A Parcel Level Model for U.S. Agricultural Land Use

Jonathan Lisic*

Abstract

Agricultural land use prediction at the parcel level is important in many applications, such as natural resource management, survey development, and agricultural economics. Unfortunately, in the United States no complete set of agricultural land use parcels covering the entire contiguous 48 states are publically available. Here a computationally feasible method for forming and classifying parcels from geospatial data sources using a modified mean shift algorithm is presented. These parcels are incorporated into a Bayesian hierarchical multinomial probit model that predicts agricultural land use. This prediction is based on crop rotations. Crop rotations are sequences of land use where current land use is dependent on prior land use. These crop rotations provide a useful temporal process to assist in prediction. A novel component of this model is the incorporation of spatially correlated crop rotation preferences through a spatial autoregressive prior. This research is of immediate use to the United States Department of Agriculture's National Agricultural Statistical Service (NASS) for survey planning and data collection in the June Agricultural Survey, and may provide a useful tool in further research in agricultural land use.

Key Words: Spatial Autoregression, GIS, Agriculture, Multinomial Probit, Mean Shift Clustering, Crop Rotation, Bayesian Predictive Inference

1. Introduction

Prediction of land use at the parcel level is useful for numerous purposes, including: agricultural production; natural resource management [Thenail et al., 2009]; survey development [Zimmer et al., 2012]; predicting and measuring changes in the local environment [Castellazzi et al., 2007]; residential land conversion; and land use response to local market changes, such as the establishment of ethanol plants [Livingston et al., 2008], [Livingston et al., 2012]. In this article, an approach to predicting future land use through an areal spatial-temporal autoregressive probit model will be presented. This approach relies on image segmentation to construct parcel-like units from high resolution imagery and an agricultural production process known as *crop rotations*.

Given a set of parcels, the prediction of future land use can be facilitated by the agricultural practice of crop rotations. Crop rotations are a repeating sequence of crops used to promote yield while retaining soil quality and mitigating pests and disease. Common rotations in the United States include corn to soybeans to corn, which has been studies by [Livingston et al., 2012] and others. In the context of parcel-based models, crop rotations are especially useful because they are a prevalent, locally observable phenomenon. The prevalence of these rotations was studied by Sahajpal et al. [2014] where it was noted that the U.S. cropping patterns could be represented with 90% accuracy by 82 distinct three year sequences.

Modeling agriculture via crop rotations is certainly not novel and has seen employment in simulations and agricultural econometric models. Simulation models have been used by Schonhart et al. [2011], Castellazzi et al. [2008] and Detlefsen

*National Agricultural Statistical Service United States Department of Agriculture, 1400 Indepence Ave. S.W., Room 6412B, Washington, DC 20250

[2004] who utilized stochastic matrices to model rotations for constructing landscapes and planning. One drawback to these models is that little attention is given to the estimation of transition parameters, and due to the simple structure of the stochastic matrices, it is not possible to directly admit covariates.

The framework of agricultural economic models of crop rotations has been explored at great length by [Hennessy, 2006]. This framework is based on expected return, which can not be practically applied at a parcel level due to return being a function of yield, which is highly volatile even at the county level for many crops. Application to optimal crop rotations of corn and soybeans has been explored by [Livingston et al., 2012], and a more general framework has been considered by [Cai et al., 2013]. Simplifications to [Hennessy, 2006] have been considered by [Ji and Rabotyagov, 2015].

The most similar approach in the literature is the graphical model based approach [Osman et al., 2015]. In this approach the spatial units are provided through the French Registre Parcellaire Graphique Land Parcel Identification System (RPG LPIS). Because parcels provided by RPG LPIS may include multiple land uses, the authors chose to restrict the modeling units to the subset of parcels that contain only a single crop and field, potentially introducing bias. Prediction in this model is through a graphical modeling approach known as Markov Logic Network (MLN) Richardson and Domingos [2006]. A MLN is similar to other computer-science-based approaches to learning networks based such as a Bayesian Network (BN) with the exception that the graphs are not directional. In the MLN, each node is a random variable with the Markov property, conditionally independent of other nodes in the graph given its neighbors. Parameters for the model are updated through frequencies of observed logic statements holding true, such as a particular land use sequence.

This MLN is inherently non-spatial, does not readily accept continuous covariate information, and depends on the existence of classified parcels. The approach in this paper resolves these issues. The dependence on pre-classified parcels is removed through constructing parcel-like units, land cover units (LCUs) that contain a single land use, avoiding removal of spatial units containing multiple fields. Covariate inclusion and spatial dependence are handled through a spatial-temporal probit approach to model the rotation. The probit provides a straight forward way to add covariates and produce statistically meaningful parameter estimates. This approach breaks the land use prediction into three tasks:

1. Image Segmentation - to estimate LCU boundaries from high resolution imagery.

2. Classification - to classify LCU estimates based on coarse, pre-classified pixels and other exogenous data sources.

3. Prediction - to predict LCU land use contents using crop rotations through a spatially auto-regressive process.

Through this paper, each of these tasks will be addressed independently.

