International Food and Agribusiness Management Review
Volume 22 Issue 2, 2019; DOI: 10.22434/IFAMR2018.0072

Received: 29 June 2018 / Accepted: 20 November 2018

RESEARCH ARTICLE

The determinants of recent soybean expansion in Mato Grosso, Brazil

Otávio Lemos de Melo Celidonio①[a], Liane S. Werner[b] and Juliana Dias Bernardes Gil[c]

[a]*Chief Executive Office, National Service of Rural Extension of Mato Grosso State (Senar, MT),*
R. Eng. Edgard Prado Arze, s/n, Q1, Setor A, Centro Político Administrativo, Cuiabá, MT, Brazil

[b]*Associate Professor, Universidade Federal do Rio Grande do Sul, Av Osvaldo Aranha,*
DEPROT/PPGEP, 99 – 5º Andar – Bairro Bom Fim, Porto Alegre, RS, Brazil

[c]*Post Doctoral Research Associate, Plant Production Systems Group at Wageningen*
University, P.O. Box 430, 6700 AK Wageningen, the Netherlands

Abstract

Understanding what drives, catalyzes or constraints land use change in the Brazilian agricultural frontier is a condition for effective policy design at the local level, which in turn might have implications for food production, environmental conservation and greenhouse gas emissions worldwide. We analyzed the process of agricultural expansion observed in the state of Mato Grosso, the country's largest agricultural producer, by mapping and quantifying the incorporation of new farming areas and the conversion of existing ones into mechanized soybean fields at the farm-level. Through statistical modelling we also investigated the influence of key economic, biophysical, environmental and logistics variables on this process while accounting for recent changes in the Brazilian environmental legislation. We found that the area converted to soybean production increased almost 1.5 million hectares between 2009 and 2013, more than 70% of which in farms that already had some soybean in previous years. By comparing the explanatory power of eight regression models involving different groups of variables, we found that soybean expansion is strongly associated with the presence of other soybean fields and warehouses within 50-100 km. The model with the largest explanatory power suggests that soybean expansion is also likely to occur in areas of high conservation value. Finally, the sensitivity of soybean expansion to soybean prices indicated the potential for further agricultural growth in Mato Grosso while highlighting how crucial smart logistics investments are for regional development with environmental protection.

Keywords: land use change, agricultural expansion, soybean, Brazil
JEL code: Q15, Q18, Q28, Q58, R14, R48, R50, R51, R52, R58, O13, O18, O21, O54

---

①Corresponding author: otaviolmc@gmail.com

1. Introduction

The state of Mato Grosso is at the center of agricultural production and land use change in Brazil. Yet, the local process of territorial occupation and landscape transformation happened recently, over a relatively short period of time. Despite Mato Grosso's low population density (Ibge, 2015) and poor infrastructure (Macrologística, 2013) favorable agronomic conditions ensured its position as top producer of several commodities. Currently, Mato Grosso alone accounts for a significant share of the national production of cotton (50%), corn (35%), soybean (29%) and beef cattle (15%) (Conab, 2015; Ibge, 2015).

In terms of economic value, soybean is the most important product of Mato Grosso's agribusiness sector, representing 40.36% of its gross domestic production (GDP) (Do Nascimento *et al.*, 2018). The fact that the agribusiness sector accounts for 53.15% of the state's total GDP makes it clear that this crop largely drives the local economy.

Further, soybean expansion is expected in face of rising commodity prices pushed by world population growth and currency exchange rates favorable to Brazilian exports (Alexandratos and Bruinsma, 2012). According to the latest projections of Mato Grosso's Institute of Agricultural Economics, more than half of the area currently occupied by pastures (i.e. 15 million out of 24 million ha) could be allocated to the production of grains, leading to a 160% increase of the 8.9 million ha cultivated with soybean in 2014/15 (IMEA, 2015b). This conforms with other recent projections according to which soybean annual growth rates will be between 0.80 and 2.65% for Brazil as a whole, and between 2.97% and 4.81% for Mato Grosso specifically (Fiesp, 2015; Mapa, 2016; Masuda and Goldsmith, 2009).

While agricultural expansion is an important driver of economic growth for Brazil (soybean exports in 2015 accounted for more than 1.5% of the country's GDP) and might be needed to satisfy the world's demand for food and feed, environmental impacts potentially associated with it must be considered by policy makers. Efforts undertaken by the Brazilian Federal Government to drive agricultural expansion into already deforested areas have been largely successful and contributed to decouple this process from illegal forest clearing (OECD, 2015). However, further policy efforts are needed to ensure the persistence of such trend and, furthermore, to bring Mato Grosso's ca. 11 million ha of degraded pastures back into cultivation.

Understanding what drives, catalyzes or constraints land use change is necessary for the design of effective and sustainable policies. Previous assessments of agricultural expansion and territorial occupation have already provided important insights into land use change in Mato Grosso (Fiesp, 2015; IMEA, 2015a; Mapa, 2016) and helped reveal more stable, longer-term land use trends. Among catalyzing factors were climatic conditions, proximity to markets, land tenure profile, logistics, biophysical conditions and land use in previous years (Aguiar *et al.,* 2007; Jasinski *et al.,* 2005; Mann *et al.,* 2010). Other studies have also explored how the interaction between some of these factors affected the income and profitability levels associated with different land use choices in livestock systems (Bowman *et al.,* 2012) and crop farming (Mann *et al.,* 2014).

Some literature also exists on short-term land use decisions in recent years. While some authors focused on the identification of multiple drivers of land use change (Veldkamp and Lambin, 2001; Verburg *et al.,* 2014a), others emphasized the role of economic factors affecting the relative profitability of different land uses, such as transportation costs (Weinhold and Reis, 2008), commodity prices (Verburg *et al.,* 2014b) or land rents (Mann *et al.,* 2010, 2014). The importance of non-economic sources of utility in Mato Grosso has also been examined, particularly with regards to the persistence of pastures where the conversion to cropland could bring greater economic returns (Bowman *et al.,* 2012).

Especially due to the deforestation process occurred in Mato Grosso, the social and environmental impacts of the land use change have also received attention (Anderson *et al.,* 2003; Carvalho *et al.,* 2016; Rivero *et al.,* 2009; Simões *et al.,* 2014). These impacts cannot be ignored given the fast pace of decision making in agribusiness and the dynamic nature of land use change, especially in a state as large and globally connected as Mato Grosso. The export orientation of the state's agricultural sector, market volatilities related to climate and price shocks, as well as recent changes to the Brazilian environmental legislation exacerbate such dynamism. After 2008, important modifications to the country's Forest Code started being implemented and new environmental conservation tools such as the Rural Environmental Registry (CAR) and the Single Environmental License (LAU) posed barriers to further forest clearance – ultimately affecting land prices and land use decisions (Almeida *et al.,* 2014).

In this study, we characterized soybean expansion and analyzed the factors influencing this process through the lens of farmers themselves, by looking into the variables that tend to steer year-to-year farming decisions. We first mapped the cumulative expansion of mechanized agriculture in Mato Grosso, Brazil over 2009-2013 at the farm level. We then identified which variables influenced cropland expansion through mixed statistical approaches. Finally, we analyzed the influence of each variable under varying soybean profitability levels through sensitivity analyses on different soybean prices. The implications of our findings, limitations of the study and policy recommendations are discussed at the end.

2. Methods

Each of the steps performed as well as details on the variables considered in our analyses are explained as follows.

