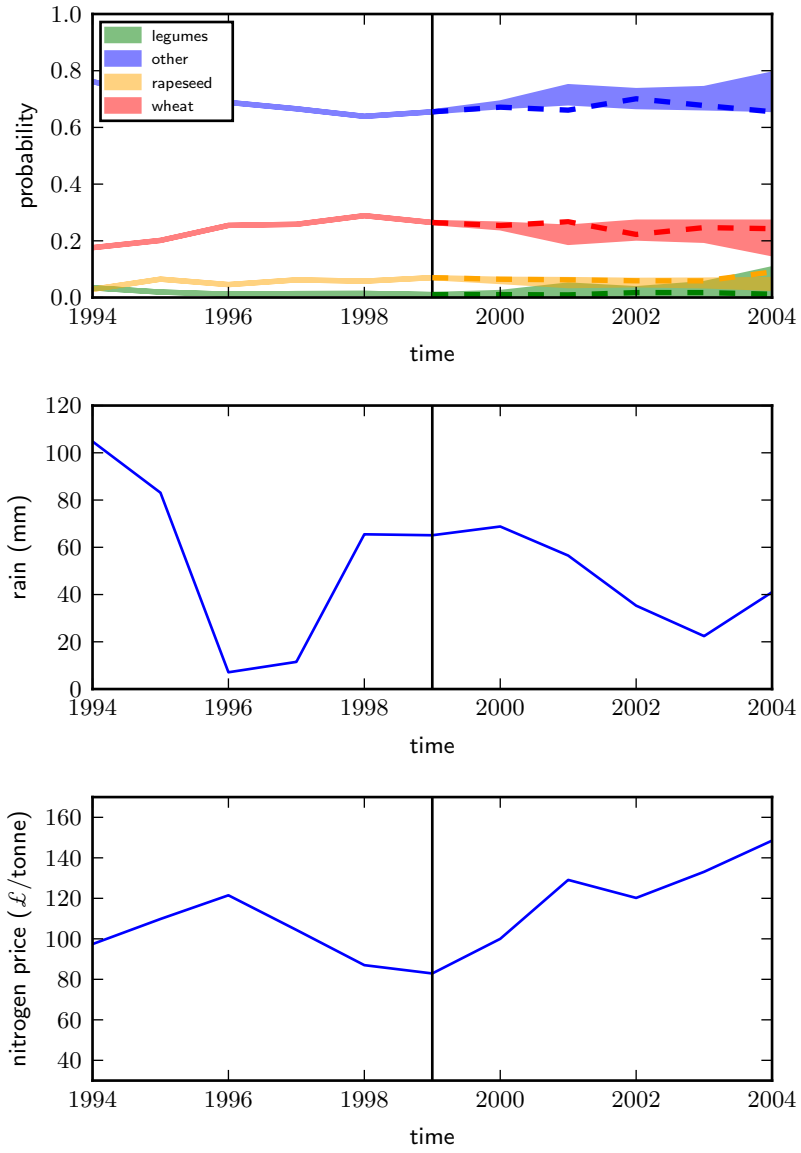


Figure 6.5: A prediction of crop distributions in 2000–2004, in Mease/Sence on medium soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999. This prediction is shown as the shaded region in the top plot, while the dashed lines show the actual crop distributions for 2000–2004.

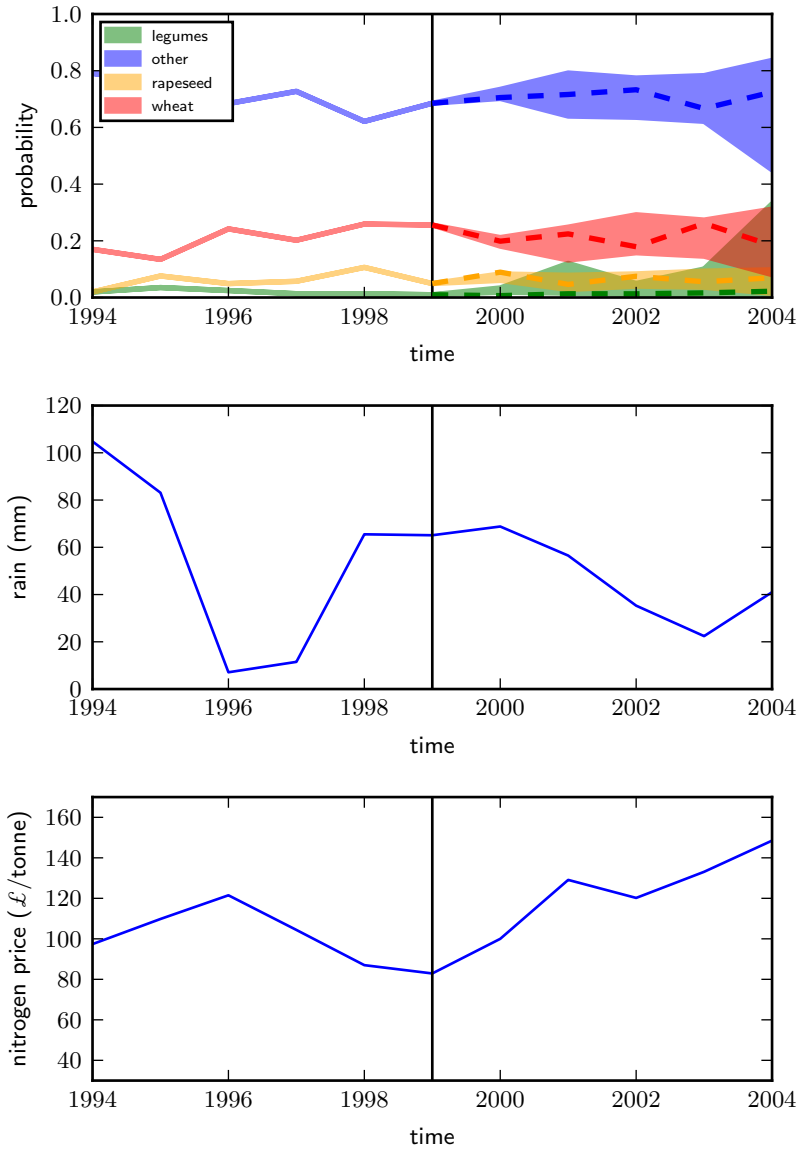


Figure 6.6: A prediction of crop distributions in 2000–2004, in Mease/Sence on heavy soil, given the actual observed rainfall and nitrogen price for that time period, and the observed crop distributions for 1994–1999. This prediction is shown as the shaded region in the top plot, while the dashed lines show the actual crop distributions for 2000–2004.

The worst performance is seen in Anglia on medium soil (figure 6.3). We must note that the failures in figure 6.3 occur before the nitrogen price goes out of range. The reasons for this are unclear, but one theory is that at the time the prediction starts, the probability of growing wheat and other is very similar, and perhaps this causes some confusion within the model. Once nitrogen price does go out of range, the model also seems to overestimate the probability of growing legumes. This is possibly because our model says that as nitrogen price increases, we expect to see more legumes grown. As our future values for nitrogen price are higher than the data that the model was trained on, it is understandable that we predict a higher amount of legumes than were actually grown, as the model will simply extrapolate the behaviour that it has seen before.

What we have done in this section is effectively another approach to model validation, this time checking how well our model predicts the future crop distributions. However, another use of models such as ours is to analyse what could happen given some future economic or climatic changes. The following section will investigate this.

6.2 A hypothetical example

A common application of land use models is to investigate the possible impact of a hypothetical scenario. There may be an interest in, for example, the impact that a rise in nitrogen prices might have on future crop distributions. What we do is select future values of the regressors in our model which constitutes a scenario of interest, and then see what the model predicts the future crop distributions will look like.