2. Image Segmentation

A *parcel* is defined by McDonnell and Kemp [1995] as a region of land usually delineated by ownership or land use. This boundary could define many possible spatial units, such as a city block, a subdivision, or an agricultural field. The ideal

parcel for land use modeling is one that has a maximal boundary surrounding a single land use type. A set of these partitions and associated land use can fully describe a region of space $R \subset \mathbb{R}^2 \times \mathfrak{C}^{\#(T)}$ with no loss of information, where the land use sequences will be an element of the product space of $\mathfrak{C}$, a set of land use categories, indexed by $\mathfrak{T}$, a set of years, where set cardinality is through the function $\#(\cdot)$. For simplicity, $\#(\mathfrak{C}) = C$. Unfortunately in the United States, two factors make identifying parcels challenging. (1) a lack of land use data sets that provide information on individual crops and (2) no publicly available or complete set of agricultural land use parcel data [Hua et al., 2005], [Cai et al., 2013].

To resolve the first problem, it is assumed that land use implies land cover. Land cover is what covers a region of land such as corn plants. Land use is how the land is used, such as corn being grown for commercial purposes. For this application, prediction is restricted to major crops on a county by county basis ensuring little difference between definitions.

For the second issue, Yan and Roy [2014] provide an approach to form parcels using $30m \times 30m$ LANDSAT imagery, but make neither the program nor output publicly available. A set of parcels with associated land use do exist, but only through voluntary participation, such as USDA Farm Service Agency (FSA) programs. This participation is voluntary, and the benefits for participating vary depending on land use and location. Furthermore, the FSA parcel data known as common land units (CLUs) are submitted under a confidentiality agreement and cannot be released to the general public. In this section, a land cover unit meeting the definition of the ideal parcel will be introduced; data sources used to construct these land cover units will be detailed; and image segmentation algorithm will be described.

2.1 Land Cover Unit

The observational unit will be called a *Land Cover Unit* (LCU). LCUs with at least one year of agricultural content will be refered to as ALCUs. Each LCU is the maximally contiguous section of land with respect to a single land cover sequence not transected by public transportation arteries or permanent hydrographic boundaries, where a land cover sequence is an ordered set of known categories indexed by a set of fixed consecutive years. Under this description and later formal definition, an LCU is dependent on both the temporal window and the categories specified, (see Figure 2.1).

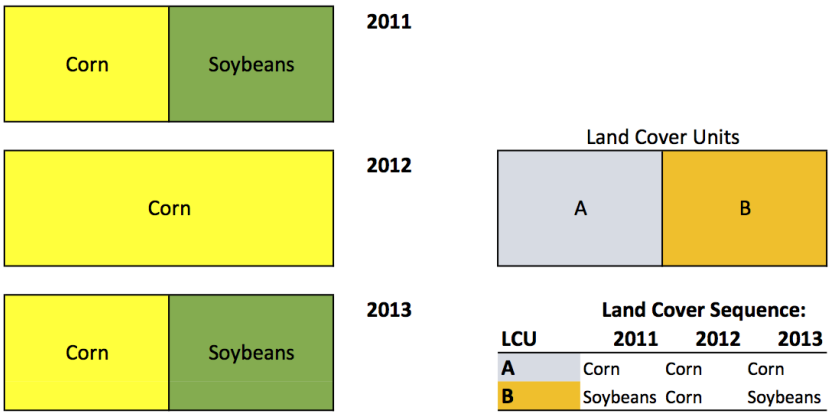

Figure 1: An example of two ALCUs.

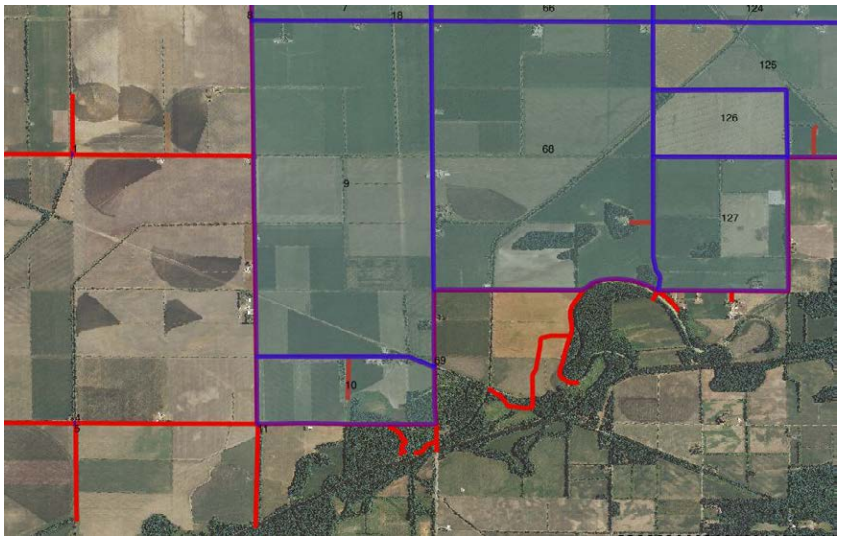

Figure 2: An example of U.S. Census Tiger GIS data plotted on top of NAIP imagery; blue shaded areas are regions and boundaries where image segmentation is applied.

To ensure LCUs can be used in GIS applications, the unit is built on a reference grid of sufficiently high resolution such that the measurement error between the area of the grid and target unit is sufficiently small. In parctice this reference grid will typically be represented by pixels from remotely sensed images.