2.1 Mapping

A land use map of Mato Grosso was generated in order to map and quantify the cumulative expansion of soybean mechanized agriculture observed over 2008-2013. Individual farms were chosen as the unit of analysis because they allow us to capture the large agro-climatic and socioeconomic variation of regions within Mato Grosso, besides reflecting the scale at which farmers' decisions are made and the environmental legislation must be enforced.

As many databases of rural properties exist in Brazil – not all in accordance with a single official agrarian registry – we juxtaposed six of them onto a map of municipality boundaries to generate a state-wide grid of farm boundaries. Non-registered areas were divided into polygons of 10,000 ha each and randomly assigned to the other datasets. The area of 10,000 ha was chosen empirically since it preserved the spatial correlation power of the analysis without leading to excessive territorial fragmentation nor requiring overly high computational processing power. ArcGIS was used to correct for the topology errors that arose from the juxtaposition. Table 1 shows basic information of the eight layers included in the final map.

Table 1. Datasets used to generate a grid of farm boundaries. The numbers under the column 'Layer' indicate the order of juxtaposition (1 is at the bottom while 8 is above all the others).

Layer	Database / Contents	Source	Year
1	Areas not registered in any of the databases	Not applicable	2014
2	Municipality boundaries	IBGE	2012
3	Rural Environmental Registry (CAR)	Sema-MT	2014
4	Single Environmental License (LAU)	Sema-MT	2014
5	Georeferenced properties (GEO)	Incra	2014
6	Settlements (ASR)	Incra	2014
7	Conservation Units (UCF)	Ibama	2014
8	Indigenous lands (UTI)	Funai	2014

Conservation units and indigenous lands (layers 7 and 8) were excluded from our analysis since the Brazilian Environmental Legislation prohibits the establishment of commercial farming in such areas. The law also determines that a minimum share of every rural property must be set aside for conservation purposes (the so-called 'Legal Reserve'). The share varies according to the biome where the property is located (until 2004, 35% in the Cerrado and 80% in the Amazon; after 2004, 20% in the Cerrado and 50% in the Amazon). In our analysis, it was not possible to check whether each property was in full compliance with the law due to uncertainties concerning (1) the exact share kept as set-aside land by each farmer, (2) the exact year when deforestation may have happened; and (3) the exact location of each farm's Legal Reserve (given the possibility to have it in another property through a lawful compensation instrument).

Although settlement properties (layer 6) are often too small for the cultivation of soybean from an economic feasibility perspective, they were considered as a potential area for soybean expansion given that they can be rented and/or incorporated to adjacent properties. Whenever inconsistencies across layers were encountered, the sequence of juxtaposition determined the predominance of a layer over another. For example, if the same grid cell appeared as a forest conservation unit in layer 7 and as indigenous land in layer 8, it would be considered an indigenous land in our analysis. The only inconsistencies found were associated with the layer CAR, for which no absolute georeferencing precision is required as in the case of land owners applying for a LAU (layer 4). The degree of precision of layers 3-5, combined with the frequency of farms with georeferenced properties (GEO), LAU and/or CAR, motivated their juxtaposition order. In particular, GEO was placed 'above' LAU given that having a farm georeferenced is the first step to obtain the LAU.

The final grid was a map of farms with potential for soybean expansion over 2009-13, determined by the presence of agricultural areas already deforested and still unoccupied by soybean (Figure 1). By cross-checking it with land use data from IMEA (2015a), it was possible to identify which polygons had undergone any land

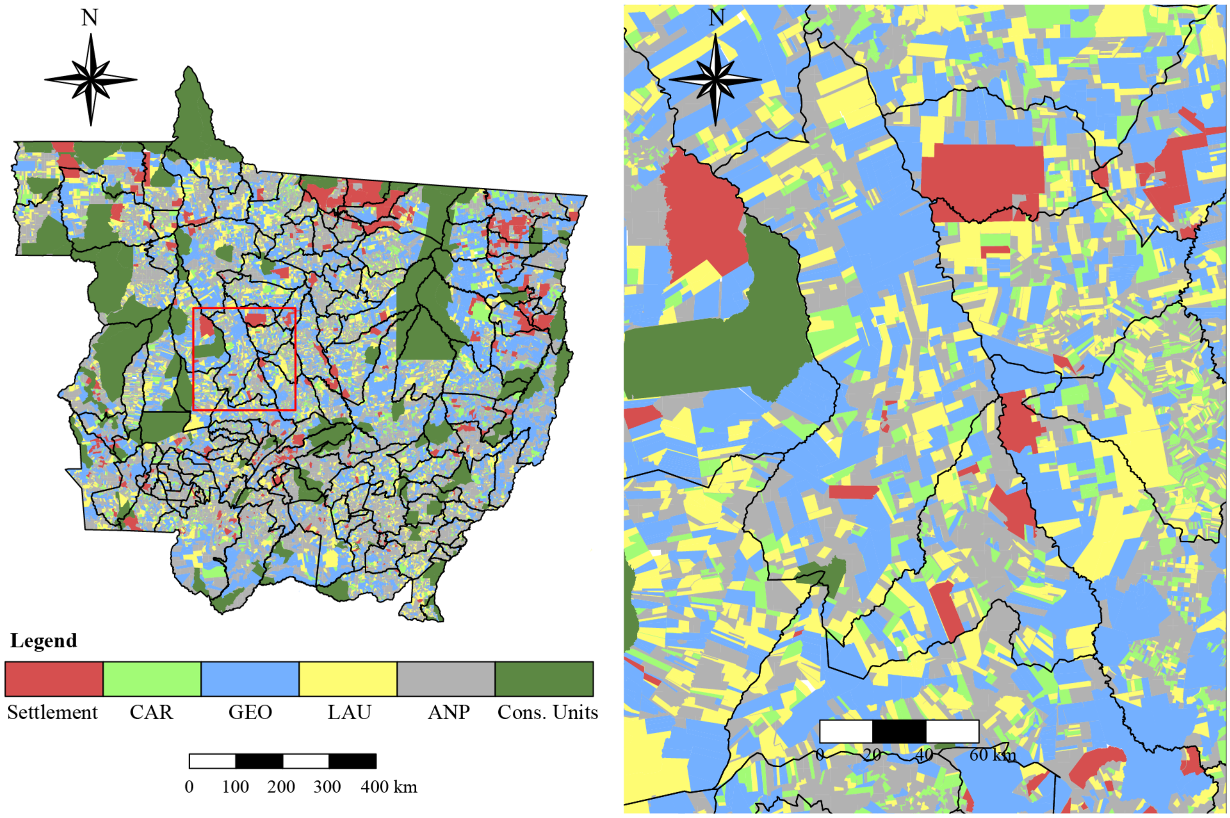

Figure 1. (Left) Rural properties in Mato Grosso state, as they appeared in the final grid used in our analysis. (Right) Inset showing map detail at the property level.

use changes, including soybean expansion. For this, we crossed MODIS satellite imagery with data on 25 thousand ground-collected coordinate points using the Support Vector Machine (SVM) classifier. The aim was to identify how land was being used in each year, from 2008 till 2014. The results were then checked against images from the Landsat satellite, which has a better resolution, to reduce errors. The hit rate of the soybean classification measured by a matrix of confusion was 96.91%.

