IIASA

Global high-resolution land-use change projections: A Bayesian multinomial logit downscaling approach incorporating model uncertainty and spatial effects

Tamás Krisztin (IIASA), Petr Havlík (IIASA), David Leclere (IIASA), Inés Moreau (UC Louvain)

Abstract

Using econometric models to estimate land-use change has a long tradition in literature. Recent contributions show the importance of including spatial information and of using a multinomial framework to take into account the inter-dependencies between the land-use classes. Few studies, however, agree on the relevant determinants of land-use change and there are no contributions so far comparing determinants on a global scale. Using multiple datasets of land use change between 2000 and 2010 – standardized to 5 arc minute resolution – and taking into account the transitions between forest, cropland, grassland and all other land covers, we estimate a Bayesian multinomial logit model, using the efficient Pólya-Gamma sampling procedure introduced by Polson et al. (2013). To identify and measure the determinants of land-use change and the strength of spatial separation, our model implements Bayesian model selection through stochastic search variable selection (SSVS) priors and flexible spatial lags of the explanatory variables.

In a second step, we combine our parameter estimates with aggregate, supra national land-use change results from the partial equilibrium agricultural model GLOBIOM and project our model in ten-year intervals up to 2100 on a spatially explicit scale along multiple shared socioeconomic pathways.

1. Introduction

GLOBAL agricultural models such as GLOBIOM are calibrated to provide realistic projections on country or above (regional) level, even if they include spatial dynamics at a finer resolution. This is due to computational and calibrational constraints. There is a strong interest, however, in exploring and visualizing agricultural land change projections at a considerably finer resolution (e.g. 5 arcminutes). These should be consistent with regional scale models. To obtain meaningful downscaled results at such a resolution, it is necessary to resort to a set of explanatory variables, which are easily available at the high resolution level and to relate these to observed land-use change. This paper puts forward a multinomial logit (MNL) model, in order to select a subset of such variables and quantify – in a Bayesian fashion – their impacts on land-use change – between the classes grassland, cropland, forest and other land – based upon multiple satellite-based land-use change maps.

2. A multinomial logit land-use change model

CONSIDER $J+1$ distinct land-use classes, with each land use-class denoted by j $(j=1,...,J+1)$. Our area of interest is subdivided into distinct parcels i $(i=1,...,N)$, which are referred to as so-called simulational units (SimU). Let us denote our main object of interest, the percentage of SimU i dedicated to land-use class j, as $y_{i,j}$.

Our model assumes that a specific SimU i has a share of land-use class j with the probability $y_{i,j}$. This is the MNL model and is specified as:

$$\Pr(y_i=j)=\frac{\exp(\mu_{i,j})}{1+\sum_{j=1}^{J}\exp(\mu_{i,j})} \tag{1}$$

where $\mu_{i,j}$ denotes the log-odds associated with land-use class j. We model $\mu_{i,j}$ through K explanatory variables $\boldsymbol{X}_{i,j}$. Thus, we can write in matrix notation:

$$\boldsymbol{\mu}_j=\boldsymbol{X}\boldsymbol{\beta}_j+\boldsymbol{W}(\phi)\boldsymbol{X}\boldsymbol{\theta}_j+\boldsymbol{\iota}_N\alpha_j. \tag{2}$$

where $\boldsymbol{\mu}_j$ is an N by 1 vector, $\boldsymbol{X}$ is an N by K matrix with associated K by 1 coefficient vectors $\boldsymbol{\beta}_j$. $\boldsymbol{\iota}_N$ represents an N by 1 vector of ones and α_j represents the intercept. The function $\boldsymbol{W}(\phi)$ is defined as:

$$\boldsymbol{W}(\phi)=\frac{\phi^m\boldsymbol{W}^{(m)}}{\sum_{m=1}^{M}\phi^m}. \tag{3}$$

where $\boldsymbol{W}^{(m)}$ is an N by N row-stochastic matrix with zeros on the main diagonal and m $(m=1,...,M)$ denotes the number of nearest neighbours (measured by e.g. geodesic distance) considered. $W_{i,s}^{(m)}$ $(s=1,...,N)$ greater than zero signifies that SimUs i and s are considered to be neighbours. The parameter $\phi\in[0,1]$ can be interpreted as a spatial decay parameter. The term $\boldsymbol{W}(\phi)\boldsymbol{X}\boldsymbol{\theta}_j$ in Eq. (1) is referred to as the spatial lag and $\boldsymbol{\theta}_j$ is the K by 1 vector of associated coefficients.

