A Robust Bayesian Analysis
of the Impact of Policy Decisions on Crop Rotations

Lewis Paton
Durham University, UK
l.w.paton@durham.ac.uk

Matthias C. M. Troffaes
Durham University, UK
matthias.troffaes@gmail.com

Nigel Boatman, Mohamud Hussein
The Food and Environment Research Agency, UK
{nigel.boatman, mohamud.hussein}@fera.gsi.gov.uk

Abstract

We analyse the impact of a policy decision on crop rotations, using the imprecise land use model that was developed by the authors in earlier work. A specific challenge in crop rotation models is that farmer's crop choices are driven by both policy changes and external non-stationary factors, such as rainfall, temperature and agricultural input and output prices. Such dynamics can be modelled by a non-stationary stochastic process, where crop transition probabilities are multinomial logistic functions of such external factors. We use a robust Bayesian approach to estimate the parameters of our model, and validate it by comparing the model response with a non-parametric estimate, as well as by cross validation. Finally, we use the resulting predictions to solve a hypothetical yet realistic policy problem.

Keywords. multinomial logistic regression, stochastic process, robust Bayesian, conjugate, maximum likelihood, crop, decision

1 Introduction

This paper investigates a specific actual real-world problem, namely how imprecise probability can be used to inform policy, in a way that reflects limited data and lack of information to policy makers. In general, policy decisions aim to balance the greater good to society with the welfare of the individual, in terms of economic costs and benefits from a policy implementation. For example, farmers typically grow crops to maximise their profits, however governments can influence this decision through policy interventions to meet the needs of society, such as biodiversity, economic resilience, and security of supply.

An issue which has received a lot of attention recently concerns changes in crop rotations, which are linked to negative environmental impact, reduced diversification of crops and reduced self-sufficiency in feed and food.

Concerning animal feed, protein demand has increased a lot, due to increasing meat demand from developing countries. Also, the use of European legumes such as peas and beans [11] has declined. At the moment, the UK imports most of its protein; however, these prices are going up due to growing global demand for soya [7]. Simultaneously, growing more protein can improve diversity, and thereby increase resistance against disease and climate change, and improve supply security [8]. 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 [2].

We will look at a hypothetical scenario to see how nitrogen price affects the amount of legumes being grown. Legumes produce their own nitrogen, and so require little nitrogen based fertiliser. As such, one expects that farmers tend to grow less fertiliser dependant crops as nitrogen prices increase. We will formulate and answer a hypothetical decision problem which illustrates the types of problems that can be solved using land use models.

Farmers generally grow crops in rotation to prevent build-up of pests and diseases, and thereby to maximise yields and profit margins. The optimal crop choices vary with soil type and climate conditions. The rotation is generally driven by the length of the period required between successive plantings of the most valuable crop that can be grown, in order to allow pests and diseases to decline to non-damaging or readily controllable levels. Rotating crops also spreads risk in the face of weather variability and annual fluctuations in commodity prices.

Modelling crop distributions across time and space is highly non-trivial. Building a statistical model for farmers' crop choices is difficult, because there are so many factors that influence a farmer's choice. We need to take care in picking the relevant major influencing factors. Moreover, although we have a reasonably sized database, some crop types and factor levels are quite

rare. Furthermore, prior expert information is difficult to obtain. Thus, building a model capable of making reasonable inferences about future crop distributions is a difficult problem.

Building on the work of Luo [9], and Chen and Ibrahim [4], we previously developed a land use model that accurately captures uncertainty in the modelling process [16, 13]. In that work, a non-stationary stochastic process models crop choice, where crop transition probabilities are multinomial logistic functions, and predictions are based on sets of conjugate priors and MAP estimates for efficient sensitivity analysis. Here, we will use this model to answer the hypothetical policy question discussed earlier.

Compared to our earlier work in this domain [16, 13], the novel contributions of this paper are: (i) We train our model on a much larger data set, and handle a larger number of crop types. (ii) We deal with numerical stability issues resulting from near-zero counts. (iii) We propose a non-parametric estimation method, which is, as far as we know, new in the literature. (iv) We validate our model, using two different approaches: formally through classification based accuracy measures, and heuristically through comparison with non-parametric estimates. (v) We propose a new method for the decision analysis based on MAP estimation. (vi) We apply our model to a hypothetical policy decision problem.

The paper is structured as follows. Section 2 describes the land use model from [13]. Section 3 explains the set of priors and posterior inferences. Section 4 shows some of the results from the model. Section 5 describes the model validation. Section 6 analyses a decision problem. Section 7 concludes the paper.