LCUs are identified by two components. (1) a spatial index $i \in \Delta$, where Δ is the set of LCU indexes in region $R \subset \mathbb{R}^2$, and (2) a land cover sequence y_i (vector valued) for the indexed LCU. Similar to a pixel, each i identifies a subset of $\mathbb{R}^2$; however, the LCUs are not necessarily uniform in shape. As an example, an LCU y_i could rotate between land cover classes for five years, producing the sequence (C, S, C, S, C) where i indexes a subset of $R \times \mathfrak{C}^{\#\mathfrak{T}}$.

2.2 Data Sources and Problem Reduction

Although ideal land cover units are not publicly available across the entire 48 contiguous states, NAIP (National Agriculture Imagery Program) aerial images and U.S. Census Edge Data are.

- NAIP (National Agriculture Imagery Program) aerial imagery: an orthorectified 1-2 meter imagery product that covers the entire lower 48 states every 3-5 years (Freely available through the USDA/NRCS geospatial data gateway).
- U.S. Census Edge Data: a collection of polygon boundaries of geographic features (e.g. lakes), linear features such as roads and hydrography (e.g. rivers and streams) (Freely available through the US Census FTP site).

To reduce the memory requirements of segmenting large areas of land, permanent boundaries from the U.S. Census edge data are used here to allow work to be done on smaller areas of land. NAIP imagery is further processed into greyscale imagery due to no appreciable gain in image segmentation results relative to using three color bands.

2.3 Mean Shift

Mean shift (MS) is a local maxima searching algorithm similar to Newton's Method for finding roots of a smooth function but restricted to kernel density estimators. To classify using MS, a set of *query* points Q are classified by steepest ascent to a local maxima from the kernel density estimator of another set of *reference* points R where each point v in R or Q is assumed to be an observation of the random variable $\mathbf{x}$ with pdf f and support in $\mathbb{R}^d$. Q and R are assumed to both have the same support, and frequently $Q \equiv R$. Because the number of local maxima determine the number of clusters, there is no *a priori* requirement on identifying the number of distinct clusters. Instead, the choices of kernel and bandwidth parameters determine the number of local maxima in the kernel density estimate (KDE).

The MS algorithm is based on the derivative of a KDE,

$$\hat{f}_{N_R,H}(v) = \sum_{j=1}^{N_R} |H|^{-1} n^{-1} K\left(H^{-1}(v - x_j)\right) = \sum_{j=1}^{N_R} b_{j,H} k\left(||v - x_j||_H^2\right). \quad (1)$$

where

K	=	a kernel,
k	=	a function proportional to the kernel,
H	=	a symmetric positive definite bandwidth matrix used to set smoothness of the KDE,
$\|\|a_1 - a_2\|\|_H$	=	Mahalanobis distance between two vectors a_1 and a_2,
N_R and N_Q =	=	shorthands for $\#(R)$ and $\#(Q)$ respectfully, and
$b_{j,H}$	=	a constant ensuring that $g\left(\|\|v^i - x_j\|\|_H^2\right) c_{j,H} = K\left(\|\|v^i - x_j\|\|_H^2\right)$.

The derivation is surprisingly simple and intuitive. Using the KDE in (1), the derivative can be equated to zero as in (2):

$$0 = \frac{\sum_{j=1}^{N_R} x_j b_{j,H} g\left(||v - x_j||_H^2\right)}{\sum_{j=1}^{N_R} b_{j,H} g\left(||v - x_j||_H^2\right)} - v \quad (2)$$

where $g = -k'$. This leads to a fixed point iterator as in (3), which is the mean shift. Fixed point iteration is obtained through repeated application of (3) from an initial point in Q.

$$v^{i+1} = \frac{\sum_{j=1}^{N_R} b_j c_{j,H} g\left(||v^i - x_j||_H^2\right)}{\sum_{j=1}^{N_R} b_{j,H} g\left(||v^i - x_j||_H^2\right)}. \quad (3)$$

This iteration continues until either a fixed number of iterations or the difference between two iterations is below a threshold.

The MS algorithm was first proposed by Fukunaga and Hostetler [1975], but did not see wide spread adoption until the publication of Cheng [1995]. Cheng used the term "Natural Clustering" to describe the procedure, as the data points themselves self-organize around the mode of the density. This self-organization is interesting in a number of ways. First, unlike K-means where the choice of initial centroids may have a considerable influence on the final classification, in the MS algorithm, the mode that any given point in the density will shift to is entirely deterministic. Because of this, if any convergent sequence, shares any value of another convergent sequence then they converge to the same mode.

Convergence of the mean shift sequence to a stationary point was addressed by Cheng [1995]. If the stationary point of the kernel density estimate $\hat{f}$ is a local maxima, then convergence of the local maxima to a maxima of the pdf of $\mathbf{x}$ is

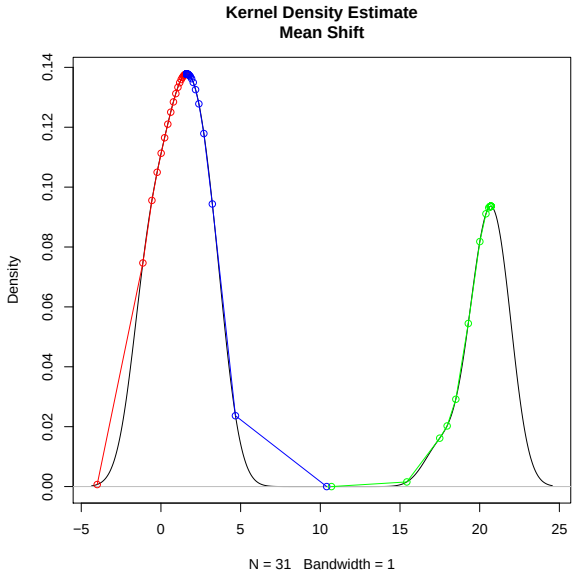

Figure 3: Simple mean shift example on a bivariate normal density sample.