2.2 Statistical analysis

■ *2.2.1 Principal component analysis*

Based on the existing literature of land use in Mato Grosso (Aguiar *et al.,* 2007; Jasinski *et al.,* 2005; Mann *et al.,* 2010), we hypothesized 38 variables to be related to farmers' land use decisions (see complete list in Supplementary Table A1). The agronomic variables selected for the analysis indicated the potential for mechanization, pH correction requirements and the presence of favorable temperature and precipitation for crop development. The logistic conditions offered a proxy for supply chain infrastructure and transport cost. We estimated three economic parameters related to the profitability of soybean production in each location (see supplementary materials – economic variables and modeling parameters). It is important to note that the profitability of each farm was calculated relative to the already deforested area; areas with no forest but with agronomic potential to produce soybean were also considered, and had their profitability estimated as well.

We then used Principal Component Analysis (PCA) to reduce the number of explanatory variables and identify the most crucial factors behind soybean expansion. In the PCA, we considered only the grid cells where soybean expansion took place. To avoid distortions, the data were standardized and each grid cell was weighted according to the size of the expansion.

■ *2.2.2 Logistic regression and model comparison*

In parallel to the PCA, eight logistic regression models were run with different subsets of variables. The comparison of the explanatory power of the different models gave us insights into the importance of each variable for soybean expansion over the analyzed period. For three of the models, variables were selected through statistical processes; another four included thematic clusters of variables (economic, agronomic and logistics); and the remaining model included variables selected by local experts (Table 2).

Table 2. Description of variables in each statistical model.

Model	Variables[1]	Origin / nature
1	roo, coo, rbo, dcv, chm, msc, svd, csl, depmi, dpmi, djc, dpf, rsv200	PCA component 1
2	ra50, rsv050, rsq050, rsq100	PCA component 2
3	roo, coo, rbo, dcv, chm, msc, svd, csl, depmi, dpmi, djc, dpf, ra50, rsv050, rsv200, rsq050, rsq100	Combination of Models 1 and 2
4	roo, coo, rbo	Economic variables
5	dcv, chm, msc, svd, csl	Agronomic variables
6	denmi, depmi, dami, dsmmi, dpmi, djc, dpf, rsv025, rsv050, rsv100, rsv200, rsq025, rsq050, rsq100, rsq200	Logistics variables
7	dmia, ra025, ra050, rsq025, rsq050	Logistics variables of neighboring farms
8	roo, dcv, chm, msc, svd, csl, lvr, pst, dmia, denmi, depmi, dami, dsmmi, dpmi, djc, ra250, ra050, ra100, rsv025, rsv050, rsv100, rsq025, rsq050	Variables selected by local experts

[1] See Supplementary Table A1 for abbreviations.

In the regressions, the binary dependent variable could assume the values of 1 if the farm had experienced soybean expansion or 0 otherwise, as stated below.

$$Y_i = \frac{1}{1 + exp\{\sum_{j=0}^{n} \beta_j X_{ij}\}} + e_i$$

where Y_i is the probability of soybean expansion in farm i, β_j is the coefficient of j, X_{ij} represents the value of variable j and e_i is the error term.

All eight regression models reflect the same period (2009-2013). Data on land use conversion from 2014/2015 was used to check the model fit. Since the probabilities of soybean expansion in all eight models were not normally distributed (see Supplementary Figures A1 and A2), we employed non-parametric tests. The following tests were performed with the software R (https://www.r-project.org): Komogolov-Smirnov, Mann-Whitney-Wilcoxon and nonparametric multiple comparisons. We then plotted a 'Roc curve' for each model to check which of them best predicted the conversion to soybean occurred in 2014/2015. The area below each curve represents the explanatory power of the corresponding model, indicating if the land use change occurred where the model predicted a higher probability of change.

■ *2.2.3 Sensitivity analysis*

Sensitivity analyses were performed at the end to explore what would happen to the results of each regression model (in terms of the relative influence of different variables) if soybean prices changed. Heat maps displaying the deciles of probabilities of soybean expansion were also generated to elicit spatial patterns and visual similarities across the models.

3. Results

3.1 Mapping

Mato Grosso's potential farming area – excluding indigenous lands, conservation units and settlements – encompassed 38,393 rural properties. Together, these properties corresponded to 66 million ha (or 73.2% of the state's total area) (Table 3). Registered farms (present in the GEO, LAU and/or CAR datasets) represented 44.8% of the state's total area and 61.1% of the analyzed area.

Soybean expansion was observed throughout the entire analyzed period (Figure 2). Most of the expansion occurred in 2012 (368 thousand ha) in the Northeast and Middle-North regions. In 2013, 52.7% of the area of all rural properties still had primary vegetation from at least one of the three biomes spanned by Mato Grosso (i.e. the Amazon rainforest, the Cerrado and the Pantanal).

Table 3. Number of properties and corresponding area mapped through each dataset.

Dataset[1]	**# Grid cells**	**%**	**Area (ha)**	**%**
GEO	9,046	23.6	22,609,611	34.3
LAU[2]	7,718	20.1	11,240,423	17.0
CAR[3]	9,341	24.3	6,499,357	9.8
ANP[4]	12,293	32.0	25,659,323	38.9
Total	38,398	100	66,008,714	100

[1] GEO = georeferenced properties; LAU = single environmental license; CAR = rural environmental registry; ANP = non-registered areas.