2 The Model

We model crop rotations on a particular field as a non-stationary stochastic process, with J states, corresponding to J crop choices. The crop grown at time k is denoted by Y_k. The choice of Y_{k+1} is influenced by regressors $X_k = (X_{k0}, X_{k1}, \ldots, X_{kM})$, as well as by Y_k, but is otherwise independent of the history of the system. As usual in a regression analysis, we set $X_{k0} = 1$. We denote the transition probabilities by

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

We assume a multinomial logistic regression model for $\pi_{ij}(x)$, with $J^2(M+1)$ model parameters β_{ijm}, where $i \in \{1, \ldots, J\}$, $j \in \{1, \ldots, J\}$, and $m \in \{0, \ldots, M\}$:

$$\pi_{ij}(x) = \frac{\exp(\beta_{ij}x)}{\sum_{h=1}^{J} \exp(\beta_{ih}x)} \qquad (2)$$

i	x_1	x_2	$n_i(x)$	$k_{i1}(x)$	$k_{i2}(x)$	$k_{i3}(x)$	$k_{i4}(x)$
1	93	112	2	0	1	0	1
2	56	154	1	0	0	1	0
1	85	110	1	0	0	0	1
3	30	90	1	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Crop rotation data for a particular soil type, where 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, and $k_{ij}(x)$ is the number of crop j being grown.

with $\beta_{ij}x := \sum_{m=0}^{M} \beta_{ijm}x_m$. Without loss of generality we can set $\beta_{iJm} = 0$ for all i and m, and call this the *baseline category logit model* [3].

Soil type is a significant driver of crop choice. Following [9], we split our data by soil type, and perform a separate analysis for each soil type. For ease, we do not index our model parameters by soil type.

For estimation, we have $n_i(x)$ observations where the previous crop was i, and the regressors were x. Obviously $n_i(x)$ will be zero at all but a finite number of $x \in \mathcal{X}$, where $\mathcal{X} = \{1\} \times \mathbb{R}^M$. Of these $n_i(x)$ observations, the crop choice was j in $k_{ij}(x)$ cases. Obviously, $n_i(x) = \sum_{j=1}^{J} k_{ij}(x)$ for each i. Table 1 shows an extract from the data set.

The following conjugate prior for the model parameters β was proposed in [13]:

$$f_0(\beta|s_0, t_0) \propto \exp\left(\sum_{i=1}^{J} \sum_{x \in \mathcal{X}} s_{0i}(x)\left[\right.\right.$$
$$\left.\left. \sum_{j=1}^{J} t_{0ij}(x)\beta_{ij}x - \log \sum_{j=1}^{J} \exp(\beta_{ij}x)\right]\right) \quad (3)$$

where s_{0i} and t_{0ij} are non-negative functions such that $s_{0i}(x) = t_{0ij}(x) = 0$ for all but a finite number of $x \in \mathcal{X}$, with $0 \le t_{0ij}(x) \le 1$ and $\sum_{j=1}^{J} t_{0ij}(x) = 1$ on those points x where $s_{0i}(x) > 0$. This conjugate prior matches the form of the likelihood, and the posterior distribution and parameters are [13]:

$$f(\beta|k, n, s_0, t_0) = f_0(\beta|s_n, t_n) \qquad (4)$$
$$s_{ni}(x) = s_{0i}(x) + n_i(x) \qquad (5)$$
$$t_{nij}(x) = \frac{s_{0i}(x)t_{0ij}(x) + k_{ij}(x)}{s_{0i}(x) + n_i(x)} \qquad (6)$$

3 Inference

Because prior expert opinion is very difficult to obtain in our problem, we use sets of prior densities, similarly to Walley's IDM [18]. Here, we study inferences resulting from a fixed prior function for $s_{0i}(x)$:

$$s_{0i}(x) = \begin{cases} s & \text{if } x \in \mathfrak{X}, \\ 0 & \text{otherwise,} \end{cases} \qquad (7)$$

for some $\mathfrak{X} \subset \mathcal{X}$ and a near vacuous set $\mathfrak{T}$ of prior functions for t_0. Note that in earlier work [13] we used a full vacuous set, however we found that we need to bound the $t_{0ij}(x)$ parameters away from zero at those points where $s_{0i}(x) > 0$ in order to maintain numerical stability in cases where we have very few observations; we chose this bound $\epsilon = 0.01 > 0$ small enough to have no observable impact on the analysis.

$\mathfrak{X}$ is the set of regressor values where we specify prior beliefs. It can be any finite subset of $\mathcal{X}$, but we note that the inferences appear more intuitive if $\mathfrak{X}$ is chosen to sensibly cover the range of observed x values [16]. As in the imprecise Dirichlet model [18, Section. 2.5], smaller values of s typically produce tighter posterior predictive bounds. For further discussion of why this choice of priors makes sense, we refer to [13].