Figure 6.7 shows one such scenario. Here, we have assumed that "future" rainfall will remain at the historical mean, although we have added some Gaussian noise to it, and we have chosen "future" nitrogen prices to be at the upper range of the historical observed values. Then, the prediction of the future crop distributions should give us some idea as to the impact of rising nitrogen prices, according to our model. The results are presented for Anglia on heavy soil. The main thing to take from this analysis is that legumes could increase quite dramatically. This is

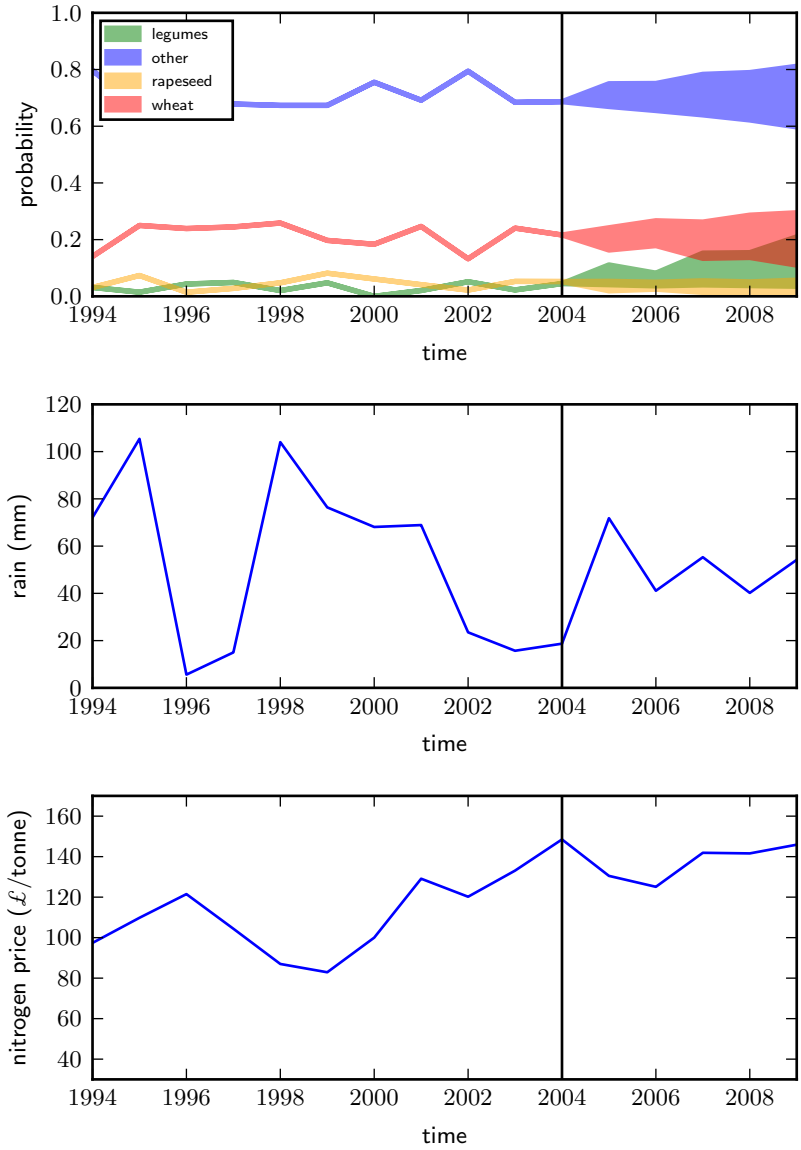


Figure 6.7: A prediction of future crop distributions in 2005–2009 in Anglia on heavy soil, given values of rainfall and nitrogen price for that period. Here, we consider a future scenario involving high nitrogen price and average rainfall.

in keeping with our beliefs that legumes are more preferable when nitrogen price is high as they do not require nitrogen fertiliser.

We must note that the inferences we take from an analysis such as this are not currently useful, due to the issue we have with the data ending in 2004. However, the work presented in this section is a common application of land use models, and we have shown that our model can perform such an analysis. If we had more modern data, this application of our model could help a government or policy maker predict the impact that changes in certain regressors may have on future crop distributions. In order to investigate this impact, we would need to obtain future scenarios for those regressors of interest. How these scenarios may be chosen is discussed in the next section.

6.3 Future data sets

Instead of performing scenarios based on fixed, hypothetical scenarios, we could use future predictions for the regressors generated by other models. One simple modelling technique we could use is a time-series analysis on our rainfall and economic data. However, there are also many existing models for the climate and the economy which also provide suitable predictions for use in our model.

Due to the importance of changes in climate, there are a large number of existing models that we could use. One particularly interesting model is the UK Climate Projections model [69] which provides predictions of future rainfall (among other climatic variables) across the United Kingdom, under various possible emission scenarios. It presents results in the form of probability distributions, therefore giving some idea of the uncertainty of the projections, making it an appealing model to consider for future rainfall predictions.

For economic data, there are also a large number of models that could be used. It is tricky to predict the future crop profit margin, as there are a number of factors that make up the margin. There are however predictions for some of the constituent parts. For example, the OECD-FAO agricultural outlook [75] provides regional, national and global predictions of agricultural commodity markets until 2023. This

includes, but is not limited to, predictions of the selling price of certain main crops (such as wheat), and predictions of variable costs such as fertiliser prices. From this, we could find predictions of profit margin for some of the more common crops, although this would not be possible for rarer crops like legumes, for which predictions are not available for.

We do not use any of these models here, as making predictions from 2015 onwards, based on a model built on data which stops in 2004, is not particularly useful. However, if field level data was collected again, then these data sets could be used to make relevant predictions.

6.4 Chapter summary

In this chapter, we have investigated one use of our land use model that could be helpful to land use experts. That is, predicting bounds on future crop distributions, given a future scenario for the regressors. Using future scenarios is a common feature of many land use models. In order to do this with our model, we used the methodology for making predictions in a non-stationary stochastic process, which was presented in section 4.4.

Due to our data ending in 2004, a 5 year "future" prediction involves predicting the crop distribution until 2009. This means that we could use the actual observed rainfall and nitrogen price as our scenario. Unfortunately, we could not compare the actual crop distribution with our predictions, as field level data stops in 2004. So, later in section 6.1, we split our data set in order to build the model on the observations from 1993 to 1999, and then we predicted the crop distribution in the following five years, and compared with what actually happened. Generally, we saw that the model captured the actual distributions well.

In section 6.2 we carried out the type of analysis we would ideally like to do, which is inputting a hypothetical scenario into the model, and seeing what it predicts. We used a scenario in which future nitrogen price was high, and saw that, for rainfall at the historic mean value, the model predicts that legumes may increase somewhat. Finally, in section 6.3 we briefly discussed some models which predict future values

of our regressors. These could be useful if we had current field level crop data.

This chapter presented one potential application of our land use model. The following chapter will present another way which in which our model could be used. We will describe how decision theory can be used in conjunction with the model to provide a tool which could be useful for policy makers.

Chapter 7

Potential relevance to policy making

In this chapter, we will investigate how our robust Bayesian land use model can be used in a way that may be applicable to policy makers. That is, we will show how we can use the theory of decision making under uncertainty in order to aid a policy decision, such as the level of subsidy or taxation to set, with the aim of stimulating growth of a particular crop. In order to do this, we will need to specify a utility function. This will be discussed in section 7.1. Section 7.2 details, following [91], how we can obtain optimal decisions in our robust Bayesian setting. Finally, section 7.3 applies this theory, in conjunction with our land use model, to a hypothetical policy question. This is a novel application of a land use model, a basic version of which was presented in [78].

7.1 Utility functions

One potential use of models such as ours is to aid policy makers. We assume farmers are interested in maximising profit margins, and changes to policy can have an impact on these margins. These policy adjustments could include, for example, levels of taxation or alterations to the subsidy level of certain crops. Land use models can help predict the impact that these changes may have on a farmer's decision.

We could use the scenario analysis method, as discussed in chapter 6, to see what

the model predicts will happen given a future scenario. However, we can also analyse the problem from the policy makers viewpoint; that is, what change in policy might make certain crops more favourable to the farmers. A policy maker must also take into account the benefit to society as a whole. For example, a very high subsidy level may be detrimental to wider society, either by costing too much money, or perhaps leading to an overproduction of the crop. We will assume that a policy maker has control over some economic factor. We choose not to consider climatic factors, as it is less reasonable to assume the policy maker has control over these.

We assume that a policy maker can specify utilities for different outcomes. That is, they can specify a function that represents their preferences. We can then use these utilities to make a recommendation as to what values of the regressor of interest achieves the highest expected utility. In our robust Bayesian setting, we investigate the effect of a wide range of priors on the optimal decision. As some crops are fairly rare in some cases, this allows us to identify situations where we do not have sufficient information in order to arrive at a conclusion.

We assume the utility function has the form $U(a, b, \kappa)$, where a is the fraction of the crop of interest and b is the cost to the policy maker – for example, nitrogen price. The fraction a is a function of β, our model parameter, and also b. $\kappa \geq 0$ is chosen by the decision maker, to control how the price is weighed against the level of the crop of interest. For example, a simple linear utility function may have the form

$$U(a, b, \kappa) = 100a - \kappa b \tag{7.1}$$

Note that a is multiplied by 100. This ensures a reasonable scale for the utility, but otherwise makes no technical difference, as utility functions are unique up to positive affine transformations (that is transformations of the form $cU + d$ for some $c, d \in \mathbb{R}$, $c > 0$ – see, for example, [61, Section 5.4.10]).