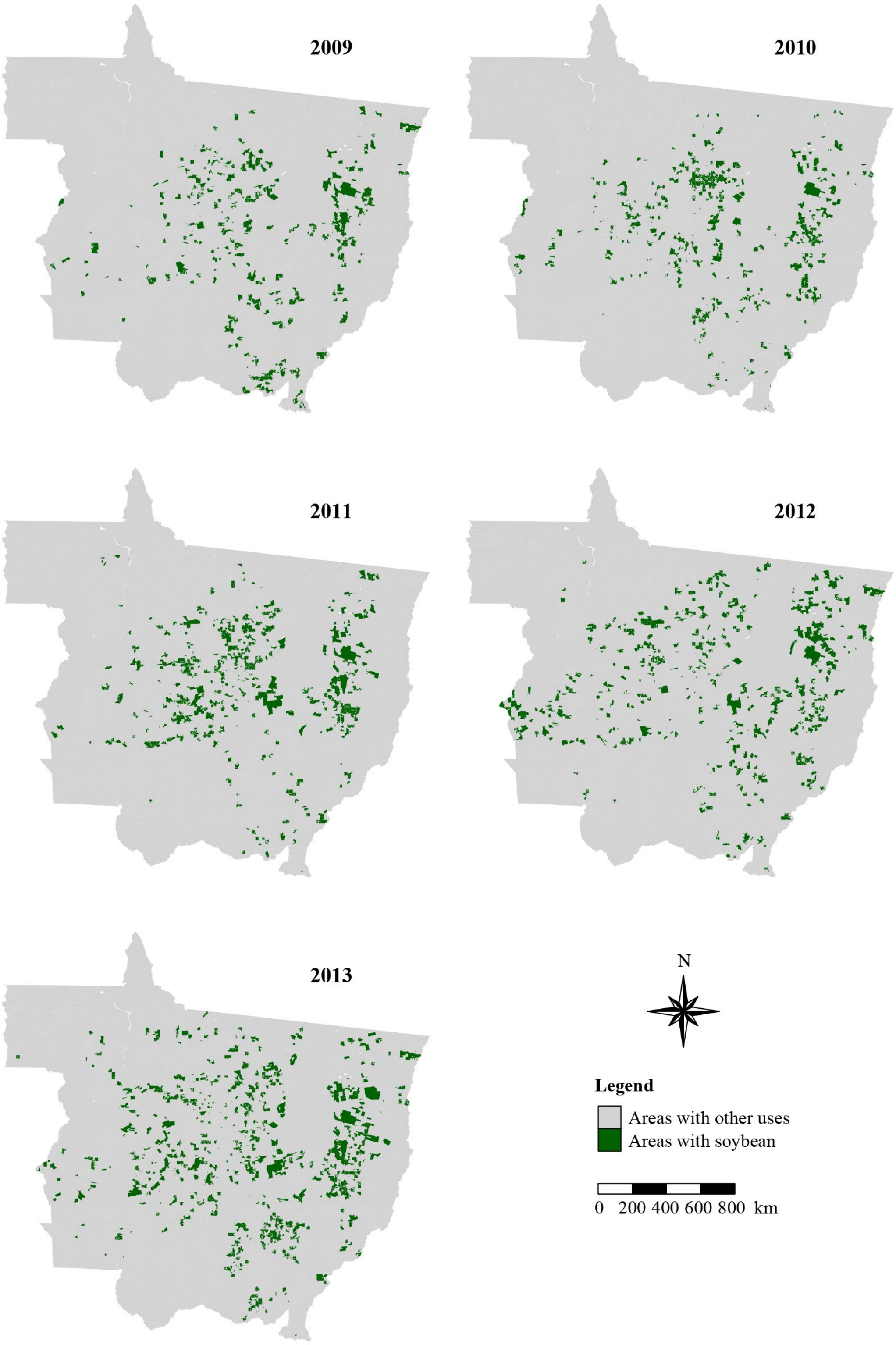

Figure 2. Soybean expansion in Mato Grosso between 2009 and 2013.

Although non-registered areas (ANP) represented 40% of the total area over 2009-2013, only 25% of the soybean expansion happened in these places. In other words, 75% of the soybean expansion occurred in areas with CAR, GEO and/or LAU. Although 2012 was the year with the largest expansion of soybean production, it was also the year with the lowest proportional growth over forest areas (3.4%). Most of this growth (87.7%) occurred over pasture areas. The largest proportion of soybean expansion over remaining forest areas was observed in 2013 (7.8%), when expansion over pasture and other land cover classes accounted for 77.6% and 14.6%, respectively. Both the area and the number of farms where soybean expansion took place increased over 2009-2013, although not at the same rate (Table 4). On average, the area converted per farm oscillated between 186 ha in 2013 and 277 ha in 2012. The share of soybean expansion in farms with no previous soybean cultivation reached its peak in 2012.

3.2 Principal component analysis

The results of the PCA are shown in Table 5. The visual correlation between the different datasets of historic soybean cultivation and the heat maps of the probability of soybean expansion (Figure 5) reflects the natural suitability of certain areas for soybean expansion. Principal component 1 explained 40% of the variance in our sample. In total, 12 variables showed values greater than 0.8 (Table 5). Although the coefficient obtained for the variable ‘storage capacity’ was slightly lower, the variable ‘storage capacity in 200 km’ showed the largest negative value and contributed to the model’s overall explanatory power. Principal component 2 explained only 14% of the variance in our sample; still, it featured five variables with values above 0.75 – namely storage capacity in 50 km, storage capacity in 100 km, number of storage capacity in 50 km, number of storage capacity in 100 km, and the total farming area in 50 km. Economic variables (*roo*, *coo* and *rbo*) had the largest influence on soybean expansion, followed by average soil class (*csl*) and storage capacity within a 200 km-radius (*rsv200*) (Figure 3).

The scale of production interfered directly in the estimated production costs, particularly because the size of the farm does not always allow for the optimal employment of machinery and labor, thus leading to idleness and higher costs with depreciation and insurance. Likewise, logistics affected production costs and revenues. Costs accounted for the effect of farm location on the prices of fertilizer and lime, since these inputs are utilized in large amounts and have relatively low added-value (the interference of transport cost is considerable). The farm location is also important in determining the revenue accrued by farmers since farm-gate prices are calculated as the soybean price referenced to the international market (e.g. Chicago Stock Exchange, USA) minus freight to the nearest port.

Table 4. Indicators of soybean expansion over 2009-2013.

Year	# Farms where soybean expanded	Total converted area (ha)	Average converted area by farm (ha)	Share of soybean expansion in farms with no previous soybean (%)
2009	1,045	250,660	240	22
2010	1,080	271,121	251	20
2011	1,270	270,701	213	20
2012	1,331	368,732	277	43
2013	1,817	338,420	186	27

Table 5. Results of the Principal Component Analysis.[1]

Nome	**Code**[2]	**Dim.1**	**Dim.2**	**Dim.3**	**Dim.4**	**Dim.5**	**Dim.6**	**Dim.7**	**Dim.8**
		40.2%	**14.4%**	**9.3%**	**5.4%**	**4.3%**	**3.6%**	**2.9%**	**2.6%**
Total area of the property	Área	0.630	0.289	0.692	-0.025	-0.071	0.101	-0.067	0.008
Net profitability	roo	0.893	-0.009	-0.181	0.014	-0.104	0.026	-0.053	-0.006
Total costs	coo	0.948	0.065	-0.147	0.056	0.051	0.034	-0.078	0.001
Total revenue	rbo	0.955	0.025	-0.173	0.034	-0.036	0.031	-0.067	-0.003
Av. slope class (%)	dcv	0.913	0.066	-0.148	0.138	0.000	0.036	-0.067	-0.003
Av. volume of annual rainfall	chm	0.968	0.032	-0.134	0.041	0.074	0.023	-0.050	0.004
Number of dry months	msc	0.958	0.048	-0.098	0.077	-0.005	0.024	-0.101	-0.007
Av. rating of severity of drought	svd	0.932	0.090	-0.138	0.042	-0.079	0.006	-0.129	-0.004
Soil class	csl	0.888	0.065	-0.074	0.199	0.000	0.064	0.029	-0.001
Sugarcane area	cna	-0.033	-0.010	0.005	0.018	-0.057	0.026	0.099	0.908
Eucalyptus area	ect	0.019	0.011	0.127	-0.147	-0.025	0.045	0.391	-0.095
Soybean area	lvr	0.352	0.010	0.506	-0.122	0.504	0.267	-0.178	0.021
Irrigated area	piv	0.193	0.418	0.500	-0.526	0.028	0.016	0.292	-0.005
Pasture area	pst	0.600	0.333	0.637	0.046	-0.263	0.012	-0.006	0.003
Forest area	rem	0.628	0.327	0.637	-0.020	-0.220	0.047	-0.033	0.000
Rubber tree area	srg	0.509	0.134	0.457	0.434	-0.363	0.023	-0.252	0.004
Teak tree area	tka	0.018	-0.044	-0.026	-0.033	-0.085	0.086	0.082	-0.400
Other uses area	otr	0.339	0.429	0.605	-0.382	-0.052	0.004	0.238	-0.004
Distance to nearest soybean farm	dmia	0.483	0.085	0.133	0.386	0.071	-0.456	0.388	-0.003
Min. distance to non-paved road	denmi	0.466	-0.117	-0.022	-0.242	0.661	0.176	-0.041	0.019
Min. distance to paved road	depmi	0.809	-0.053	-0.001	0.163	0.135	-0.293	0.242	-0.007
Min. distance to a storage facility	dami	0.631	-0.138	-0.072	0.302	0.299	-0.341	0.391	-0.012
Min. distance to a city	dsmmi	0.747	-0.087	0.008	-0.173	0.474	0.024	-0.099	0.019
Distance to nearest port	dpmi	0.944	0.167	-0.133	-0.022	0.056	0.009	-0.040	0.005
Min. distance to the nearest limestone deposit	djc	0.914	-0.010	0.001	0.199	0.004	-0.060	-0.058	-0.004
Distance to the port of Santos	dpf	0.965	-0.002	-0.102	0.077	0.051	0.031	-0.121	0.006
Sum of agricultural area up to 25 km	ra025	0.663	-0.283	-0.304	-0.234	-0.248	0.252	0.185	0.039
Sum of agricultural area up to 50 km	ra050	0.008	0.786	-0.170	-0.350	0.097	-0.203	-0.093	0.030
Sum of agricultural area up to 100 km	ra100	0.138	0.597	-0.275	-0.470	0.009	-0.344	-0.170	0.031
Sum of agricultural area up to 200 km	ra200	0.571	-0.205	-0.452	-0.367	-0.322	0.125	0.087	0.019
Sum of storage capacity up to 25 km	rsv025	-0.121	0.682	-0.217	0.193	0.054	0.404	0.248	-0.014
Sum of storage capacity up to 50 km	rsv050	-0.151	0.849	-0.203	0.073	0.097	0.066	0.001	-0.041
Sum of storage capacity up to 100 km	rsv100	-0.265	0.804	-0.159	-0.031	0.001	-0.348	-0.180	0.001
Sum of storage capacity up to 200 km	rsv200	-0.714	-0.036	0.444	0.260	0.221	-0.120	-0.120	-0.013
Number of storage facilities up to 25 km	rsq025	-0.151	0.678	-0.259	0.283	0.043	0.449	0.239	0.011
Number of storage facilities up to 50 km	rsq050	-0.201	0.814	-0.260	0.233	0.052	0.181	0.038	-0.014
Number of storage facilities up to 100 km	rsq100	-0.313	0.804	-0.198	0.140	-0.044	-0.166	-0.096	0.012
Number of storage facilities up to 200 km	rsq200	-0.583	0.082	0.237	0.380	0.214	0.177	-0.059	0.046