A standard way to do the inference now would go via MCMC. However, as we wish to perform a sensitivity analysis against the prior, and the dimension of the parameter space is very large, MCMC is too slow for our purpose. Therefore, we simply use MAP estimation. If we can find a MAP estimate for all $t_0 \in \mathfrak{T}$, 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. Therefore,

$$\hat{\underline{\pi}}_{ij}(x) \approx \inf_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J} \exp(\beta^*_{ih}x)} \tag{8}$$

$$\hat{\overline{\pi}}_{ij}(x) \approx \sup_{\beta^* \in B^*} \frac{\exp(\beta^*_{ij}x)}{\sum_{h=1}^{J} \exp(\beta^*_{ih}x)} \tag{9}$$

are the desired lower and upper posterior probability estimates of the transition probability.

4   Case Study

We have crop rotation data from two separate regions in the UK, detailing which crop was grown in every field in each region from 1993 until 2004 [15].

We have data available for a variety of regressors: here we look at rainfall [10] before sowing and the nitrogen price [1]. Rainfall is important as some crops grow better when it is wetter, and some soil types deal with heavy rainfall better. We can assume farmers are interested in maximising their profit margin. Most fertilisers are nitrogen based, and as such a high nitrogen price will impact profit margins for crops which require large amounts of fertiliser.

We will assume a farmer is faced with a choice of $J = 4$ types of crops: wheat, legumes, rapeseed and all other crops. A common practice is to grow wheat (generally the most profitable crop) followed by a *break crop*, such as legumes or rapeseed. Transitions between legumes
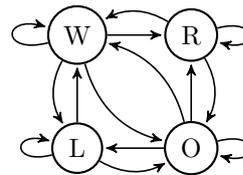

Figure 1: Possible crop transitions, where W is wheat, R is rapeseed, L is legumes, and O is other. Transitions between R and L do not occur in practice so have been excluded from the model.
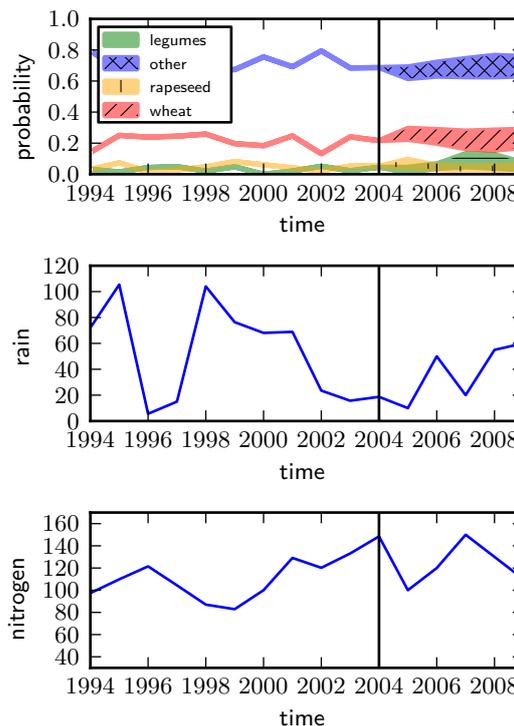

Figure 2: Prediction of future crop distributions on heavy soil given a future scenario.

and rapeseed are very rare (this only occurred 3 times in roughly 30000 observations). We could leave these transitions in, but they make negligible difference to the inferences, and experts have no interest in these transitions anyway. Therefore, we remove them from the model. Figure 1 depicts all crops and transitions in our model.

An important use of land use models is to predict what may happen in the future, given a future scenario for the regressors. For future crop distributions, we use the methodology for imprecise Markov chains developed in [6]. Our initial distribution is calculated empirically from the data and is 23% wheat, 5% rapeseed, 4% legumes and 68% others. Figure 2 shows the results for heavy soil.

The figure shows the historical crop distribution up to 2004 and the values of the regressors until that point. It then shows the predicted crop distributions for the next five years, given a future scenario for the regressors. Here, we have analysed what would happen if future rainfall will be quite low (compared to the observed historical values), and future nitrogen prices will be high. We can see that, although nothing drastic is predicted, legumes seem to increase somewhat. We could compare different scenarios to analyse the impact of changes in, say, nitrogen price. However, a government can influence regressors such as nitrogen price through policy. Thus, it is of interest to study a decision problem which aims to advise this policy. We do this analysis in section 6.

Note that our data runs from 1993 to 2004, so in fact the prediction is until 2009. It would be interesting to compare predictions with actually observed crop distributions, however field level data was no longer being collected from 2005 onwards. It may be possible in the future to validate against satellite data (we currently do not have such data in this study), and thereby to gauge the predictive power of the model.

5 Validation

We discuss two methods for validating the model. A first naive but simple way is to graphically compare the predicted transition probabilities with a non-parametric estimate from the data. A second way is to cross validate the model's predicted best response with parameters estimated from training data against the response as in the test data; this is similar to what is done in classification.