We could use a more advanced utility function than that defined in equation (7.1), which may be a more realistic reflection of a policy makers preferences. However, to obtain such a function would involve eliciting the required information from a policy maker, which, as this section is a proof of concept, is not feasible. Nevertheless, the analysis presented here easily generalises to a more realistic utility function.

As in the rest of the thesis, we do not actually calculate the expected utility, as this is computationally too expensive. Instead, we directly use the MAP estimate for β, and calculate the corresponding value for a

$$a(\beta^*, b) := \frac{\exp(\beta^*_{ij} x)}{\sum_{h=1}^{J} \exp(\beta^*_{ih} x)}$$

The (approximate) optimal decision is then the values of b that satisfy

$$\text{opt}(\mathscr{B}) := \arg\max_{b \in \mathscr{B}} U(a(\beta^*, b), b, \kappa) \tag{7.2}$$

where $a(\beta^*, b)$ is the fraction of crop of interest in the model with MAP parameter β^* and regressor b, and $\mathscr{B}$ is the set of feasible values of b. The decision criteria in equation (7.2) is known in the literature as *maximising expected utility* [91, Equation 3]. Note that since we use the MAP estimate we do not actually calculate an expectation, so this name is strictly not correct in this instance. However, for consistency with the literature we will proceed with this name. If our posterior distribution was symmetric and unimodal, then the MAP estimate would be the same as the posterior expectation, but this is not the case here. The effect of replacing the expectation with the MAP estimate is unknown, and would be worthy of further research.

7.2 Decision making under uncertainty

In our robust Bayesian setting, as discussed in chapter 4, we actually have a set B^* of β^* values. This means that we cannot simply make decisions as in equation (7.2), as we now have a set of utility functions $U(a, b, \kappa)$, one corresponding to each $\beta^* \in B^*$. In order to make decisions in this case, we can use the *lower and upper expected utilities*,

$$\underline{U}(b) := \inf_{\beta^* \in B^*} U(a(\beta^*, b), b, \kappa)$$

$$\overline{U}(b) := \sup_{\beta^* \in B^*} U(a(\beta^*, b), b, \kappa)$$

We still wish to arrive at a set $\text{opt}(\mathscr{B})$ of optimal values of b, and there are a number of decision criteria we can use in order to arrive at this set. There are many decision

criteria available in the literature ([91] provides an overview of the theory), and here we will present four criteria for selecting $\text{opt}(\mathscr{B})$.

Two simple decision criteria come from an obvious generalisation of maximising expected utility: we can maximise the lower expected utility or maximise the upper expected utility. This corresponds to taking either the most pessimistic stance, or the most optimistic stance. These decision criteria are known as Γ-*maximin* and Γ-*maximax*.

Definition 7.1 [91, Section 3.1] The set of optimal decisions given by Γ-*maximin* is

$$\text{opt}_{\underline{\Gamma}}(\mathscr{B}) := \left\{ b \in \mathscr{B} : \arg\max_{b \in \mathscr{B}} \underline{U}(b) \right\}$$

Meanwhile, the set of optimal decisions given by Γ-*maximax* is defined as

$$\text{opt}_{\overline{\Gamma}}(\mathscr{B}) := \left\{ b \in \mathscr{B} : \arg\max_{b \in \mathscr{B}} \overline{U}(b) \right\}$$

Γ-maximin and Γ-maximax are quite extreme decision criteria, which has lead to criticisms of them; Γ-maximin for being too conservative, and Γ-maximax for being too bold. Due to the fact they are extreme criteria they tend to return a single optimal decision, which can be a useful property. However, there are criteria that are more likely to return a set of decisions, which we may decide, in some circumstances, better reflects our uncertainty. One such criterion is known as *interval dominance*:

Definition 7.2 [91, Equation 8] The set of optimal decisions given by *interval dominance* is

$$\text{opt}_I(\mathscr{B}) := \left\{ b \in \mathscr{B} : \overline{U}(b) \geq \max_{b \in \mathscr{B}} \underline{U}(b) \right\}$$

Interval dominance is a rather weak criterion: it excludes all values of b for which the upper utility is lower than the maximal lower utility over all b, and as such it generally returns a large set of optimal decisions.

A stronger criterion than interval dominance, and one which approaches the problem from a different direction, can be described as a robust Bayesian criterion. We have a set B^* of model parameters, and for each $\beta^* \in B^*$ we can find the value of b that maximises our utility, as in equation (7.2). We can then consider every

$\beta^* \in B^*$, which will return a set $\text{opt}_{RB}(\mathscr{B}) \subseteq \text{opt}_I(\mathscr{B})$ of optimal decisions (for proof of this result, see [91, Theorem 1, p. 25]).

Definition 7.3 The set of optimal decisions given by a *robust Bayesian* decision criterion is:

$$\text{opt}_{RB}(\mathscr{B}) := \left\{ \beta^* \in B^* : \arg\max_{b \in \mathscr{B}} U(a(\beta^*, b), b, \kappa) \right\}$$

Clearly, our decision will depend heavily on the decision criteria that we choose. We refer to [91, Section 4] for a detailed discussion of which criteria might be the best choice in a particular situation. Here, we will use all 4 criteria we have defined, and compare the results, on the hypothetical policy aim of increasing the amount of legumes being grown.

7.3 Application to increasing legumes

As mentioned in chapter 2, there is currently a real interest in the United Kingdom in increasing the amount of legumes being grown. This is due to the desire to grow more home grown sources of protein. We can use the methodology presented in the previous two sections, in conjunction with our land use model, to analyse how a farmer's behaviour may change in response to government policy.

Both legumes and rapeseed are break crops from wheat, so we are particularly interested in the transitions from wheat. As we found in section 5.2.2, if we include nitrogen price rather than profit margin in the model, then our results seem more intuitive. We saw that, for all soil types, as nitrogen prices increased, the amount of legumes grown after wheat increased too, so we will use this model for our policy analysis. To inform policy makers, we will consider a series of scenarios with varying nitrogen price, and thereby investigate the hypothetical impact on crop transitions. Because legumes require far less fertiliser than rapeseed, we expect that an increase in nitrogen price leads to an increased growing of legumes.

There is perhaps a more obvious way to approach this problem. A more intuitive way a government could aim to stimulate legume growth is by offering a subsidy to grow them. We have the data available to us to attempt this. We can perform

an analysis where we alter the subsidy level of legumes directly. One would expect that as legume subsidy increases, more farmers would plant legumes as a break crop. The results for this particular model are shown in figure 7.1.

Figure 7.1 shows $[\underline{U}(b), \overline{U}(b)]$, in green, when moving from wheat, for various values of κ, in Anglia on heavy soil. Of course, in reality, a government would not base policy on previous crop or soil. However, we present this analysis as it shows a variety of interesting features, and also compares well with the validation plots in figure 5.21.

Figure 7.1 also shows our four decision criteria. The horizontal black line represents $\max_{b\in\mathscr{B}} \underline{U}(b)$, and the value(s) of b for which this coincides is $\text{opt}_{\underline{\Gamma}}(\mathscr{B})$, while the horizontal grey line represents $\max_{b\in\mathscr{B}} \overline{U}(\mathscr{B})$, and coincides with $\text{opt}_{\overline{\Gamma}}(\mathscr{B})$. Values of b for which $\overline{U}(b)$ lies above the black line are optimal by interval dominance, and are shown by the union of the brown and yellow regions. Values of b in the brown region only correspond to $\text{opt}_{RB}(\mathscr{B})$.

When $\kappa = 0$, our utility does not take into account the subsidy paid for legumes. Even in this case, our optimal values of legume subsidy, according to Γ-maximin and Γ-maximax, are low. We also see that $\text{opt}_{RB}(\mathscr{B}) = \text{opt}_{I}(\mathscr{B}) = \mathscr{B}$. That is they are vacuous. For all $\kappa > 0$, we do include subsidy level in our utility function, and we can see that, as κ increases, our decision criteria increasingly favour lower values of subsidy. This is somewhat counter-intuitive.

A possible explanation for this behaviour is that the policy maker simply deems any legume subsidy detrimental to society as a whole. Although using nitrogen price directly makes the analysis less intuitive, it produces interesting (and perhaps more sensible) results, and so for this reason we proceed with using nitrogen price directly.