[1] Marked cells indicate the variables selected to compose the models.

[2] See Supplementary Table A1 for abbreviations of the codes.

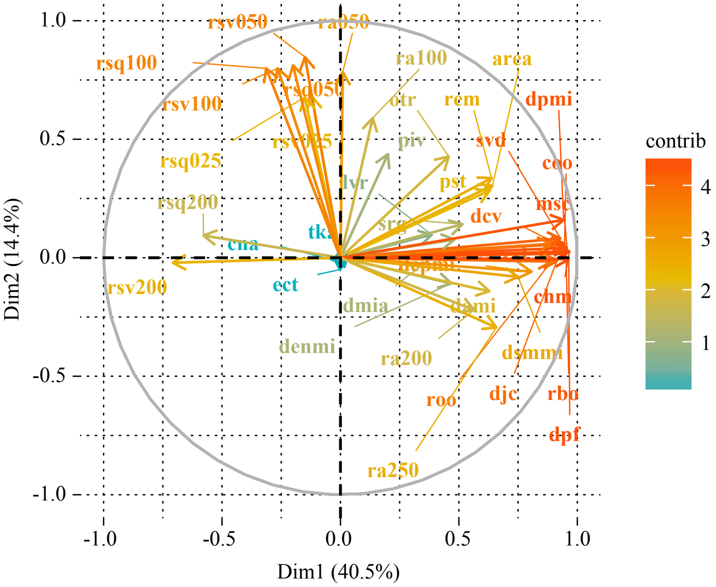

Figure 3. Principal Component Analysis results: map of variable factors for principal components 1 and 2.

3.3 Logistic regressions and model comparison

The impact of the variation of each variable on the probability of soybean expansion is displayed in Table 6 and 7. Such impact was calculated as the absolute difference in soybean expansion probability, measured in percentage points and caused by variations in each variable. Two cases were analyzed: when the variables assume their minimum and maximum (Table 6) values vs. when they assume median values that correspond to the 1st or 3rd quartiles of their distribution function (Table 7). The quartile measure facilitates the analysis of variables whose distribution function is concentrated around specific classes. For example, if a variable such as 'slope' was set to its minimum and maximum values, model result would be strongly affected; however, that would hardly be the case in the real world given Mato Grosso's natural conditions. In general, *biophysical* variables did not have a major effect on the probability of soybean expansion. Variations in soil class (csl) were the main source of sensitivity for Model 5, second main source for Model 1 and third main source for Models 3 and 8. The severity of droughts (svd) exhibited high relative importance to Models 1 and 5. Variables controlling for *previous land use* (pst, lvr) were part of Model 8 only and led to small conversion probability changes under current conditions (i.e. 1st and 3rd quartiles) but large probability changes (up to 97%) when minimum and maximum values were considered. Proxies of the *surrounding farming area* (dmia, ra025 and ra050) corresponded to the greatest sensitivity in Models 2, 3, 7 and 8. Only two of the three *economic* variables originally included in Model 4 were found to be significant. The difference between soybean expansion probabilities where it did occur vs. not was considerable, just like the median between the first and third quartiles (73%). The probability of conversion was shown to be particularly sensitive to *logistics* variables, particularly siloes and storage facilities, indicating the importance of this type of structure for future agricultural expansion and development.

The model with the highest explanatory power (Model 2) contained two variables representative of logistics and infrastructure: siloes and neighboring soybean fields. Intermediate distances (50km and 100km) were found to be more important than the shortest and longest distances (25km and 200km). Although access to farm is commonly pointed out by specialists as one of the main determinants of farmers' land use decisions, the distance to paved and unpaved roads (denmi and depmi, respectively) did not have a large impact either. The same was true concerning the distance to the centroid of the municipality (demmi), ports (dpf / dpmi) and lime mills (djc).

Table 6. Impact of the variation of each variable on the probability of soybean expansion under each regression model, measured as the absolute difference in percentage points between the maximum and minimum values.[1,2]

Code	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
roo	7.24***		6.77***	7.91***				4.86***
coo	Exc		Exc	1.01***				
rbo	Exc		Exc	Exc				
dcv	1.76***		1.61**		2.59***			1.27***
chm	0.47*		0.59***		1.79***			2.05***
msc	4.34***		5.07***		6.55***			2.56***
svd	2.79***		1.25***		2.02***			1.30***
csl	2.75***		2.68***		3.58***			2.29***
cna								
ect								
lvr								97.26***
piv								
pst								97.38***
rem								
srg								
tka								
otr								
dmia							6.72***	4.60***
denmi						1.76***		1.49***
depmi	2.14***		2.36***			Exc		6.28***
dami						6.29***		3.56***
dsmmi						9.80***		Exc
dpmi	2.08***		1.90***			3.37***		2.26***
djc	Exc					Exc		2.75***
dpf	0.76**		0.81**			Exc		
ra025							2.48***	2.28***
ra050		41.13***	13.72***				6.24***	1.21*
ra100								2.54***
ra200								
rsv025						Exc		0.76*
rsv050		2.33***	1.13***			Exc		19.52***
rsv100		7.37***	13.35***			44.68***		2.47***
rsv200	1.30***		2.82***			4.44**		
rsq025						Exc	1.04	Exc
rsq050		1.90	Exc			Exc	3.51***	2.72***
rsq100		3.14***	3.66***			4.73***		
rsq200						1.55		

Low 🟩🟨🟧🟥 high impact

[1] 'Exc' indicates non-significant variables excluded from the model. $P<0.001$ (***), $P<0.01$ (**), $P<0.05$ (*).