3. Estimation and prior set-up

THE Pólya-Gamma distribution Polson et al. (2013) can be used to sample directly from the MNL in Eq. (1). The main tactic employed is to introduce a Pólya-Gamma random variable in the joint distribution in such a fashion that the marginal resulting from the joint distribution leaves the original model intact.

For this purpose the likelihood of all the coefficients of land use class j, $\boldsymbol{\delta}_j=[\boldsymbol{\beta}_j',\boldsymbol{\theta}_j',\alpha_j]'$ can be written conditional on $\boldsymbol{\delta}_{-j}$, where $\boldsymbol{\delta}_{-j}$ denotes the $2K+1$ by J parameter matrix $\boldsymbol{\delta}$ without the j-th column (Holmes and Held, 2006):

$$\mathcal{L}(\boldsymbol{\delta}_j|\boldsymbol{\delta}_{-j},y)=\prod_{i=1}^{N}\left(\frac{e^{\eta_{i,j}}}{1+e^{\eta_{i,j}}}\right)^{y_{i,j}}\left(\frac{e^{\eta_{i,j}}}{1+e^{\eta_{i,j}}}\right)^{n_i-y_{i,j}} \tag{4}$$

where

$$\eta_{i,j}=\boldsymbol{Z}_i\boldsymbol{\delta}_j-C_{i,j}$$
$$C_{i,j}=\log\sum_{r\neq j}\exp(\mu_{i,r}).$$

Given the conditional likelihood in Eq. (**??**) and an additional set of priors, we can easily formulate a Gibbs sampler for our model. The rest of our prior set up and the particular elicitation is as follows:

$$\delta_{l,j}\sim(1-\gamma_{l,j})\mathcal{N}\left(0,\tau_0^2\right)+\gamma_{l,j}\mathcal{N}\left(0,\tau_1^2\right)\ \text{with}\ \tau_0=7/10^4,\ \tau_1=7/10$$
$$\gamma_{l,j}\sim Be(p_\gamma)\ \text{with}\ p_\gamma=1/2$$
$$\omega_{i,j}\sim PG(n_i,0)$$
$$\phi\sim U(0,1)$$
$$M\sim U(1,M_{max})\ \text{with}\ M_{max}=30.$$

As a prior for the regression parameters $\delta_{l,j}$ $(l=1,...,2K+1)$ we use the stochastic search variable selection (SSVS) specification introduced by George and McCulloch (1993), which is a mixture of normals prior. Following Polson et al. (2013) we set a Pólya-Gamma prior for the variance parameter $\omega_{i,j}$. τ_0 and τ_1 denote the prior variance and $\gamma_{i,j}$ is a mixture indicator, with corresponding Bernoulli prior. Our choices for the spatial parameter priors are motivated by LeSage and Pace (2009).

4. Estimation results

OUR global dataset consists of $212,707$ observations on changes between four land-use classes in the time period 2000-2010. Additionally we use three regional datasets for Europe, US and Ukraine, to improve the accuracy of our results. Table 1 provides an overview over our main datasets. Table 2 lists our explanatory variables.

Dataset source	Resolution	Coverage
ESA CCI data (Liu et al., 2012); processed by UC Louvain	5 arcminutes	Global
Satellite data from NLCD (Fry et al., 2007)	30m	USA
Wageningen (Fuchs et al., 2015); based on satellite and statistical data	30m	EU-15 and Switzerland
Satellite data, processed for Ukraine by (Skakun et al., 2015)	30m	Ukraine

Table 1: *Datasets of land-use change for the dependent variables; all data available for 2000 and 2010. ESA CCI - European Space Agency Climate Change Initiative, NLCD - National Land Cover Database.*

Short name	Variable	Short name	Variable	Short name	Variable
Alti2_300	avg. altitude (300m)	HarvWood	wood harvest (tons)	Slp3_6	avg. slope (6^0)
Alti3_600	avg. altitude (600m)	HI_MEAN	high fertilization (ha)	Slp4_10	avg. slope (10^0)
Alti4_1100	avg. altitude (1100m)	IR_MEAN	irrigated crops (ha)	Slp5_15	avg. slope (15^0)
CNTRY	country dummy	LI_MEAN	low fertilization (ha)	Slp6_30	avg. slope (30^0)
cropland	past cropland (ha)	meanTimeToMarket	dist. to next market (min)	Soil2_Medium	Soil type medium
CrpLnd	past cropland (pct)	MEANYLD	crop yield (tons)	Soil3_Heavy	Soil type heavy
forest	past forest (ha)	other	past other land (ha)	Soil4_Stony	Soil type stony
Forest	past forest (pct)	OthNatLnd	past other land (pct)	Soil5_Peats	Soil type peats
Grass	past grassland (pct)	popDens	population density	SS_MEAN	subsistance farming (ha)
grassland	past grassland (ha)	REGION	region dummy	STDYLD	std. crop yield (tons)
GRASYLD	grass yield (tons)	SimUarea	area of SimU		
HarvCost	wood harvest costs (USD)	Slp2_3	avg. slope (3^0)		