Figures 7.2, 7.3 and 7.4 show $[\underline{U}(b), \overline{U}(b)]$ and our decision criteria when moving from wheat on each soil type and for various values of κ, when we have included nitrogen price in the model instead of the level of subsidy.

The same trends are observable across all soil types. When $\kappa = 0$, we are saying that the policy maker is indifferent to changes in nitrogen price. As such a high nitrogen price is desirable, as the model predicts this leads to an increase in legume growth. Thus, optimal values of b are high – in fact in this case $\text{opt}_{\underline{\Gamma}}(\mathscr{B}) =$

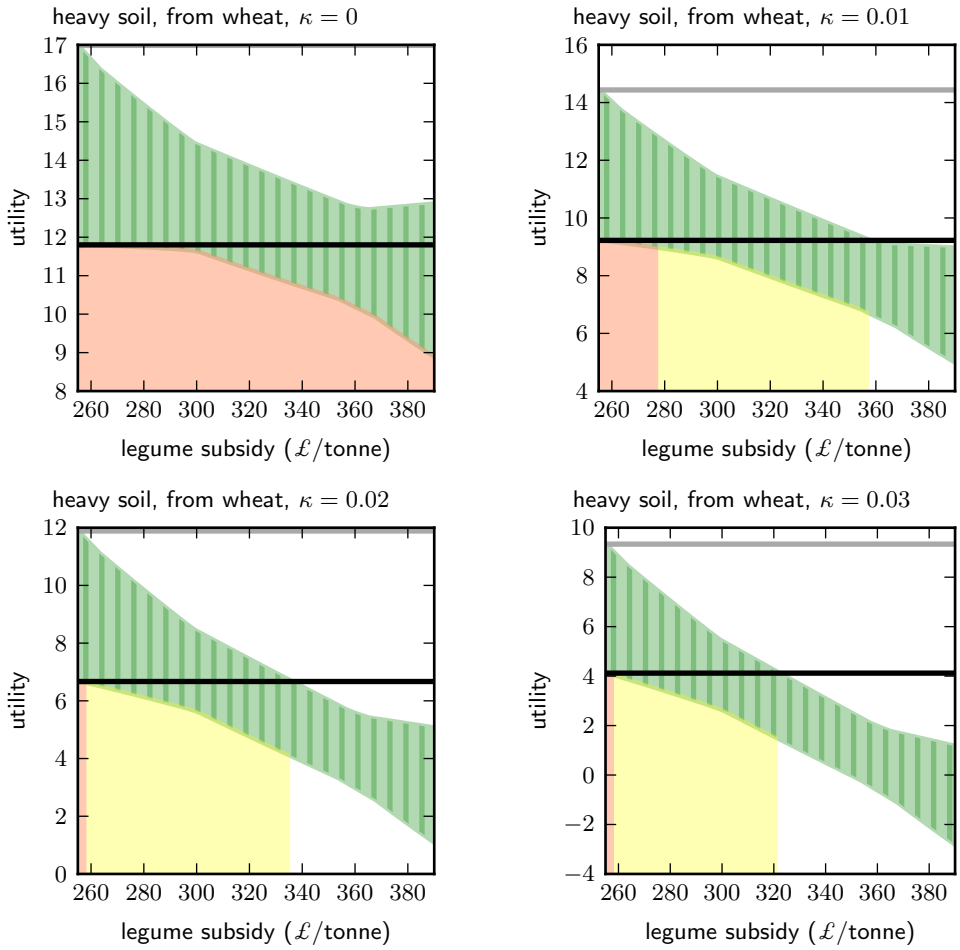


Figure 7.1: Utility functions and decision criteria for a model which contains legume subsidy directly, under four different values of κ. The black line represents $\max_{b \in \mathscr{B}} \underline{U}(b)$, and the value(s) of legume subsidy for which this coincides is $\mathrm{opt}_{\underline{\Gamma}}(\mathscr{B})$. Meanwhile the grey line represents $\max_{b \in \mathscr{B}} \overline{U}(b)$, and the value of legume subsidy for which this coincides is $\mathrm{opt}_{\overline{\Gamma}}(\mathscr{B})$. The brown region represents $\mathrm{opt}_{RB}(\mathscr{B})$, while the union of the brown and yellow regions represents the set $\mathrm{opt}_{I}(\mathscr{B})$.

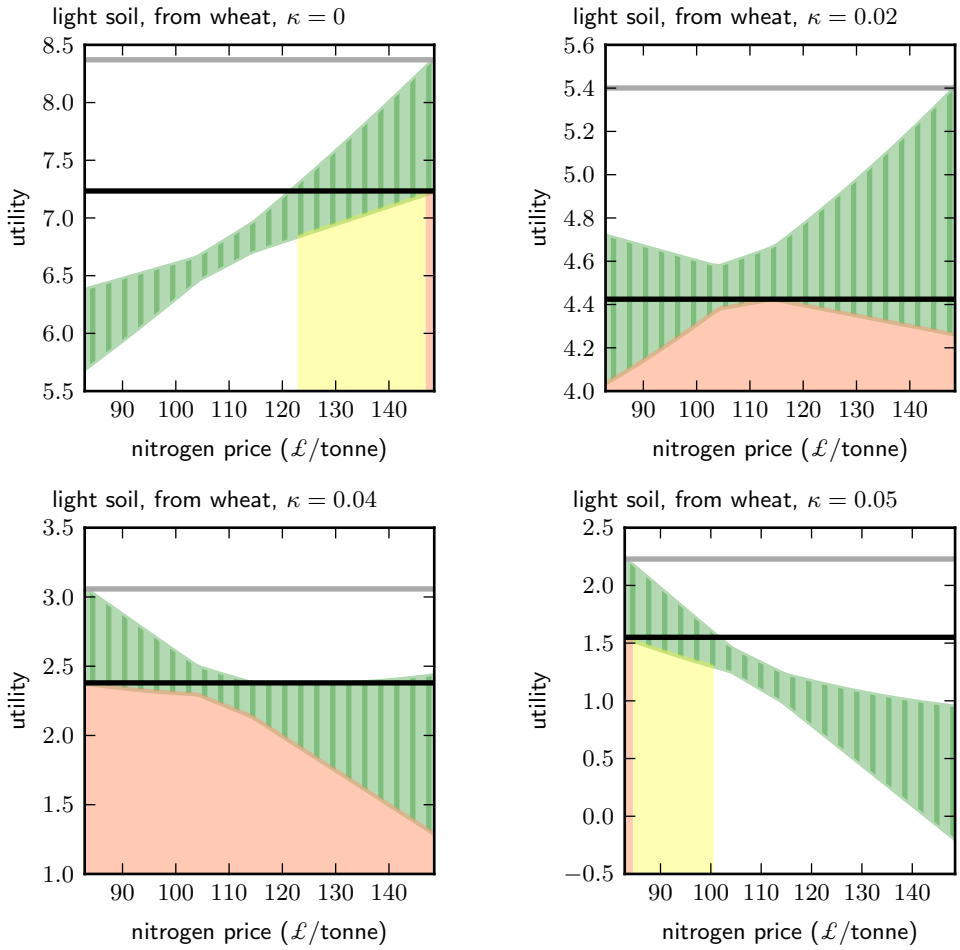


Figure 7.2: Utility functions and decision criteria for a model which includes nitrogen price, on light soil in Anglia, under four different values of κ. The black line represents $\max_{b \in \mathscr{B}} \underline{U}(b)$, and the value(s) of nitrogen price for which this coincides is $\mathrm{opt}_{\underline{\Gamma}}(\mathscr{B})$. Meanwhile the grey line represents $\max_{b \in \mathscr{B}} \overline{U}(b)$, and the value of nitrogen price for which this coincides is $\mathrm{opt}_{\overline{\Gamma}}(\mathscr{B})$. The brown region represents $\mathrm{opt}_{RB}(\mathscr{B})$, while the union of the brown and yellow regions represents the set $\mathrm{opt}_{I}(\mathscr{B})$.