5.1 Non-Parametric Estimates

A simple non-parametric estimate of $\pi_{ij}(x)$ takes a weighted average of the observations around x:

$$\tilde{\pi}_{ij}(x) := \frac{\sum_{x' \in \mathcal{X}} w(x - x') k_{ij}(x')}{\sum_{x' \in \mathcal{X}} w(x - x') n_i(x')} \qquad (10)$$

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. We took a multivariate Gaussian kernel:

$$w(x) := |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \qquad (11)$$

with

$$\Sigma^2 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 20^2 & 0 \\ 0 & 0 & 20^2 \end{bmatrix} \qquad (12)$$

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.

Figure 3 depicts $\tilde{\pi}_{ij}(x)$ as calculated from eq. (10) and $[\hat{\underline{\pi}}_{ij}(x), \hat{\overline{\pi}}_{ij}(x)]$ as calculated from eqs. (8) and (9), for all cases of previous crop i and soil type, as a function of nitrogen price and for a fixed value of rainfall (we chose the historic mean, 55mm). 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.

Figure 4 shows a smoothed version of $n_i(x)$, that is:

$$\sum_{x' \in \mathcal{X}} w(x - x') n_i(x') / w(0) \qquad (13)$$

These plots give an idea of the size of the denominator in eq. (10), 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 around 20. The highest data density is observed from other on light soil type, where we see numbers between 1000 and 2300. 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.

The worst fits are observed in the two bottom right plots, where the robust Bayesian model seems to slightly overestimate the slopes of the curves. We currently have no good explanation as to why this behaviour occurs.

5.2 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 fig. 3, 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; fig. 3 just shows 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 [5]. The *determinacy* is the percent-
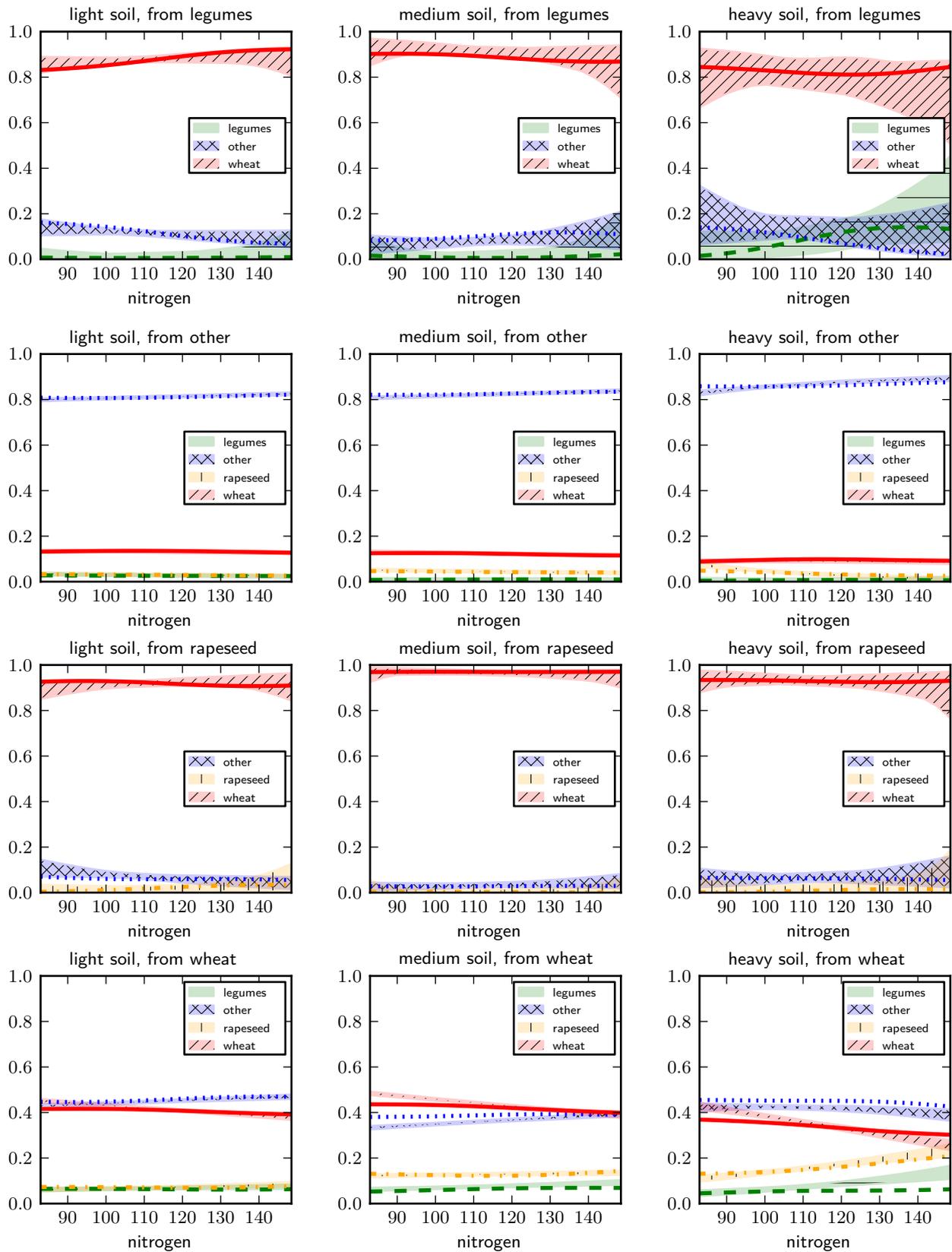
Figure 3: Non-parametric estimates $\tilde{\pi}_{ij}(x)$ and robust Bayesian interval estimates $[\hat{\underline{\pi}}_{ij}(x), \hat{\overline{\pi}}_{ij}(x)]$ for all previous crops i, soil types, as a function of nitrogen price, for fixed rainfall. Probability lies on the y axis.