[2] See Supplementary Table A1 for abbreviations of the codes.

Table 7. Impact of the variation of each variable on the probability of soybean expansion under each regression model, measured as the absolute difference in percentage points between the values of the 1st and 3rd quartiles.[1,2]

Code	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
roo	2.28***		2.21***	2.54***				1.76***
coo	Exc		Exc	0.19***				
rbo	Exc		Exc	Exc				
dcv	0.29***		0.25**		0.48***			0.18***
chm	0.16*		0.19***		0.56***			0.64***
msc	0.00***		0.00***		0.00***			0.00***
svd	0.89***		0.41***		0.66***			0.43***
csl	1.87***		1.78***		2.53***			1.42***
cna								
ect								
lvr								0.01***
piv								
pst								0.17***
rem								
srg								
tka								
otr								
dmia							4.75***	3.14***
denmi						0.28***		0.23***
depmi	0.43***		0.50***			Exc		0.35***
dami						3.13***		1.20***
dsmmi						0.68***		Exc
dpmi	0.53***		0.48***			0.86***		0.57***
djc	Exc					Exc		0.81***
dpf	0.21**		0.23**			Exc		
ra025							1.05***	0.97***
ra050		3.48***	1.74***				1.22***	0.31*
ra100								0.69***
ra200								
rsv025						Exc		0.05*
rsv050		0.63***	0.19***			Exc		0.89***
rsv100		1.18***	1.65***			3.98***		1.11***
rsv200	0.47***		0.92***			1.41**		
rsq025						Exc	0.08	Exc
rsq050		0.29	Exc			Exc	1.38***	1.14***
rsq100		1.97***	2.65***			3.36***		
rsq200						0.87		

Low high impact

[1] 'Exc' indicates non-significant variables excluded from the model. $P<0.001$ (***), $P<0.01$ (**), $P<0.05$ (*).

[2] See Supplementary Table A1 for abbreviations of the codes.

Table 8 depicts the relative and absolute differences pointed out by each model predicting soybean expansion in 2014. The intersection between quartiles (displayed on the right column) indicates the level of accuracy of each model's predictions. The smaller the intersection, the greater the ability of the model to distinguish areas of soybean expansion from others. Model 8 presented the largest differences between the mean of converted vs. non-converted areas. Concerning median values, Model 3 presented the largest differences between converted vs. non-converted areas. Model 2 showed the smallest overlap between the first and third quartiles (1%). Additional model comparisons can be found in Supplementary Figures A3 and A4.

The comparison of the *roc* curves (Figure 4) and the area under each of them once again suggested that Model 2, based on variables extracted from component 2 of the PCA, was the best predictor of soybean expansion which occurred in 2014. The area under the curve of Model 2 was 0.8157, followed by Model 1 (0.8083); Model 3 (0.7972); Model 8 (0.7949); Model 5 (0.7553); Model 6 (0.7453); Model 7 (0.7346); and finally Model 4 (0.6154).

3.4 Sensitivity analysis

The heat maps in Figure 5 allow us to visualize the differences and similarities across each model projection of soybean expansion. Clearly, economic variables alone (Model 4) did not show a clear spatial pattern of the probability of soybean expansion. Although no sharp contrast can be seen across models, Model 5 (based

Table 8. Comparison of the probability of soybean expansion occurrence vs. non-occurrence in 2014, as predicted by each regression model.

Model	**# variables**	**Median**		**Mean**		**Intersection between 1st and 3rd quartiles**
		Abs. difference	**Rel. difference**	**Abs. difference**	**Rel. difference**	
1	10	0.04347	148%	0.03556	92%	8%
2	5	0.02108	87%	0.02068	67%	1%
3	14	0.04496	154%	0.03925	96%	14%
4	2	0.00472	11%	0.00578	14%	73%
5	5	0.02049	91%	0.01527	55%	24%
6	8	0.01536	58%	0.01544	56%	28%
7	5	0.02011	60%	0.01928	55%	36%
8	21	0.04368	142%	0.04602	109%	18%

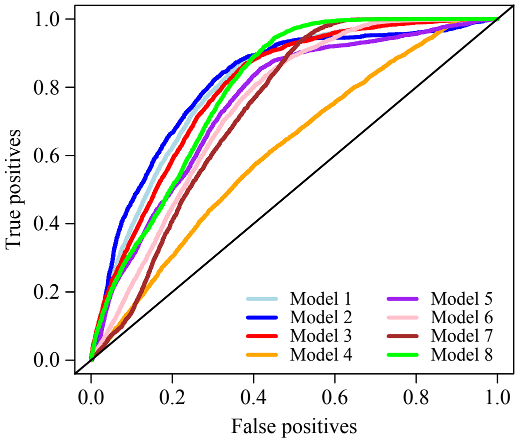

Figure 4. Roc curves showing the explanatory power of each simulated model.