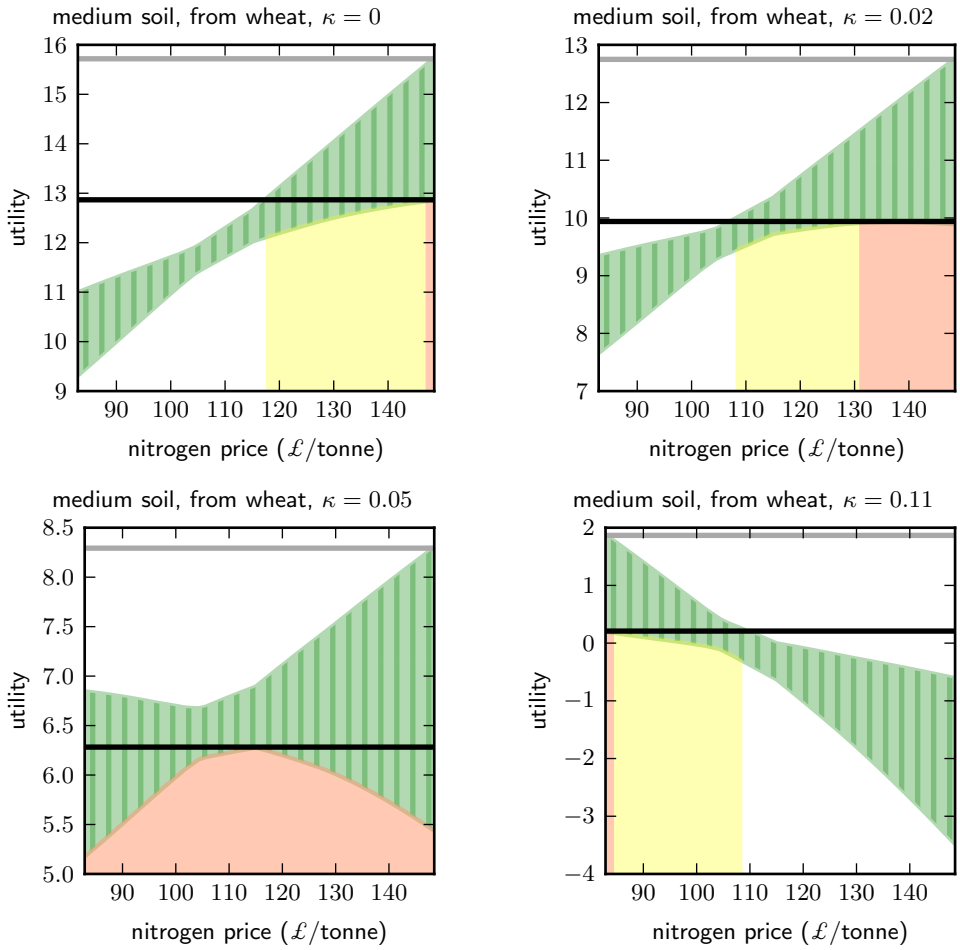


Figure 7.3: Utility functions and decision criteria for a model which includes nitrogen price, on medium soil in Anglia, under four different values of κ. The black line represents $\max_{b \in \mathscr{B}} \underline{U}(b)$, and the value(s) of nitrogen price for which this coincides is $\text{opt}_{\underline{\Gamma}}(\mathscr{B})$. Meanwhile the grey line represents $\max_{b \in \mathscr{B}} \overline{U}(b)$, and the value of nitrogen price for which this coincides is $\text{opt}_{\overline{\Gamma}}(\mathscr{B})$. The brown region represents $\text{opt}_{RB}(\mathscr{B})$, while the union of the brown and yellow regions represents the set $\text{opt}_{I}(\mathscr{B})$.

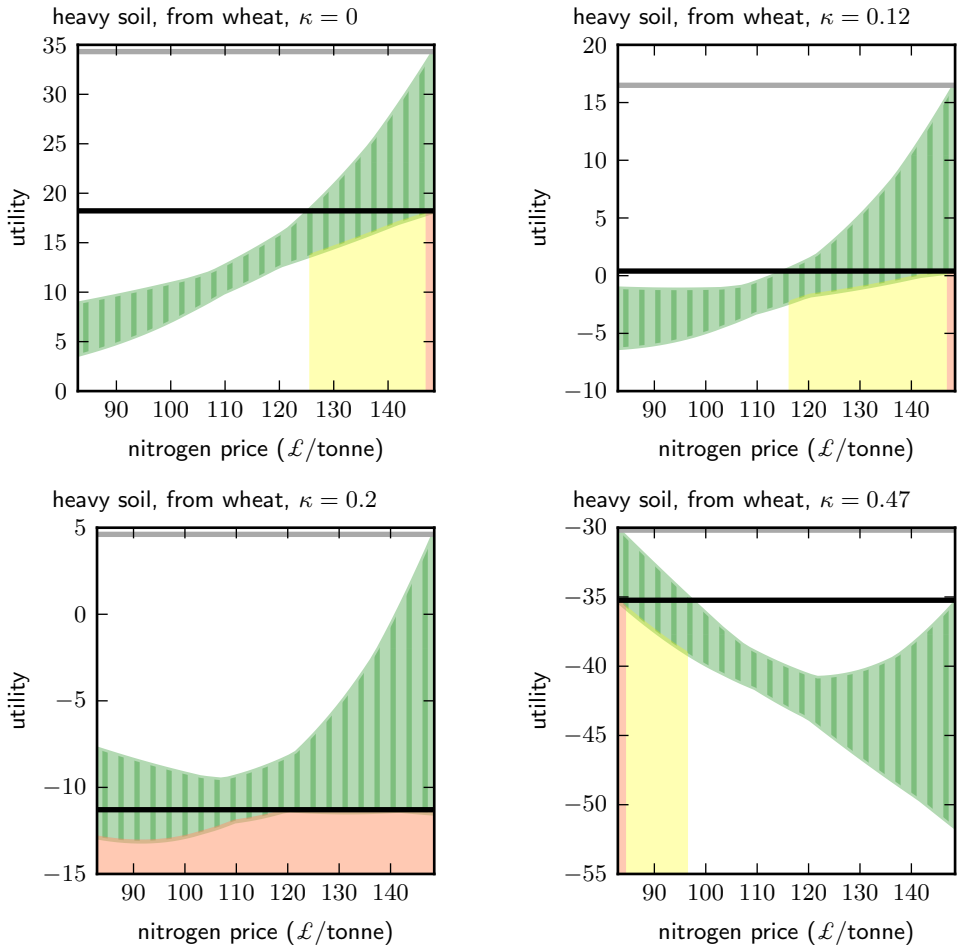


Figure 7.4: Utility functions and decision criteria for a model which includes nitrogen price, on heavy soil in Anglia, under four different values of κ. The black line represents $\max_{b \in \mathscr{B}} \underline{U}(b)$, and the value(s) of nitrogen price for which this coincides is $\text{opt}_{\underline{\Gamma}}(\mathscr{B})$. Meanwhile the grey line represents $\max_{b \in \mathscr{B}} \overline{U}(b)$, and the value of nitrogen price for which this coincides is $\text{opt}_{\overline{\Gamma}}(\mathscr{B})$. The brown region represents $\text{opt}_{RB}(\mathscr{B})$, while the union of the brown and yellow regions represents the set $\text{opt}_{I}(\mathscr{B})$.

$\text{opt}_{\overline{\Gamma}}(\mathscr{B}) = \text{opt}_{RB}(\mathscr{B}) = \max(\mathscr{B})$ on all soil types.

As we increase κ, equation (7.1) says that a higher nitrogen price is becoming more detrimental to society. As such, we expect lower values of b to become optimal. Eventually, we reach a point for which all b are optimal by interval dominance and the robust Bayesian criterion. For example, on light soil this occurs at $\kappa = 0.02$, although we can see that $\text{opt}_{\underline{\Gamma}}(\mathscr{B})$ and $\text{opt}_{\overline{\Gamma}}(\mathscr{B})$ still contain only one value of b. As $\text{opt}_{I}(\mathscr{B})$ and $\text{opt}_{RB}(\mathscr{B})$ are vacuous, this suggests that there is a large amount of uncertainty, so we may question the use of Γ-maximin and Γ-maximax in this situation.

Eventually, for all soils, we reach a stage where a high nitrogen price is highly undesirable for society, regardless of the benefits that it may bring with respect to a predicted increased in legume growth. On light soil, this occurs when $\kappa = 0.05$, and for $\kappa = 0.11$ on medium soil. On heavy soil, we must consider relatively large values of κ before reach a situation where *only* small b are optimal, and this change happens for $\kappa \geq 0.47$. Interestingly, $\kappa = 0.2$ in figure 7.4 displays an example when even Γ-maximin provides a range of optimal b values, implying we are highly uncertain in this situation.

For a policy maker, once decided on a value of κ (which would be determined by the policy maker determining what scenarios they are indifferent between), then they would need to decide how to actually alter the nitrogen price to suit society's needs. For example, on heavy soil with $\kappa = 0.12$, a higher nitrogen price is deemed beneficial to society by all of our decision criteria, according to our linear utility function. As such, a government could perhaps increase the tax on nitrogen in order to increase the price of it. On the other hand when $\kappa = 0.47$, lower values of b are deemed optimal, so a government could maybe decrease the tax on nitrogen to reduce the price of nitrogen based fertiliser to the farmer.