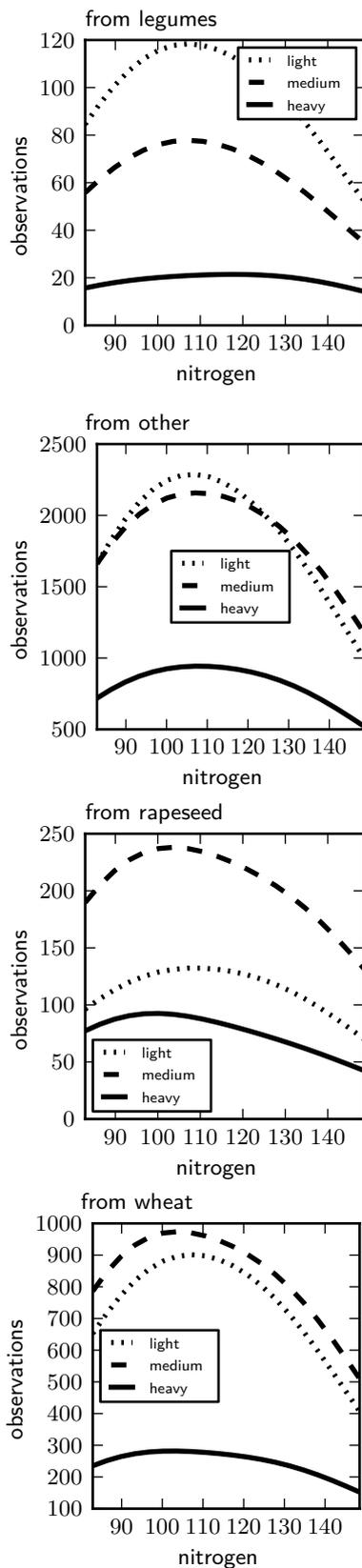from legumes

from other

from rapeseed

from wheat

Figure 4: Smoothed $n_i(x)$ as a function of x. This gives an idea of the accuracy of the non-parametric estimates plotted in fig. 3.

region	deter-minacy	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 2: Cross validation results

age 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 part.

We can use a similar approach to validate our model. We have two distinct geographical regions in our dataset. We perform cross validation within each region, and also combine the two regions together and perform cross validation on the entire data set.

Table 2 presents the results. The single accuracy is quite excellent: our model predicts the correct crop in 70–75% of the cases. The set accuracy is even better, around 80–90%. 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. If we split our analysis by soil and previous crop, we find certain combinations where the determinacy is much lower. Due to space constraints we omit this analysis here, as it only affects determinacy in a substantial way. Indeed, we can already now tell that the set accuracy will on average remain about 80–90%, which indicates that the model performs very well. 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.

To assess the predictive power of the model, we compared our multicategorical logistic model with a much simpler multinomial model, without covariates, using the imprecise Dirichlet model [18] with $s = 2$ for our priors. Due to the amount of observations in our data set, this 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, and 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, in our view, 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.

6 Policy Example

An important use of land use modelling is to aid policy makers. Changes in policy affect farmer's decisions, and so land use models can predict the impact of these changes. As mentioned in Section 1, there is an interest in the UK in increasing the amount of legumes being grown. Changes in government policy can help to achieve this.

To inform policy makers, we 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. We emphasize that we have not built a causal model [14], thus one must be wary not to give too strong an interpretation to the inferences presented here.