provided by Chen et al. [2014] under reasonable regularity conditions. Bandwidth selection can be performed under typical rule-of-thumb estimates for derivatives of multivariate kernels or cross validation. Within this application, bandwidth was selected via cross validation using a set of 200 parcels over 15 fully delineated regions bounded by known boundaries and an adjusted Rand Index as an objective function [Rand, 1971].

A notable problem with the MS algorithm is that the algorithm tends to converge slowly and has a computational complexity of $\mathcal{O}(mN_RN_Q)$ where m is the numnber of iterations. In this application, a number of methods were used to speed up the algorithm. These includes running MS on a sample of the original image, creating a generalized iterator with properties of both the MS and Newton's Method, and constructing a dual-tree implementation of MS to greatly reduce the number of points to iterate. An overview of these enhancements are provided in the documentation of the meanShiftR package created to support this research [Lisic, 2015].

3. Classification

Land use sequences for the LCUs are estimated through classified pixel level data from the United States Department of Agriculture's Cropland Data Layer (CDL), a yearly 30-56 square meter thematic map over all contiguous 48 states from 2008 to present. The CDL is available prior to 2008 for a small number of states [Boryan et al., 2011]. Unfortunately, land use sequences cannot be directly applied from CDL pixels due to the large difference in resolution between the CDL pixels and NAIP derived boundaries.

The problem of determining LCU estimate classification from pixels is within the domain of the *polygon overlay problem*. The polygon overlay problem is a geographic problem where a set of target units are desired but the properties of these units are not available. Instead, a set of source units are available with known properties, but do not share the same areal units. In this context, the CDL pixels

are considered the source units, and the LCU estimates are the target units. The set of approaches to this problem can be found in Gotway and Young [2002]. In this research a simple initial approach within the class of pixel aggregation and areal weighting is chosen, namely mapping CDL pixels interior to an LCU's boundary to the LCU. Interior CDL pixels are used to reduce the measurement error from CDL pixels that lie between LCUs and misalignment of the geospatial data sources used to classify the CDL pixels [Dean and Smith, 2003] [Boryan et al., 2011]. LCUs that do not have sufficient dimensions to have an interior CDL pixel, are unlikely to have agricultural land use and are removed.

The boundaries suggested by the MS or other segmentation algorithms are not perfect. LCU estimates may contain portions of other LCUs or may be a proper subset of another LCU. These errors cause loss of information with respect to the LCU's land use sequence, which makes interpreting relationships between LCUs more difficult. To reduce these issues with the LCU estimators, three post-processing steps are performed after the assignment of interior CDL pixels to LCU estimates:

1. Identification of under-segmentation, forcing segmentation of an LCU estimate, currently through a heuristic.

2. Classification of land-use sequences into agricultural and non-agricultural land use through logistic regression.

3. Identification of over-segmentation, merging of LCU estimates through maximization of the joint likelihood of adjacent LCU classifications.

The sequence of these steps is important, especially to avoid the merging of non-agricultural LCUs to ALCUs. Methods to perform these three steps are currently under development and will be presented in future work.

4. Prediction

A spatial temporal multinomial probit model is used to predict agricultural land use through crop rotations. The choice of this approach was motivated by initial investigations into the spatial correlation of the number of popular crop rotations of each LCU, namely corn-to-corn and corn-to-soybean rotations in a subset of La Porte county in Indiana (Figure 4). This investigation focused on measures of spatial association through Moran's I, (Table 4), and simple plots of the prevalence of particular rotations (Figure 4). Both the Moran's I and the plots seem to suggest the presence of some spatial dependence.

A multinomial probit model was chosen due to its ability to link the categorical response (major commercial crops) to a multivariate normal linear model. This multivariate normal form is easy to extend as a hierarchical model and has a simpler error structure than multinomial logits and related approaches. Within this section, a brief introduction to multinomial probit models and spatial autoregressive models will be presented followed by the crop rotation model.

Rotation	Moran's I	p-value
Corn-to-Corn	0.25	0.005
Corn-to-Soy	0.24	0.005

Table 1: Measures of spatial association for specific rotation counts at each LCU under a 'Rook' based neighborhood.