We must note here that the model that has been developed throughout this thesis is not a causal model [79]. That is, our model tells us of associations between the regressors and the crop choice, but we can not draw any causal conclusions. As such, one must be wary not to give too strong an interpretation to the inferences presented here.

7.4 Chapter summary

In this chapter we have shown one application of land use models, namely how they can be combined with decision theory in order to aid policy makers. As far as we are aware, this is a novel application of land use modelling. In order to perform this analysis, we had to assign a utility function for the policy maker with respect to the level of legumes being grown. We specified a simple linear utility function as this analysis is a proof of concept only, but including a more realistic utility function is simple (although perhaps not so simple to elicit from the policy maker). We then, using the theory presented in [91], discussed some decision criteria which can be used to make decisions in our robust Bayesian framework. As we use the MAP estimate for our posterior inferences, this is a novel method for decision analysis, based on this estimate.

Finally, we applied the methodology presented to a hypothetical yet realistic decision policy, namely aiming to increase the amount of legumes being grown. We used two separate models – one using legume subsidy, one using nitrogen price – to analyse the policy decision. We found the analysis carried out using legume subsidy directly produced surprising results. However, when using nitrogen price, the results were more as expected. While there were differences between the soil types, the same trend was observable, namely that as κ increased, lower values of nitrogen price became optimal.

Chapter 8

Concluding discussion

This chapter will summarise the work presented in this thesis, and will suggest some potential areas of future work.

8.1 Summary of the thesis

This thesis was concerned with building a novel approach to land use modelling. The model consists of a non-stationary stochastic process which fulfils the Markov property. We assumed the transition probabilities in this process were influenced by external non-stationary random variables, and we assumed a multinomial logit model for these transition probabilities. We then performed a robust Bayesian analysis on the parameters of the logit model. We applied this model to real-world data, and explored ways in which the model can be used to aid agricultural scientists and policy makers.

In chapter 2, we discussed a brief overview of land use and land use modelling. We investigated some factors which may drive a farmer's crop choice, and talked about why modelling this process is important. We discussed why it is difficult, which is primarily due to limited data for rare crop types, and also the difficulty in obtaining prior expert opinion. We also mentioned some existing land use models, and discussed their strengths and weaknesses.

Chapter 3 laid the theoretical ground work for our model. We introduced stochastic processes and the Markov property. We then described how we can view the tran-

sition probabilities of this stochastic process as being dependent on non-stationary external variables, which represent the various economic and climatic factors that influence the crop choice. We assumed a multinomial logit model for these transition probabilities, and discussed some properties of this model. We proposed a conjugate prior distribution for the model parameters, which built on the work of Chen and Ibrahim. However, we adapted their work, by transforming the structure of the data, to remove their requirement of knowing the locations of the data a priori. The resulting conjugate prior distribution belongs to the curved exponential family, and we discussed some properties of this distribution.

Chapter 4 discussed what a robust Bayesian analysis is, how we perform one in this setting, and the implications of carrying out such an analysis. To perform a robust Bayesian analysis, a set of prior distributions had to be specified, and we argued why this was an applicable approach in our setting. Then, inspired by Walley's imprecise Dirichlet model, we proposed a broad class of priors for our model, representing a near-ignorance model, and discussed the choice of this set of priors. A robust Bayesian analysis gives a set of inferences, which turned our stochastic process into an imprecise stochastic process. We adapted the theory of imprecise Markov chains to discuss how we make forward inferences in our process.

Due to the number of distributions involved, a key requirement of our robust Bayesian analysis was efficient calculation. In chapter 4 we discussed two possible approaches: an MCMC simulation, and the MAP estimate. Due to the number of distributions we had to evaluate, an MCMC simulation was very slow. However, the MAP estimate, as it is the solution of a system of equations, could be computed very quickly. Although the MAP approximation is crude – we assumed that the mass of the posterior distribution is all centred at the posterior mode – we found that estimates of the posterior transition probabilities resulting from the MAP estimate were similar to the probabilities generated by the MCMC simulations.

In chapter 5 we introduced the data sets that we had access to, and we performed some initial exploratory analysis. We then built various versions of our land use model, with a particular focus on a break crop model. We investigated the effect that various different regressors had on the transition probabilities from wheat. We

found that when we used the profit margin as a regressor we got non-intuitive results, but when we used nitrogen price instead, the results appeared more sensible. Why this is is open to conjecture, but we discussed various hypotheses which may have caused this behaviour.

Later in chapter 5, we discussed three approaches to model validation: a comparison with a non-parametric estimation of the transition probabilities; a cross-validation approach; and a comparison with a multinomial model with no covariates. All three of these approaches indicated that our model performed well, particularly in regions of sparse data.

In chapter 6 we concentrated on how we can use the model to make predictions about future crop distributions. As the field level data that our model was built on stopped in 2004, predictions from the current date are of little use. One solution to this was to build our model on 1993–1999, and then predict the crop distributions for 2000–2004. We then compared these predictions with the actual crop distributions of those years (for which we had data for), and we found that the model performed well. We then discussed a hypothetical scenario of real-world interest, namely whether a higher nitrogen price would lead to an increase in the amount of legumes being grown. This is the type of analysis that land use models are commonly used for, although due to the issue with the data, this was mainly included as a proof of concept. Finally, we briefly discussed some existing models which predict future values for our regressors of interest. These models could be used in conjunction with our land use model if we had more up-to-date data.

In chapter 7 we presented another use of our land use model. This chapter focussed on how our model could be used in conjunction with decision theory in a way that could be useful to policy makers. We assumed a utility function could be specified, and, using the theory of decision making under uncertainty, investigated a hypothetical yet realistic scenario, involving a desire to increase the amount of legumes being produced in the United Kingdom.

8.2 Potential areas of future work

With research, there are always avenues for future development or ideas. As this thesis represents a novel approach to land use modelling, there are areas which could be improved upon. In this section, we will briefly discuss some ideas where this work could be developed.

Theoretically, a major part of our model is the use of the MAP estimate for our model parameter β. While the posterior transition probabilities generated by the MAP estimate appear to approximate the probabilities resulting from the MCMC simulations very well (see section 4.3.2), it is clearly a crude approximation. We could use a more advanced MCMC method than the Metropolis-Hastings algorithm, which would speed up the MCMC simulations. However, while each simulation would take less time, we would still have to carry out one simulation for each posterior distribution generated by our broad set of priors. We suspect that this means any MCMC simulation would be too computationally intensive for our purposes, although a full study of advanced MCMC methods would be interesting to carry out. Also of interest for further investigation would be whether a method such as approximate Bayesian computation would be applicable to our model.

While we explicitly aimed to build a near-ignorance model by choosing a near-vacuous set of prior distributions, there could be scope for including some elicitation techniques in the model. This may lead to a smaller set of priors, which could make the calculations somewhat less complex.

While the methodology presented in this thesis was specifically built with land-use modelling in mind, it could easily be used in other settings. The model could be used anywhere a Markov chain is influenced by non-stationary external variables. Applying the model to other data sets would be an interesting area to look at.

On the applied side, there is an obvious, although unavoidable, weakness – the data set stops in 2004. This means that any conclusions we have drawn need to be considered in the light of changes since 2004, which have been significant. For example, there is now just a single subsidy payment for all crops, while in 2004 each crop had a separate payment. Furthermore, there are certain environmental management requirements now, which if not fulfilled lead to a farmer having part

of their subsidy withdrawn, a policy known as cross-compliance. In addition, in the period of our data, there was a requirement for a farmer to leave approximately 10% of their arable land fallow. This policy, known as set-aside, was an attempt to stop overproduction of crops, and farmers often used this period to "clean" the field of pests and weeds. However, this obligation was stopped in 2008, which of course would alter the behaviour of the farmers. These are just some of the ways that farming has changed since 2004, in addition to a changing climate, advances in technology, and so on.

There is however an interesting development with regards to the collection of modern field level data. There has been some recent work at Fera with regards to collecting field level data from satellite images [18]. The project aimed to build a classifier that, having being trained on a sample dataset (in fact, they used the same IACS database that we have), produced a supervised classification, given 3 images of a region of arable land use. The method has an accuracy of around 80%, and has a higher accuracy than that for certain crop types, including wheat, rapeseed, and legumes. So, it seems that there may be scope for using this methodology in our land use model.