Both legumes and rapeseed are break crops, so we are particularly interested in transitions from wheat, depicted in the bottom three plots of fig. 3. We see that, for all soil types, as nitrogen prices increase, the amount of legumes grown after wheat increases too. We use these three plots in our policy example.

There is perhaps a more obvious way to approach this problem. The usual way a government would aim to increase levels of legumes is by offering a subsidy to grow them. We have the data available to us to attempt this. By including profit margin as a regressor, we performed an analysis where we altered the profit margin of legumes relative to rapeseed. One would expect that as legumes became relatively more profitable, for example through increased subsidy, more

farmers would plant legumes as a break crop instead of rapeseed. However, the results in fact showed the opposite happening.

One potential explanation for this is the format of the data. The profit data we use [12] is actually the predicted profit for the next year. We use this as that is the information farmers have available when making their decision. As such, if there is expected to be an increase in legumes for the next year, then because of supply and demand, there may be a predicted decrease in the profitability of legumes. As such, we suspect there is a confounding variable. In fact, using nitrogen price directly produces more sensible results. Although this makes the analysis less intuitive, for this reason, we proceed with nitrogen price directly.

We are interested in analysing how a farmer's decision responds to changes in nitrogen price. Thus, we assume that the policy maker has some control over the nitrogen price, and we analyse the decision problem from the policy maker's point of view (rather than the farmer's). If the policy maker can specify utilities for different outcomes, then we can use these utilities to make a specific recommendation as to which nitrogen price achieves the best expected utility. In our robust Bayesian setting, we investigate the effect of a wide range of priors on the optimal decision. As legumes 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.

For the purpose of this paper, we choose a very simple form for the utility function:

$$U(a,b) = 100a - \kappa b \tag{14}$$

where a is the fraction of legumes across all farms, b is the nitrogen price, and κ is chosen to control how this price is weighed against the level of legumes. 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. Also, we do not fix any particular value for κ; instead, we investigate our decision problem across a range of κ values.

As before, 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

$$a(\beta^*, b) := \frac{\exp(\beta^*_{ij} \cdot (1, r, b))}{\sum_{h=1}^{J} \exp(\beta^*_{ih} \cdot (1, r, b))} \tag{15}$$

where $(1, r, b)$ is x; r is rainfall, which for the purpose of this analysis is kept fixed. Varying r makes no substantial difference to the conclusions of our study. As here we are only interested in transitions from

wheat to legumes, i represents wheat and j represents legumes. The (approximate) optimal decision is then

$$\arg \max_{b \in [b_1, b_2]} U(a(\beta^*, b), b) \qquad (16)$$

where $a(\beta^*, b)$ is the fraction of legumes in the model with MAP parameter β^* and nitrogen price b.

In our robust setting, we actually have a set B^* of β^* values. We use interval dominance, due to the simplicity by which it can be computed and graphically represented. Specifically, with

$$\underline{U}(b) := \inf_{\beta^* \in B^*} U(a(\beta^*, b), b) \qquad (17)$$

$$\overline{U}(b) := \sup_{\beta^* \in B^*} U(a(\beta^*, b), b) \qquad (18)$$

all $b \in [b_1, b_2]$ that satisfy

$$\overline{U}(b) \geq \max_{b \in [b_1, b_2]} \underline{U}(b) \qquad (19)$$

are deemed optimal. We have taken the values b_1 and b_2 to be the lowest and highest observed historical nitrogen price. These are the values our model is built on. Therefore, in our decision problem we vary nitrogen price over the range of values we have previously observed. Figure 5 shows $[\underline{U}(b), \overline{U}(b)]$ when moving from wheat on each soil type and for various values of κ. The horizontal black line represents $\max_{b \in [b_1, b_2]} \underline{U}(b)$. Values of b for which $\overline{U}(b)$ lies above this line are optimal by interval dominance. 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 fig. 3.

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, the values of b which are optimal are high.

As we increase κ, eq. (14) 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. For example, on light soil this occurs at $\kappa = 0.02$.

Eventually we reach a stage where a high nitrogen price is highly undesirable for society, regardless of the benefits that it brings with respect to increased legume growth. For example, on medium soil and $\kappa = 0.07$, only b values less than 100 are optimal.

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 the job would be to determine how to alter the nitrogen price to suit society's needs. For example, on heavy

soil with $\kappa = 0.06$, a high nitrogen price is beneficial to society. As such, a government could increase tax on nitrogen to increase the price of it. On the other hand, for heavy soil and $\kappa = 0.2$ government could decrease tax on nitrogen.

We stress again that the above analysis is purely hypothetical. We made unrealistic assumptions, and made no attempt at modelling causal relationships, so the conclusions drawn above in no way represent realistic policy proposals. Instead we demonstrated mathematical techniques for aiding policy making. Only if we had suitable data, a suitable utility function, and a suitable choice of causal covariates, could we draw hard policy conclusions from the results.