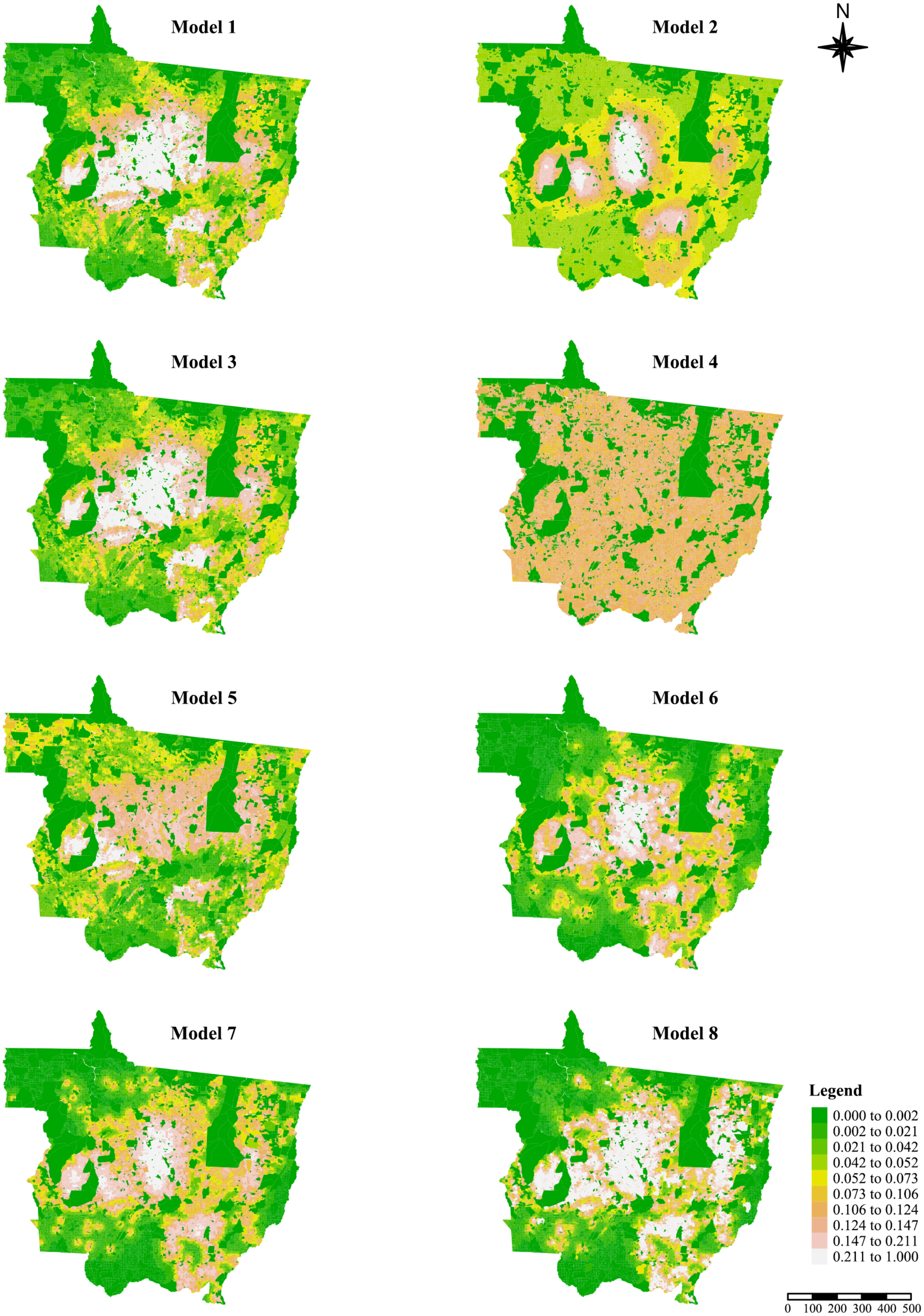

Figure 5. Heat maps showing the probability of soybean expansion across different regions of Mato Grosso as predicted by regressions models 1-8.

on environmental/agronomic variables) is the most discrepant, pointing out a relatively high probability of land conversion in the Northwest of Mato Grosso (in the Amazon region). Models 6 and 7 suggested some heat points in areas located near existing soybean fields.

The best predictive models (1, 2, 3 and 8) are quite similar with regard to the highest probabilities of soybean expansion around the center of the state, near the highway BR-163. The main distinction found between Model 2 and the others was the homogeneity of probabilities around conversion hotspots.

4. Discussion

Our results elicited some hotspots of soybean expansion and suggested the persistence of such trend in the future. In more traditional soybean production regions, where infrastructure already exists and the soybean supply chain is more developed, some expansion is still expected due to agglomeration economy effects reported in the literature to favor the emergence of a positive feedback loop for agricultural development in existing production hubs (Garrett *et al.,* 2013). Further expansion in traditional production regions is also likely to be linked to yearly price variations and their impact on farmers' willingness to further invest in the farm.

Yet, as our analysis also indicates, a significant share of soybean expansion in the next years might take place in less traditional regions with adequate agronomic conditions and possibly lower land prices. The increase of the share of soybean expansion in farms with no previous soybean reflects farmers' reaction to positive economic return, followed by their decision to acquire new lands. This is aligned with other studies showing the greater liquidity of land markets over the same period (Informa Economics, 2013) and the dissemination of rental contracts involving the temporary conversion of pastures into cropland (Cohn *et al.,* 2016).

It is hard to disentangle the effect of different groups of variables (economic, agronomic, logistics, etc.) on farming decisions since there are interactions among them. For instance, even the models with no agronomic variables might implicitly account for them given that soybean is currently produced in regions with favorable natural conditions (as indicated by the similarity between the map of Model 5 and the others). Analogously, the proximity to siloes and good logistic conditions influence soybean production costs in certain regions. In addition, the explanatory power of economic variables is relative to the explanatory power of the other variables they are analyzed with (mainly logistics-related variables, due to their influence on the cost of labor and other inputs). Nevertheless, the proximity to existing soybean fields and to supply chain infrastructure were found to be quite influential of soybean expansion. The same goes for logistics variables: the distance from municipality to ports was not that crucial, but the distance from each farm to the nearest siloes and to other soybean farms were both significantly associated with land conversion. Even though farmers may perceive distance differently and be more or less tolerant to it, their willingness to invest in the construction of storage facilities decreases if the farm is already close to a commercial silo that can be rented (IMEA, 2016). Soybean expansion where storage facilities are already available is less risky.

The discussion of which of the variants is the most important is attractive, but it is dangerous to assume that one has more impact than the other. Some have an important weight in all models that have been present, such as the variable roo and csl. Others had an even greater impact in some models, such as dmia, but it was a variable not selected by the PCA models. There were also cases of variables that had a very high importance in some models, but of less relevance in others, such as ra050. Finally, some variables were important only in extreme situations, when in their maximum or minimum, as the case of variables lvr and pst.

Regardless of the effect that each individual variable may have on soybean expansion, our analysis shows the importance of looking into them in an integrated manner. Despite how influential potential costs and revenues are for farmers' land use decisions (our sensitivity analysis showed that changes in profitability were the main source of variation for several models), economic variables alone are not good predictors of soybean expansion unless combined with other variables, such as agronomic indicators.

Our analysis of the importance of economic variables might have been partly compromised by the lack of an official, consistent database of farm boundaries. Exactly because of the importance of farm scale for soybean production, future mapping of soybean expansion should follow strict farm boundaries. Yet, the probabilities of conversion statistically modeled were well aligned with real-world observations in 2014. Similarly, variables that can proxy for cultural aspects as well as information on second harvest crops and complementary economic activities could lead to somewhat different conclusions concerning farmers' land use behavior. Despite the uncertainty on specific points, our regression models shows a higher probability of soybean expansion in two areas of Mato Grosso of high conservation value (the Amazon rainforest in the state's Northwest and in the Pantanal in the state's Center-South) thus requiring attention and revealing an opportunity to influence land use dynamics. While price oscillations are harder to predict and act upon, investments in logistics – a crucial factor for soybean expansion – should be planned while taking that into account.

Our results should be carefully interpreted concerning compliance with the environmental legislation. Although a small share of the soybean expansion occurred in farms which did not appear in the CAR, LAU and/or GEO databases, this does not necessarily reflect farmers' compliance. In fact, those who have established a 'Term of Conduct Adjustment' (TAC – in Portuguese: '*Termo de Ajuste de Conduta*') committing themselves to a compliance plan already appear in these databases irrespective of the environmental liability they bear. It could also be that the land conversion associated with soybean expansion took place before a 'TAC' was established. Similarly, the fact that some farmers do not appear in these databases does not necessarily mean that they failed to set aside a share of their properties for conservation purposes.

The fact that soybean prices were generally high throughout the entire analyzed period limits our understanding of the influence of economic losses or less favorable market prices on the expansion or contraction of soybean farming. On the other hand, it elicits how sensitive soybean expansion is to small profitability changes as well as other variables of interest, such as farmers' perception of production scale and how that affects production decisions. Although higher profitability levels tend to be associated with larger farms, we showed that this may vary due to yearly price oscillations. Throughout the analyzed period, 2012 was the year with the highest profitability (when a profit margin of 200% could be reached by farms of approximately 4500 ha) whereas 2010 showed the lowest profitability (requiring a larger soybean area to achieve the same profit margin as in other years). The fact that soybean expansion in 2012/13 happened in greater proportion in farms that already produced soybean can be largely explained by the expansion that had already taken place in 2011 and increasing land prices (up to 75% higher in some regions (Informa Economics, 2013)).