One question which would need answering is how the uncertainty may propagate through the model. Our model is built on observed (and therefore, we assume, accurate) data, rather than predictions, which although generally accurate, will sometimes be incorrect. There perhaps is some scope to include a credal classification method [26] in the satellite imaging project. However, this would raise another issue. Our model is built on precise data, so we would have to answer how we could modify our model so that it can be trained on sets of crops, rather than precise observations.

There is good opportunity for extending the scope and range of the model. There are many more arable regions in the United Kingdom, and the model could easily be used on those regions where data is available. The model on the whole gives quite sensible results, however, the choice of regressors could be made more sophisticated. While the regressors we included was driven by expert opinion, our measure of climate is rather simplistic, as its influence on the crop choice is more complex than simply rainfall in the month preceding the average drilling time. In reality, the

relationship between the level of rainfall, and the drainage capability of the soil, is very important to the farmer, and it would possibly improve the model if we could have a more realistic measure of this. How to go about this is not immediately obvious.

Furthermore, there is the issue we get with non-intuitive results when we include profit margin in our model. Perhaps the regressors we used simply were not important to the farmers, and that they are much more influenced by nitrogen price. It is also possible that the unusual behaviour is caused by the format of the data, as we used the predicted margins rather than the observed margins. If data is available for the actual observed margins then we could run the model again, and see if we still obtain the unusual results. The behaviour could be caused by variables we have not considered in our model, such as locality to processing plants or long-term contractual obligations, both of which are particularly important in the production of peas. Whether we could obtain relevant data for these factors is not clear. Using nitrogen price instead of profit margin does give more sensible results, and it is a perfectly reasonable factor to include, but the model could perhaps be improved by investigating this issue further.

Bibliography

[1] Aggarwal, P., Roetter, R., Kalra, N., van Keulen, H., Hoanh, C., and van Laar, H. *Land use analysis and planning for sustainable food security*. International Rice Research Institute, 2001.

[2] Agresti, A. *Categorical Data Analysis*, second ed. John Wiley and Sons, Hoboken, 2002.

[3] Agriculture and Horticulture Development Board. Guide to rotation. `http://cereals-2.ahdb.org.uk/publications/documents/cropresearch/Rotation.pdf`.

[4] Agriculture and Horticulture Development Board. Spring barley. `http://cereals-2.ahdb.org.uk/publications/documents/cropresearch/Spring_Barley.pdf`.

[5] Agriculture and Horticulture Development Board. Take-all. `http://cereals.ahdb.org.uk/cereal-disease-encyclopedia/diseases/take-all.aspx`.

[6] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*, second ed. John Wiley and Sons, 1984.

[7] Benavoli, A., and Zaffalon, M. A model of prior ignorance for inferences in the one-parameter exponential family. *Journal of Statistical Planning and Inference 142*, 7 (2012), 1960 – 1979.

[8] Berger, J. O. The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses*, J. B. Kadane, Ed. Elsevier Science, 1984, pp. 63–144.

[9] Berger, J. O. Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference 25* (1990), 303–328.

[10] Berger, J. O. An overview of robust Bayesian analysis. Tech. rep., Department of Statistics, Purdue University, October 1993.

[11] Bernado, J. M., and Smith, A. F. M. *Bayesian Theory.* John Wiley and Sons, 1994.

[12] Bernardo, J. M. Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B* (1979), 113–147.

[13] Bhattacharya, R. N., and Waymire, E. C. *Stochastic Processes with Applications.* John Wiley and Sons, 1990.

[14] Billingsley, P. Statistical methods in Markov chains. *The Annals of Mathematical Statistics* (1961), 12–40.

[15] Bock, R. D. Estimating multinomial response relations. In *Contributions to Statistics and Probability*, R. C. Bose, Ed. University of North Carolina Press, 1970, pp. 453–479.

[16] Boole, G. *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities.* Walton and Maberly, 1854.

[17] Box, G. E., and Tiao, G. C. *Bayesian inference in statistical analysis*, vol. 40. John Wiley and Sons, 2011.

[18] Brown, P. Remote sensing of crops. Internal Report, Fera, 2015.

[19] Bruns, H. A. Concepts in crop rotations. In *Agricultural Science*, G. Aflakpui, Ed. InTech, 2012.

[20] Castellazzi, M. S., Brown, I., Poggio, L., and Gimona, A. Multi-scale modelling of ecosystem services - an iterative approach. In *International Congress on Environmental Modelling and Software* (Ottawa, Canada, July 2010), D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli, and T. Filatova, Eds., pp. 396–405.

[21] Castellazzi, M. S., Matthews, J., Angevin, F., Sausse, C., Wood, G. A., Burgess, P. J., Brown, I., Conrad, K. F., and Perry, J. N. Simulation scenarios of spatio-temporal arrangement of crops at the landscape scale. *Environmental Modelling Software 25*, 12 (2010), 1881–1889.

[22] Castellazzi, M. S., Wood, G. A., Burgess, P. J., Morris, J., Conrad, K. F., and Perry, J. N. A systematic representation of crop rotations. *Agricultural Systems 97* (2008), 26–33.

[23] Castle, E., and Becker, M. *Farm business management, the decision-making process.* Macmillan, 1962.

[24] Chen, M.-H., and Ibrahim, J. G. Conjugate priors for generalized linear models. *Statistica Sinica 13* (2003), 461–476.

[25] Cooke, R. M. *Experts in uncertainty: opinion and subjective probability in science.* Oxford University Press, 1991.

[26] Corani, G., and Zaffalon, M. Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2. *Journal of Machine Learning Research 9*, 4 (2008), 581–621.

[27] Cressie, N., and Wikle, C. K. *Statistics for Spatio-Temporal Data.* John Wiley and Sons, 2011.

[28] Crossman, J., Whitehead, P., Futter, M., Jin, L., Shahgedanova, M., Castellazzi, M., and Wade, A. The interactive responses of water quality and hydrology to changes in multiple stressors, and implications for the long-term effective management of phosphorus. *Science of The Total Environment 454* (2013), 230 – 244.

[29] de Cooman, G., and Hermans, F. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence 172*, 11 (2008), 1400 – 1427.

[30] de Cooman, G., Hermans, F., and Quaeghebeur, E. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences 23*, 4 (2009), 597–635.

[31] Department for Environment, Food and Rural Affairs. Data collected for the index of the purchase prices of the means of agricultural production.

[32] Diaconas, P., and Ylvisaker, D. Conjugate priors for exponential families. *The Annals of Statistics 7*, 2 (1979), 269–281.

[33] Eickhout, B., and Prins, A. G. EURURALIS 2.0. technical background and indicator documentation. `http://www.eururalis.eu/index.htm`, 2008.

[34] European Commission. Nitrogen in agriculture. `http://ec.europa.eu/agriculture/envir/report/en/nitro_en/report.htm`.

[35] European Commission. The Common Agricultural Policy after 2013. `http://ec.europa.eu/agriculture/cap-post-2013/`.

[36] Fonnesbeck, C., Patil, A., Huard, D., and Salvatier, J. PyMC: Bayesian stochastic modelling in Python. `https://pymc-devs.github.io/pymc/`.

[37] Food and Agriculture Organization of the United Nations. What is the best cropping sequence for the farm? `http://www.fao.org/docrep/006/Y5146E/y5146e0a.htm`.

[38] Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association 100*, 470 (2005), 680–701.

[39] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. *Bayesian data analysis*, 2 ed. Chapman and Hall, 2004.

[40] Genest, C., and Zidek, J. V. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science 1*, 1 (1986), 114–135.

[41] Gilbert, C. L., and Morgan, C. W. Food price volatility. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 365*, 1554 (2010), 3023–3034.

[42] Goldstein, M., and Wooff, D. A. *Bayes Linear Statistics: Theory and Methods*. John Wiley and Sons, 2007.

[43] Good, I. J. *Probability and the Weighing of Evidence*. Griffin, 1950.

[44] Good, I. J. Rational decisions. *Journal of the Royal Statistical Society, Series B 14*, 1 (1952), 107–114.

[45] Good, I. J. Explicativity, corroboration, and the relative odds of hypotheses. *Synthese 30* (1975), 39–73.

[46] Haberman, S. J. *The Analysis of Frequency Data*. University of Chicago Press, 1974.

[47] Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*, 1 (1970), 97–109.

[48] Helming, J. F. M., Jansen, S., Van Meijl, H., and Tabeau, A. *European farming and post-2013 CAP measures; A quantitative impact assessment study*. No. 2010-085/2103. LEI, Wageningen, 2010.