7 Summary and Conclusions

In this paper we further developed the previously proposed land use model from [13]. The model uses multinomial imprecise logistic regression with sets of conjugate prior distributions, on a non-stationary stochastic process. We obtained robust Bayesian bounds on the posterior transition probabilities of growing wheat, legumes, rapeseed or anything else, as functions of rainfall and nitrogen price. Compared to previous work we trained our model on a much larger data set. We addressed numerical stability issues by use of a near vacuous set of priors to bound probabilities away from zero.

We validated our model in two ways: comparing a nonparametric estimate with the robust Bayesian interval estimate, and by performing cross-validation. The results show that our model performs well, particularly in areas where there are few observations.

We formulated and answered a hypothetical decision problem with real-world relevance. We investigated what level of nitrogen price is most beneficial to society to promote legume growth. We used interval dominance to identify optimal policies due to its graphical representability and computational simplicity. We demonstrated how land use modelling can aid policy makers, and how imprecise probability can help to solve real world problems.

On a critical note, we may wonder about what is the advantage of using an imprecise probability model as opposed to a precise non-parametric model, or a precise Bayesian model? Indeed, confidence intervals on the parameters could be easily obtained through the non-parametric model that we introduced in eq. (10)— albeit with all the issues that come with such estimates particularly in regions where the data is sparse and where we do not believe that eq. (10) is accurate. Similarly, credible intervals could be obtained
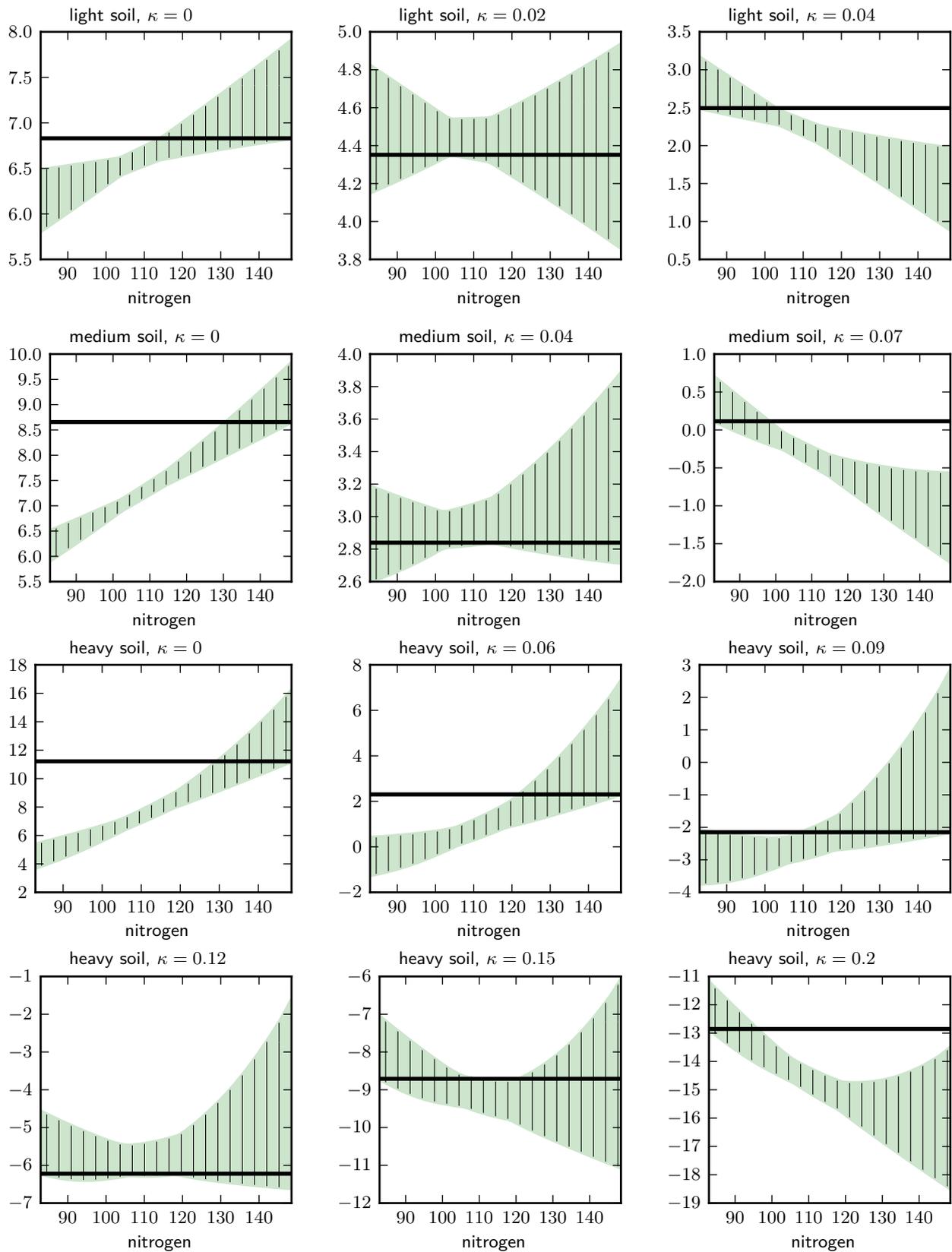
Figure 5: $[\underline{U}(b), \overline{U}(b)]$ when moving from wheat to legumes on all soil types, for various values of κ. Utility lies on the y axis.