Table 2: *List of explanatory variables and their short names in our model. All variables except country and region dummy are SimU specific.*

The total set of SimUs is subdivided into 30 regions. We estimate a separate model for every region for our baseline data. For the regions where we have more than one dataset, we use the region-specific land-use change datasets in addition to the global dataset and treat both as independent observations. We treat both datasets as equally likely a priori. By incorporating information about multiple datasets in our parameters we reduce the chance of biased parameter estimates through observation errors.

Cropland			Grassland			Forest		
Coeff.	pmean	pip	Coeff.	pmean	pip	Coeff.	pmean	pip
cropland	1.09	1.00	grassland	1.08	1.00	forest	1.09	1.00
other	-1.07	1.00	other	-1.59	1.00	other	-1.43	1.00
meanTimeToMarket	-0.67	1.00	W_meanTimeToMarket	-1.14	1.00	W_HI_MEAN	0.61	0.78
grassland	0.67	1.00	W_grassland	0.07	0.25	cropland	-0.17	0.58
W_IR_MEAN	0.24	0.59	forest	-0.03	0.08	W_STDYLD	-0.27	0.56
ϕ	0.79							
M	9.48							

Table 3: *Estimation results for Brazil region. pmean - posterior mean of coefficient; pip - posterior inclusion probability of coefficient.*

Table 3 shows as an example a summary of coefficient estimates for the Brazil region. The first five rows contain the coefficients with the highest posterior inclusion probabilities, based on SSVS priors. A posterior inclusion probability of one signifies that the variable is included in all models, whereas a value close to zero signifies that the variable is virtually partialled out. The coefficients for other natural land are set to zero, and the coefficients in Table 3 are to be interpreted in relation to other natural land not changing. The last two rows contain the posterior estimate for the spatial decay parameter ϕ and M the maximum number of k-nearest neighbour matrices considered. The estimated value for ϕ indicates that spatial neighbourhood plays a significant role, with the 8-th neighbour having $\sim 1/5$-th of the influence of the first neighbour.

5. Downscaling land-use projections

WE use our coefficient estimates to downscale land-use change projections from GLOBIOM. The projections are available in ten year time steps from 2010 until 2100 and along three Shared Socio-Economic Pathways (SSP), which provide an exogenous framework for the agricultural model on socio-economic developments. SSP1 represents sustainable development, SSP2 is a middle of the road scenario, while SSP3 is characterized by continued divergence in economic growth.

For the projections we use the posterior mean of $\boldsymbol{y}$ (denoted as $\hat{\boldsymbol{y}}_0$ from Eq. (1). To arrive at $\boldsymbol{y}_1$, we replace the past observations on land-use with $\hat{\boldsymbol{y}}_0$, update the yield and population density variables and set α_j, so that the regional average composition of land-use change corresponds to GLOBIOM's regional projections. As an example of our projections Fig. 1 show the differences in cropland for SSP1-3 in the period 2010-2100.