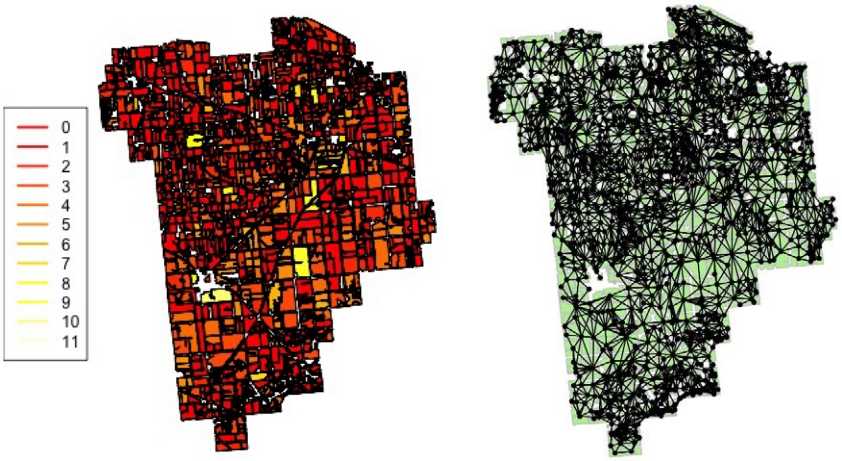

Figure 4: **Left:** number of Corn-to-Corn Rotations 2001-2013 in a subset of La Porte County Indiana. **Right:** initial spatial neighborhood using Rook based association.

4.1 Multinomial Probit Model

Multinomial probit models provide a link function between a categorical response and a $J = C-1$ dimensional linear model with a multivariate normal error structure. This linking is performed by partitioning the support of the linear model (Figure 5),

$$\mathbf{y}_i = \begin{cases} c = 1 & \max\left(\mathbf{z}_{i,j}\right) \leq 0 \qquad \forall j \in \mathfrak{J} \\ c = j+1 & \text{argmax}_{j \in \mathfrak{J}} \left\{ \mathbf{z}_{i,j} \mathbb{I}_{\{\mathbf{z}_{i,j} > 0\}} \right\}) \end{cases} \tag{4}$$

where the observation of category $c \in \mathfrak{C}$ is determined by $z_{i,j}$ is the j^{th} element of the latent vector z_i with $j \in \mathfrak{J} = \{1, \ldots, J\}$ [McCulloch and Rossi, 1994].

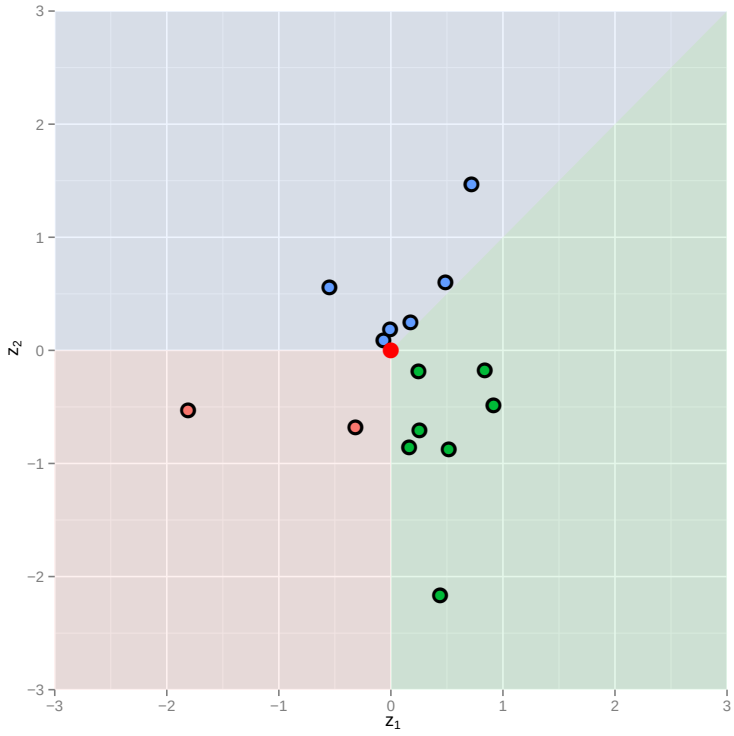

Figure 5: An example of data augmentation in MNP, intercept in red and other points represent deviates generated from the latent variable.

Inference in multinomial probit models is conducted primarily through the data augmentation within the Bayesian paradigm, an extension of the univariate approach suggested by Albert and Chib [1993]. In this approach, the latent variable $\mathbf{z}$ is simulated conditioned on the observed states of $\mathbf{y}$. Some notable issues are present with these MNP models, namely the linear model parameters are not identifiable,

$$\Pr\left(\mathbf{z}_{i,j} > a\right) = \Pr\left(\alpha\mathbf{z}_{i,j} > \alpha a\right), \ \alpha \in \mathbb{R}. \tag{5}$$

The standard approach is to fix the first element of the covariance structure to one as detailed in McCulloch et al. [2000]. However, a strict restriction may create excessive correlation between subsequent draws from the posterior distributions of parameters when using Gibbs sampling, requiring a great deal of thinning and longer run times. The approach used by [Imai and van Dyk, 2005] known as marginal data augmentation provides a more computationally efficient method. They reduce correlation by introducing a "working parameter" that allows draws from unrestricted posterior distributions. The working parameter α is used to scale the unidentified scale and location parameters of the latent variable and then is marginalized out of the posterior distribution.

The multivariate extension of the univariate approach of Albert and Chib [1993] was first suggested by McCulloch and Rossi [1994] with latent variable form

$$Z^* = U\beta + \boldsymbol{\epsilon}, \tag{6}$$

where

β = a vector of coefficients of length p,
$\boldsymbol{\epsilon}$ $\sim$ $\mathcal{TN}\left(0, \Sigma, A_j, B_j\right)$ where A_j and B_j are upper and lower bounds respectfully defined by class j, and
U = an $nJ \times p$ matrix of covariates.