through MCMC on just a single prior. However, for decision making, we need expected utility (or, loss), not confidence intervals or credible intervals. A precise Bayesian model always gives an exact expectation, and one would still worry about sensitivity against the prior, thereby ending up doing exactly what we do in the paper. Moreover, it is well known that the simplest way to find admissible frequentist decisions goes through a robust Bayesian analysis [17]. So, frequentists should find our analysis also quite appealing, provided they accept the parametric model.

Future work will concentrate on analysing decision problems in a more realistic way. Our data set is quite old—after 2004 field level data was not collected in the UK. However, it is planned to start again in the near future, meaning the model can be built on more relevant data. We plan on obtaining legume subsidy price, and including that as a regressor to see if that stops the confounding error discussed in section 6. The profit margin of a crop is simply a function of various factors, including subsidy level. Thus, including subsidy directly in the model as a regressor will be straightforward. We also plan to investigate other decision criteria, such as maximality and E-admissibility, particularly when interval dominance leads to vacuous decisions. The utility function could also be enhanced to account for risk aversion, and other factors that influence the benefits to society.

Acknowledgements

References

[1] Data collected by DEFRA for the index of the purchase prices of the means of agricultural production.

[2] Legume futures report. `http://www.legumefutures.de/`. Accessed: 18/02/2015.

[3] Alan Agresti. *Categorical Data Analysis*. John Wiley and Sons, third edition, 2013.

[4] Ming-Hui Chen and Joseph G. Ibrahim. Conjugate priors for generalized linear models. *Statistica Sinica*, 13:461–476, 2003.

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

[6] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, October 2009. doi: 10.1017/S0269964809990039.

[7] C. L. Gilbert and C. W. Morgan. Food price volatility. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1554):3023–3034, 2010. doi: 10.1098/rstb.2010.0139.

[8] Erik Steen Jensen, Mark B. Peoplese, Robert M.Boddey, Peter M. Gresshoff, Henrik Hauggaard-Nielsen, Bruno J.R. Alves, and Malcolm J. Morrison. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. a review. *Agronomy for Sustainable Development*, 32 (2):329–364, 2012. doi: 10.1007/s13593-011-0056-7.

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

[10] MET. Data collected by the Met Office. `http://www.metoffice.gov.uk/climate/uk/stationdata/`. Accessed: 11/02/2013.

[11] Thomas Nemecek, Julia-Sophie von Richthofen, Gaëtan Dubois, Pierre Casta, Raphaël Charles, and Hubert Pahl. Environmental impacts of introducing grain legumes into european crop rotations. *European Journal of Agronomy*, 28:380–393.

[12] John Nix. *Farm Management Pocketbook*. Agro Business Consultants Ltd., 1993–2004.

[13] Lewis Paton, Matthias C. M. Troffaes, Nigel Boatman, Mohamud Hussein, and Andy Hart. Multinomial logistic regression on Markov chains for crop rotation modelling. In Anne Laurent, Oliver Strauss, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Proceedings of the 15th International Conference IPMU 2014 (Information Processing and Management of Uncertainty in Knowledge-Based Systems, 15–19 July 2014, Montpellier, France)*, volume 444 of *Communications in Computer and Information Science*, pages 476–485. Springer, 2014. doi: 10.1007/978-3-319-08852-5_49.

[14] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009. doi: 10.1214/09-SS057.

[15] RPA. Data collected by the Rural Payments Agency under the integrated administration and control system for the administration of subsidies under the common agricultural policy.

[16] Matthias C. M. Troffaes and Lewis Paton. Logistic regression on Markov chains for crop rotation modelling. In F. Cozman, T. Denœux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 329–336, Compiègne, France, July 2013. SIPTA. URL `http://www.sipta.org/isipta13/index.php?id=paper&paper=033.html`.

[17] Abraham Wald. Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, 10(4):299–326, December 1939. doi: 10.1214/aoms/1177732144.

[18] Peter Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58(1):3–34, 1996. URL `http://www.jstor.org/stable/2346164`.