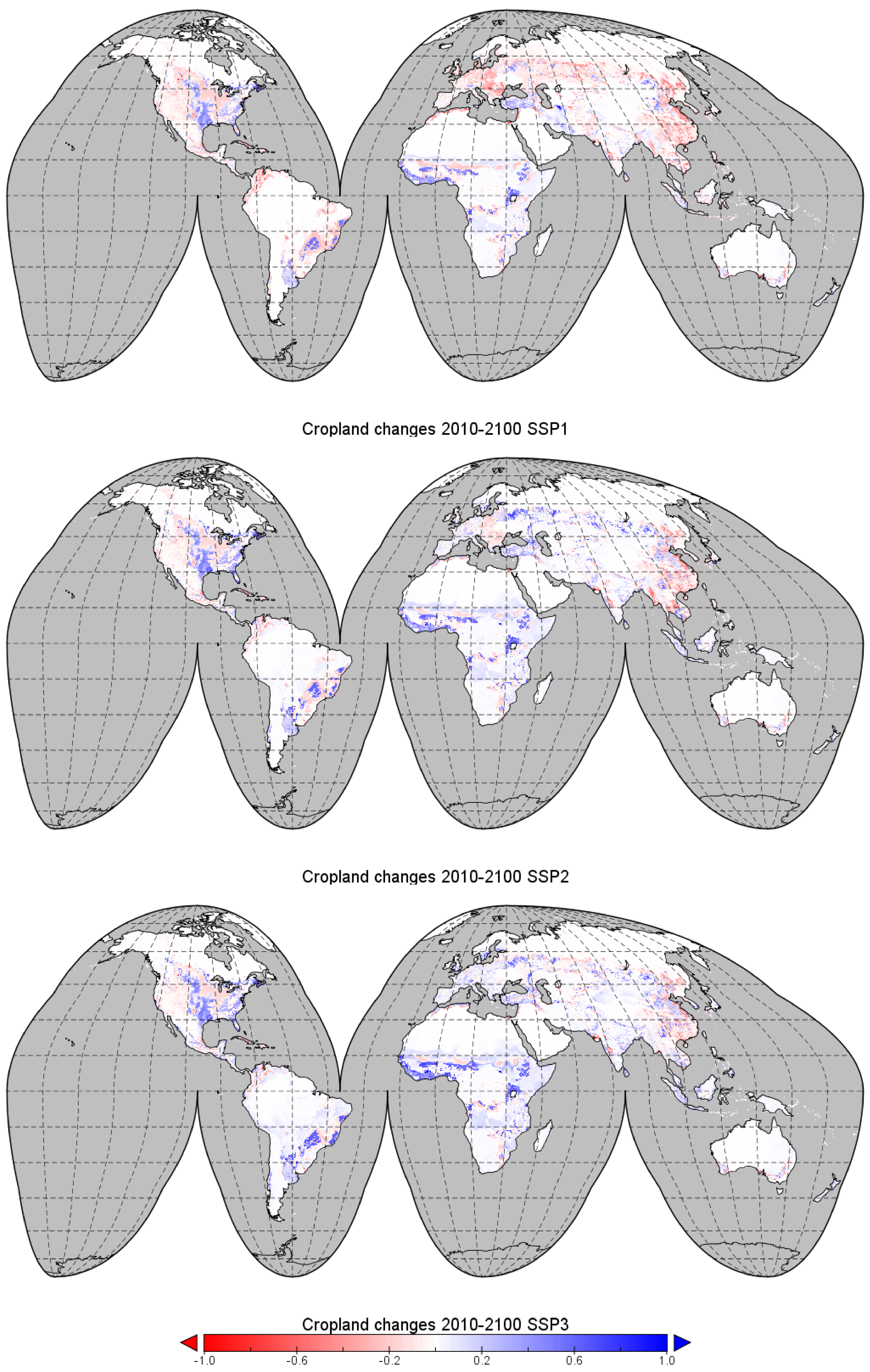

Figure 1: *Posterior mean of SimU-level downscaled SSP1, SSP2 and SSP3 GLOBIOM land-use change projections showing the percentage of cropland change per SimU for the interval 2010-2100.*

6. Concluding remarks

FIRST, our results offer valuable insight into the dynamics of land-use change: chiefly, while past values of land-use (and surrounding land-use) are undoubtedly important, other factors such as proximity to market seem to also play a central role in most regions. Second, we show the influence of spatial proximity per region on land-use change. Third, we demonstrate the applicability of our method by downscaling GLOBIOM land-use projections along multiple SSP scenarios.

References

Fry Ja, Xian G, Jin S, Dewitz Ja, Homer CG, Yang L, Barnes Ca, Herold ND and Wickham JD (2007) Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 73(4), 337–341

Fuchs R, Herold M, Verburg PH, Clevers JGPW and Eberle J (2015) Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Global change biology* 21(1), 299–313

George E and McCulloch R (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88(423), 881–889

Holmes CC and Held L (2006) Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* 1(1), 145–168

LeSage PJ and Pace RK (2009) *Introduction to Spatial Econometrics*. CRC Press, Boca Raton London New York

Liu Y, Dorigo W, Parinussa R, de Jeu R, Wagner W, McCabe M, Evans J and van Dijk A (2012) Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment* 123(1), 280–297

Polson NG, Scott JG and Windle J (2013) Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association* 108(504), 1339–1349

Skakun S, Kussul N, Shelestov A, Lavreniuk M and Kussul O (2015) Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.1109/JS