4.2 Spatial Autoregressive Multinomial Probit Models

Two common spatial autoregressive Gaussian models are the simultaneous autoregressive (SAR) Whittle [1954] model and the conditional autoregressive (CAR) Besag [1974] model. Of these two forms, the SAR model is more popular in MNP models [LeSage and Pace, 2009] [Wang et al., 2012]. The standard SAR model has the following form (called the spatial error model (SEM) by LeSage and Pace [2009])

$$\mathbf{Z} = B\mathbf{Z} + (I - B)U\beta + \boldsymbol{\epsilon}, \tag{7}$$

where

$\mathbf{Z}$	=	is a vector of length n,
U	=	is an $n \times p$ vector of covariates,
β	=	is a length p vector of location parameters,
$\boldsymbol{\epsilon}$	=	is a random vector composed of length n with distribution $\mathcal{N}(0, \Sigma)$ and
B	=	is an $n \times n$ matrix of spatial weights.

Under this approach Z is distributed as

$$\mathbf{Z} \sim \mathcal{N}\left(U\beta, (I - B)^{-1}\,\Sigma\left(I - B^{T}\right)^{-1}\right) \tag{8}$$

In practice, the $n \times n$ weight matrices B is usually simplified to ρW, where ρ is a scalar and W is an $n \times n$ matrix of fixed spatial weights. The matrix W is sometimes referred to in the literature as a spatial adjacency matrix. In the simplest of cases, pairwise neighbors in this matrix are set to one while all other values are set to zero, and to provide a more natural weighting of observations each row is scaled to sum to one. The scalar ρ is sometimes called an autocorrelation parameter, but as noted by Wall [2004], the value does not necessarily indicate any particular level of autocorrelation. The covariance matrix of $\mathbf{Z}$ is of full rank when $\rho \in \{\rho : \rho < 1/\lambda_i\}$ $i \in \{1, \ldots n\}$ where λ_i is an eigenvalue of W. In this research, the value of W has been set to a Rook based approach, but using a fixed buffered region of 50 meters to determine adjacency. This is essential to allow for neighbors across streets (Figure 4).

Spatial autoregressive MNP models or SAR MNP models have been explored by LeSage and Pace [2009], with temporal spatial approaches developed by Wang et al. [2012]. Although these models are called SAR models they differ from (7),

$$\mathbf{Z}^{*} = B\mathbf{Z}^{*} + U\beta + \boldsymbol{\epsilon}, \tag{9}$$

where

B	=	$\rho W \otimes I_J$,
W	=	$n \times n$ matrix of spatial weights (Rook based in this application), and
ρ	=	scalar parameter for W.

These models are sometimes called "mixed regressive and spatial autoregressive model" or "lagged response model" in ecology.

Estimation of parameters in these models follows a Bayesian approach, and posterior distributions are provided through Gibbs sampling. The most computationally burdensome part of this approach is in the drawing from the conditional distribution of ρ given β, Σ, and Y. Since the distribution of ρ does not follow any standard form deviates are generated either by interpolation Metropolis-Hastings, or through interpolation and inversion of the CDF, details in LeSage and Pace [2009].

5. Spatial-Temporal Model for Crop Rotation

The MNP and SAR MNP models provide a way to link categorical response to a linear model with multivariate normal error structure. These models still need to be linked with the crop rotation phenomena used for prediction. To do this, it is assumed that the categorical response is temporally stationary and conditionally independent given a prior state, sequence of prior crops, on the same LCU. Therefore, by specifying a set of prior states (rotations) P, it is then possible to create a design matrix to include the prior state information for each LCU. This design matrix is both large and sparse with dimension $nJ \times nJp$ so special care should be taken in handling this matrix for computation.

In the proposed crop rotation model, it is also assumed that the crop rotations are spatially correlated, e.g. corn-to-soybean rotations are close to corn-to-soybean rotations. This model differs from the SAR MNP models of LeSage and Pace [2009] and Wang et al. [2012], where the response is considered auto-correlated, e.g. corn grows next to corn. The autocorrelated crop rotation model approach has the form

$$Z^* = U^*\boldsymbol{\beta}^* + X\gamma + \boldsymbol{\epsilon}, \tag{10}$$

$$\boldsymbol{\beta}^* = B\boldsymbol{\beta}^* + (I - B)\beta_0 + \boldsymbol{\upsilon}, \tag{11}$$

where

U^*	=	$nJ \times nJp$ design matrix,
$\boldsymbol{\beta}^*$	=	npJ vector of covariates,
X	=	other covariates,
γ	=	other covariates coefficients,
β_0	=	hyper parameter for β_0 and
$\boldsymbol{\upsilon}$	=	random vector of length n with distribution $\mathcal{N}(0, \Sigma_\upsilon)$.

A problem with this model is the large number of parameters introduced; however, for larger values of ρ, the number of effective parameters is actually much lower than npJ, avoiding excessive over-fitting.

Parameter estimation as in other SAR MNP approaches is through Gibbs sampling, however unlike LeSage and Pace [2009] and Wang et al. [2012] that use conditional data augmentation, the more computationally efficient data augmentation method of Imai and van Dyk [2005] is used. Due to the SAR component of the model only interacting with the rest of the model through β the implementation is fairly straightforward.