[49] Iizumi, T., and Ramankutty, N. How do weather and climate influence cropping area and intensity? *Global Food Security 4* (2015), 46 – 50.

[50] IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2013.

[51] Jain, T. *Microeconomics and Basic Mathematics*. VK Publications.

[52] Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 186*, 1007 (1946), 453–461.

[53] Jensen, E. S., Peoplese, M. B., Boddey, R. M., Gresshoff, P. M., Hauggaard-Nielsen, H., Alves, B. J. R., and Morrison, M. J. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. a review. *Agronomy for Sustainable Development 32*, 2 (2012), 329–364.

[54] Jones, E., Oliphant, T., and Peterson, P. SciPy: Open source scientific tools for Python. `http://www.scipy.org/`.

[55] Jones, P. *Developing LUAM (Land Use Allocation Model) and Modelling CAP Reforms.* CAS paper. Centre for Agricultural Strategy, The University of Reading, 1995.

[56] Keynes, J. M. *A Treatise on Probability.* Macmillan, 1921.

[57] Knight, S., Somerwill, K., and Luo, W. Review of land use models suitable for consideration to aid future research in the UK. Internal Report, Food and Envirnoment Research Agency, 2011.

[58] Kuznetsov, V. P. *Interval Statistical Models.* Radio i Svyaz, 1991. In Russian.

[59] Lambin, E. F., and Meyfroidt, P. Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences 108*, 9 (2011), 3465–3472.

[60] Land Information System. NATMAP database. `http://www.landis.org.uk/data/natmap.cfm`.

[61] Lindley, D. V. *Making Decisions*, second ed. John Wiley and Sons, 1985.

[62] Lindley, D. V. *Understanding Uncertainty.* John Wiley and Sons, 2006.

[63] Link, W. A., and Eaton, M. J. On thinning of chains in MCMC. *Methods in Ecology and Evolution 3*, 1 (2012), 112–115.

[64] Living Field Institute. Biological nitrogen fixation by legumes. `http://livingfield.hutton.ac.uk/science/bnf`.

[65] Luo, W. Land use modelling. Internal Report, Food and Envirnoment Research Agency, 2010.

[66] Mantel, N. Models for complex contingency tables and polychotomous dosage response curves. *Biometrics 22*, 1 (1966), 83–95.

[67] Martin, J., Wilcox, L. C., Burstedde, C., and Ghattas, O. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *Journal on Scientific Computing 34*, 3 (2012), 1460–1487.

[68] Met Office. UK climate - historic station data. `http://www.metoffice.gov.uk/climate/uk/stationdata/`.

[69] Met Office. United Kingdom Climate Projections. `http://ukclimateprojections.metoffice.gov.uk/`.

[70] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21* (1953), 1087–1092.

[71] Murphy-Bokern, D., Watson, C., Stoddard, F. L., Lindström, K., Zander, P., Reckling, M., Preissel, S., Bues, A., and Torres, A. Outlook for knowledge and technology for legume supported cropping systems. `http://www.legumefutures.de/`, 2014.

[72] Natural England. National Character Areas. `http://publications.naturalengland.org.uk/category/587130`.

[73] Nemecek, T., von Richthofen, J.-S., Dubois, G., Casta, P., Charles, R., and Pahl, H. Environmental impacts of introducing grain

legumes into european crop rotations. *European Journal of Agronomy 28* (2008), 380–393.

[74] Nix, J. *Farm Management Pocketbook*, 23–34 ed. Agro Business Consultants Limited, 1993–2004.

[75] Organisation for Economic Co-operation and Development and the Food and Agriculture Organization of the United Nations. Agricultural Outlook 2014–2023. `http://www.agri-outlook.org/`.

[76] Paton, L., Troffaes, M. C. M., Boatman, N., Hussein, M., and Hart, A. East Anglian and Mease/Sence dataset. `http://dx.doi.org/10.15128/br86b361k`.

[77] Paton, L., Troffaes, M. C. M., Boatman, N., Hussein, M., and Hart, A. Multinomial logistic regression on Markov chains for crop rotation modelling. In *Proceedings of the 15th International Conference IPMU (Information Processing and Management of Uncertainty in Knowledge-Based Systems)* (2014).

[78] Paton, L., Troffaes, M. C. M., Boatman, N., Hussein, M., and Hart, A. A robust Bayesian analysis of the impact of policy decisions on crop rotations. In *ISIPTA'15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications* (2015), pp. 217–226.

[79] Pearl, J. Causal inference in statistics: An overview. *Statistics Surveys 3* (2009), 96–146.

[80] Powell, M. J. D. *Numerical Methods for Nonlinear Algebraic Equations*. Gordon and Breach Science Publishers, 1970, ch. A hybrid method for nonlinear equations, pp. 87–114.

[81] R Core Team. R: A language and environment for statistical computing. `https://www.R-project.org/`.

[82] Ribeiro, P. J., and Diggle, P. J. geoR: Analysis of geostatistical data. `https://cran.r-project.org/web/packages/geoR/geoR.pdf`.

[83] Rienks, W., Hoek, S., Verweij, P., Lokers, R., and Vanmeulebrouk, B. Eururalis, a discussion support tool for rural Europe. In *Proceedings of the iEMSs Fourth Biennial Meeting* (2008), pp. 988–998.

[84] Rigby, D., and Caceres, D. Organic farming and the sustainability of agricultural systems. *Agricultural Systems 68*, 1 (2001), 21 – 40.

[85] Robert, C. P., and Casella, G. *Introducing Monte Carlo Methods with R*. Springer, 2010.

[86] Ronfort, C., Souchere, V., Martin, P., Sebillotte, C., and Castellazzi, M. Methodology for land use change scenario assessment for runoff impacts: A case study in a north-western European Loess belt region (Pays de Caux, France). *CATENA 86*, 1 (2011), 36 – 48.

[87] Rural Payments Agency. Data collected under the Integrated Administration and Control System for the administration of subsidies under the Common Agricultural Policy.

[88] Stone, M. The opinion pool. *The Annals of Mathematical Statistics 32*, 4 (1961), 1339–1342.

[89] Stramer, O., and Tweedie, R. L. Langevin-type models II: Self-targeting candidates for MCMC algorithms. *Methodology And Computing In Applied Probability 1*, 3 (1999), 307–328.

[90] Theil, H. A multinomial extension of the linear logit model. *International Economic Review 10*, 3 (1969), 251–259.

[91] Troffaes, M. C. M. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning 45*, 1 (2007), 17–29.

[92] Troffaes, M. C. M. Bayesian statistics III/IV lecture notes, 2013–14. Durham University.

[93] Troffaes, M. C. M., and Paton, L. Logistic regression on Markov chains for crop rotation modelling. In *ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications* (2013), pp. 329–336.

[94] Tutz, G. *Regression for Categorical Data.* Cambridge University Press, 2012.

[95] Tzanopoulos, J., Jones, P. J., and Mortimer, S. R. The implications of the 2003 common agricultural policy reforms for land-use and landscape quality in england. *Landscape and Urban Planning 108*, 1 (2012), 39 –48.

[96] UK Agriculture. Production cycle for wheat. `http://www.ukagriculture.com/production_cycles/wheat_production_cycle.cfm`.

[97] van Rossum, G., and Drake, F. Python reference manual version 2.7.10. `http://www.python.org`.

[98] Walley, P. Coherent lower (and upper) probabilities. Tech. rep., University of Warwick, 1981. Statistics Research Report 22.

[99] Walley, P. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1991.

[100] Walley, P. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B 58*, 1 (1996), 3–34.

[101] Walley, P. Measures of uncertainty in expert systems. *Artificial Intelligence 83* (1996), 1–58.

[102] Walter, G., and Augustin, T. Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice 3* (2009), 255–271.

[103] Weichselberger, K. The theory of interval probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning 24* (2000), 149–170.

[104] Whittle, P. *Probability via Expectation*, third ed. Springer, 1992.

[105] Williams, P. M. Notes on conditional previsions. Tech. rep., School of Mathematis and Physical Sciences, University of Sussex, 1975.

[106] Williams, P. M. Notes on conditional previsions. *International Journal of Approximate Reasoning 44*, 3 (2007), 366–383.

[107] Winkler, R. L. The assessment of prior distributions in bayesian analysis. *Journal of the American Statistical Association 62*, 319 (1967), 776–800.

[108] Ya-hao, C., Gao, Z., and Seale, Jr., J. Changing structure of China's meat imports. *Journal of Integrative Agriculture 14*, 6 (2015), 1081–1091.