Using marginal data augmentation, the identified parameter $\tilde{\beta}$ under a Bayesian approach has a conditional distribution for Gibbs sampling similar in form to Algorithm 2 in Imai and van Dyk [2005] when $\beta_0 \neq 0$ and Algorithm 1 when $\beta_0 = 0$; that is,

$$\tilde{\boldsymbol{\beta}}|\Theta_{-\beta} \sim \mathcal{N}\left(\hat{\beta}, (\alpha^2)^* \left(A + \sum_{t\in\mathfrak{T}} U^T\Sigma^{-1}U\right)^{-1}\right) \tag{12}$$

where,

$$\hat{\beta} = \left(A + \sum_{t\in\mathfrak{T}} U^T\Sigma^{-1}U\right)^{-1}\left(U^T\Sigma^{-1}Z + A\beta_0\right), \tag{13}$$

A is $\left((I - B)^{-1}\Sigma_\nu(I - B^T)^{-1}\right)^{-1}$ and, Θ_{-a} is the set of parameters minus parameter a. $\boldsymbol{\Sigma}$, Z, and $(\alpha^2)^*$ all follow their respective conditional distributions in Imai and van Dyk [2005]. ρ from $B = \rho W \otimes I_{np}$ follows the same distribution as in

LeSage and Pace [2009], and implemented using the sparse techniques found in Bivand [2015].

To test the convergence of the Gibbs sampling, the process as described in Imai and Van Dyk [2005] was followed. In this process, three MCMC chains from the gibbs sampler were generated with different initial parameters, and compared against each other through the Gelman-Rubin convergence diagnostic statistic [Gelman and Rubin, 1992].

6. Results

The subset of La Porte county that was initially segmentated and explored through spatial analysis was used to provide some initial results. This county subset consisted of 1486 ALCUs.

An implementation of this model was based on a fork of [Imai and Van Dyk, 2005] with a large amount of work spent on developing sparse matrix handling. The results were compared against a non-spatial and a naïve Bayes model where the later approach did not take into account any rotation, while the former considered a small number of two year rotations based on popularity (Table 6). The categories considered for this model include corn, soybeans, winter wheat, non-agriculture and other agriculture.

Rotations	Percent of Rotations
Corn → Corn	14.9%
Corn → Soybeans	23.8%
Soybeans → Corn	24.2%
Non-Agriculture → Non-Agriculture	12.5%
Other	23.6%

Table 2: Popular rotations in a subset of La Porte county Indiana from 2001 to 2011.

The run time for the image segmentation was four hours, including post-processing. Run time for parameter estimation was six hours per chain, for a total of three chains. Only the first chain was used for predicting the subsequent year via MAP The prediction step used 1000 iterations, with a 200 iteration burnin, and performed Gibbs sampling using an unoptimized Metropolis Hastings routine for the conditional distribution of ρ. Because of excessive correlation between draws from the Metropolis Hastings routine, only one out of every 30 draws were retained, no thinning was needed for other parameters. All processing was performed on three quad core (i7-4790K) computers with 32GB of RAM each, utilizing an ad-hoc *map reduce* approach through job scheduling on a PostGIS database.

Estimated LCUs were compared against hand drawn LCUs and FSA CLU data with an initial adjusted Rand Index of 0.73. Predictions were compared, conditional on the LCU land-use sequences, using an integrated misclassification rate (IMR),

$$IMR = 1 - \frac{\sum_{i \in D} \mathbb{I}_{\hat{y}_{i,j}=y_{i,j}} a_i}{\sum_{i \in D} a_i} \tag{14}$$

where a_i is the area of ALCU i.

		IMR	
Year	Naïve Bayes*	Pooled MAP	Spatial MAP
2010	0.62	0.25	0.29
2011	0.64	0.59	0.54
2012	0.62	0.68	0.63
2013	0.60	0.47	0.44

Table 3: Integrated misclassification rate for ALCU in La Porte County Indiana for five classes (corn, soybeans, winter wheat, non-agriculture and other agriculture).

7. Discussion and Future Work

This article provides a basic method for predicting U.S. Agriculture at the parcel level. This prediction is made possible through image segmentation through an accelerated mean shift implementation. A spatial autoregressive hierarchical model was presented, providing spatially correlated coefficients to represent locally similar rotation choices. Initial testing in La Porte County, Indiana shows, on average, a 2% decrease in misclassification, relative to a non-spatial model. In the current research, only La Porte County, Indiana has been explored; future work will involve testing this method over more diverse and differing crop types. Future improvements to this method include development and testing of post processing tasks to improve the initial image segmentation, evaluation of different spatial neighborhoods for the SAR model as suggested by Bivand and Gomez-Rubio [2008], and identification of useful exogenous data to use as covariates.

References

Albert, J. H. and Chib, S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Besag, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

Bivand, R. Spatial weights objects as sparse matricies and graphs. 2015. URL https://cran.r-project.org/web/packages/spdep/vignettes/nb_igraph.html. R vignette.

Bivand, R. and Gomez-Rubio, V. *Applied spatial data analysis with R*. Springer, New York, 2008. ISBN 9780387781716.

Boryan, C., Yang, Z., Mueller, R., and Craig, M. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 26(5):341–358, 2011.

Cai, R., Mullen, J. D., Wetzstein, M. E., and Bergstrom, J. C. The impacts of crop yield and price volatility on producers' cropping patterns: A dynamic optimal crop rotation model. *Agricultural Systems*, 116:52–59, 3 2013. ISSN 0308521X. doi: 10.1016/j.agsy.2012.11.001.

Castellazzi, M. S., Matthews, J., Wood, G. A., Burgess, P. J., Conrad, K. F., and Perry, J. N. Landsfacts: Software for spatio-temporal allocation of crops to fields. In *Proceedings of 5th Annual Conference of the European Federation of IT in Agriculture, Glasgow, UK*, 2007.

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(1-2):26–33, 4 2008. ISSN 0308521X. doi: 10.1016/j.agsy.2007.10.006.

Chen, Y.-C. . C., Genovese, C. R., and Wasserman, L. Generalized mode and ridge estimation. *arXiv preprint arXiv:1406.1803*, 2014.

Cheng, Y. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8):790–799, 1995.

Dean, A. M. and Smith, G. M. An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing*, 24(14):2905–2920, 2003.

Detlefsen, N. Crop rotation modelling. In *Proceedings of the EWDA-04 European workshop for decision problems in agriculture and natural resources. Silsoe Research Institute, England*, pages 5–14, 2004.

Fukunaga, K. and Hostetler, L. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1):32–40, 1975.

Gelman, A. and Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Gotway, C. A. and Young, L. J. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648, 2002.

Hennessy, D. A. On monoculture and the structure of crop rotations. *American Journal of Agricultural Economics*, 88(4):900–914, 2006.

Hua, W., Sohngen, B., and Hite, D. Assessing the relationship between crop choice and land use change using a markov model. In *Selected paper for prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July*, pages 24–27, 2005.

Imai, K. and van Dyk, D. A. A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of econometrics*, 124(2):311–334, 2005.

Imai, K. and Van Dyk, D. A. Mnp: R package for fitting the multinomial probit model. *Journal of Statistical Software*, 14(3):1–32, 2005.

Ji, Y. and Rabotyagov, S. Estimating adoption of cover crops using preferences revealed by a dynamic crop choice model. *AAEA annual meeting, Western Agricultural Economics Association, San Francisco, CA.*, 2015.

LeSage, J. P. and Pace, R. K. *Introduction to spatial econometrics*. CRC Press, Boca Raton, 2009. ISBN 142006424x.

Lisic, J. meanshiftr. https://github.com/jlisic/meanShiftR, 2015.

Livingston, M., Roberts, M. J., and Rust, J. Optimal corn and soybean rotations. In *AAEA annual meeting, Orlando, Florida*, pages 27–29, 2008.

Livingston, M., Roberts, M. J., and Zhang, Y. Optimal sequential plantings of corn and soybeans under price uncertainty. *North Carolina State University, Working paper*, 2012.

McCulloch, R. and Rossi, P. E. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1):207–240, 1994.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99(1):173–193, 2000.

McDonnell, R. and Kemp, K. K. *International GIS dictionary*. John Wiley & Sons, 1995.

Osman, J., Inglada, J., and Dejoux, J.-F. Assessment of a markov logic model of crop rotations for early crop mapping. *Computers and Electronics in Agriculture*, 113:234–243, 2015.

Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

Richardson, M. and Domingos, P. Markov logic networks. *Mach Learn*, 62(1-2): 107–136, 1 2006. ISSN 0885-6125. doi: 10.1007/s10994-006-5833-1.

Sahajpal, R., Zhang, X., Izaurralde, R. C., Gelfand, I., and Hurtt, G. C. Identifying representative crop rotation patterns and grassland loss in the us western corn belt. *Computers and Electronics in Agriculture*, 108:173–182, 2014.

Schonhart, M., Schmid, E., and Schneider, U. A. Croprota - a crop rotation model to support integrated land use assessments. *European Journal of Agronomy*, 34 (4):263–277, 5 2011. ISSN 11610301. doi: 10.1016/j.eja.2011.02.004.

Thenail, C., Joannon, A., Capitaine, M., Souchere, V., Mignolet, C., Schermann, N., Di Pietro, F., Pons, Y., Gaucherel, C., Viaud, V., and Baudry, J. The contribution of crop-rotation organization in farms to crop-mosaic patterning at local landscape scales. *Agriculture, Ecosystems & Environment*, 131(3-4):207–219, 6 2009. ISSN 01678809. doi: 10.1016/j.agee.2009.01.015.

Wall, M. M. A close look at the spatial structure implied by the car and sar models. *Journal of Statistical Planning and Inference*, 121(2):311–324, 4 2004. ISSN 0378-3758. doi: 10.1016/s0378-3758(03)00111-3.

Wang, X. C., Kockelman, K. M., and Lemp, J. D. The dynamic spatial multinomial probit model: analysis of land use change using parcel-level data. *Journal of Transport Geography*, 24:77–88, 2012.

Whittle, P. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

Yan, L. and Roy, D. P. Automated crop field extraction from multi-temporal web enabled landsat data. *Remote Sensing of Environment*, 144:42–64, 3 2014. ISSN 0034-4257. doi: 10.1016/j.rse.2014.01.006.

Zimmer, S., Kim, J. K., and Nusser, S. Automatic stratification for an agricultural area frame using remote sensing data. *JSM Proceedings. American Statistical Association-IMS. San-Diego, CA.*, 2012.