5. Conclusions

Investigating the extent, location and determinants of soybean expansion in Mato Grosso is a condition for the implementation of effective and sustainable land use policies. Our maps and statistical analyses allowed us to (1) characterize soybean expansion in recent years at a detailed level, (2) understand the importance of environmental, agronomic and logistic factors in determining the probability of land conversion, and (3) highlight how sensitive soybean expansion can be to price changes in the short run.

Land use policies that allow the state to prevent irregular agricultural expansion through illegal deforestation – and at the same time provide the necessary infrastructure for the sustainable development of agribusinesses – should be prioritized. As this paper and other studies have shown (Aguiar, 2006; Chakir and Parent, 2009), the improvement of the supply chain infrastructure through the introduction of paved roads, for example, can foster agricultural production of higher economic value. These land use transitions could bring environmental gains as well, such as the restoration of formerly degraded pastures, although impacts on carbon emissions, water consumption and biodiversity must be assessed carefully.

Some policies focused on the creation and/or expansion of conservation units and indigenous lands are currently being discussed at the federal, state and municipal levels. The results of our analysis can aid these discussions by providing evidence on the evolution of productive chains and what drives or constraints further agricultural expansion.

It is important to promote the development of services and industries in regions where the potential for soybean expansion already exists, thereby reducing production costs, adding value to the final product and reducing the pressure for soybean expansion where environmental conservation should be the focus.

Alternative methodological approaches can help us understand land use changes and could complement the analysis presented here. For instance, social data can be explored to understand and forecast farmers´ intentions concerning management decisions (Pocewicz, 2008), just like land prices and other market-related information may act as complementary predictors of land use transitions (Ustaoglu, 2016). Additional research avenues identified throughout this study include sensitivity analyses involving refined logistics parameters of alternative transport modes and storage facilities. Several projects for the construction of waterways and railroads are already underway and have attracted investments from local and international players of the soybean supply chain. Given that transport accounts for a large share of soybean production costs in Mato Grosso, logistics are crucial for predicting crop farming expansion and assessing the environmental pressure that may arise from it. Finally, soybean expansion in frontier regions may be a distinctive process due to land prices, market speculation and actors' goals, thus warranting further investigation.

Supplementary material

Supplementary material can be found online at https://doi.org/10.22434/IFAMR2018.0072.

Table A1. Complete list of variables hypothesized as relevant for the expansion of soybean in Mato Grosso.
Materials and methods. Economic variables and modeling parameters
Figure A1. Histograms depicting the distribution of probability of soybean expansion in each statistical model.
Figure A2. Boxplots of the probability distribution of converted vs. non-converted areas in 2014.
Figure A3. Histogram of the probability distribution of areas converted in 2014.
Figure A4. Boxplots of the probability distribution of soybean conversion according to each regression model for converted areas.

Acknowledgments

We would like to thank Mato Grosso's Institute of Agricultural Economics (IMEA) for the support and data provided.

References

Aguiar, A.P.D. 2006. *Modelagem de mudança do uso da terra na amazônia: explorando a heterogeneidade intraregional.* Sensoriamento Remoto – Instituto Nacional de Pesquisas Espaciais, São Paulo, Brazil.

Aguiar, A., G. Câmara and M.I.S. Escada. 2007. Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: exploring intra-regional heterogeneity. *Ecological Modelling* 209(2-4): 169-188.

Alexandratos, N. and J. Bruinsma. 2012. World agriculture towards 2030/2050: the 2012 revision. *FAO: ESA Working paper*, *4*. Food and Agriculture Organization, Rome, Italy.

Almeida, P.S., A. Tiosso and J.S.F. Silva. 2014. Análises e considerações sobre as influências do código florestal Brasileiro na política nacional sobre mudança do clima: lei no. 12.651/12 e suas alterações. *Revista de Estudos Ambientais* 15(2): 16-24.

Anderson, L.O., E.H.M. Rojas and Y.E. Shimabukuro. 2003. Avanço da soja sobre os ecossistemas cerrado e floresta no Estado do Mato Grosso. In: *Simpósio Brasileiro de Sensoriamento Remoto,* pp. 19-25.

Bowman, M.S., B.S. Soares-Filho, F.D. Merry, D.C. Nepstad, H. Almeida and O.T. Rodrigues. 2012. Persistence of cattle ranching in the Brazilian Amazon: a spatial analysis of the rationale for beef production. *Land Use Policy* 29(3): 558-568.

Camargo, L. 2011. *Atlas de Mato Grosso: abordagem socioeconômico-ecológica.* Entrelinhas, Cuiabá, Brazil.

Carvalho, T.S., A.S. Magalhães and E.P. Domingues. 2016. Desmatamento e a contribuição econômica da floresta na Amazônia. *Estudos Econômicos* 46(2): 501-533.

Chakir, R. and O. Parent. 2009. Determinants of land use changes: a spatial multinomial probit approach. *Papers in Regional Science* 88(2): 327-344.

Cohn, A.S., J. Gil, T. Berger, H. Pellegrina and C. Toledo. 2016. Patterns and processes of pasture to crop conversion in Brazil: evidence from Mato Grosso State. *Land Use Policy* 55: 108-120.

Conab. 2015. *Companhia Nacional de Abastecimento*. Available at: http://www.conab.gov.br/.

do Nascimento, A. P. P., Figueiredo, A. M. R., & Miranda, P. R. 2018. "Dimensão do PIB do agronegócio na economia de Mato Grosso." *Ensaios FEE*: 903-930.

Federação das Indústrias do Estado de São Paulo (Fiesp). 2015. *Outlook Fiesp. 2025: projeções para o agronegócio brasileiro.* Available at: https://tinyurl.com/ybuf6dlr.

Garrett, R.D., E.F. Lambin and R.L. Naylor. 2013. The new economic geography of land use change: supply chain configurations and land use in the Brazilian Amazon. *Land Use Policy* 34: 265-275.

Informa Economics, F.N.P. 2013. *ANUALPEC 2013: anuário da pecuária brasileira.* São Paulo, Brazil.

Ibge. 2015. *Instituto Brasileiro de Geografia e Estatística.* Available at: http://www.sidra.ibge.gov.br/.

IMEA. 2015a. *Instituto Mato-Grossense de Economia Agropecuária*. Available at: http://www.imea.com.br/.

IMEA. 2015b. *Projeções do Agronegócio em Mato Grosso para 2025.* Instituto Mato-Grossense de Economia Agropecuária, Cuiabá, Brazil.

IMEA. 2016. *Fast tool Armazenagem.* Available at: http://www.aprosoja.com.br/produtor/simuladores/.

Jasinski, E., D. Morton, R. DeFries, Y. Shimabukuro, L. Anderson and M. Hansen. 2005. Physical landscape correlates of the expansion of mechanized agriculture in Mato Grosso, Brazil. *Earth Interactions* 19(6): 1-18.

Macrologística. 2013. *Projeto Centro-Oeste competitivo*. Available at: http://www.ampa.com.br/arquivos/publicacoes/19122013032704.pdf.

Mann, M.L., R.K. Kaufmann, D. Bauer, S. Gopal, M.D.C. Vera-Diaz, D. Nepstad and G.S. Amacher. 2010. The economics of cropland conversion in Amazonia: the importance of agricultural rent. *Ecological Economics* 69(7): 1503-1509.

Mann, M.L., R.K. Kaufmann, D.M. Bauer, S. Gopal, M. Womack, J.Y. Nomack and B.S. Soares-Filho. 2014. Pasture conversion and competitive cattle rents in the Amazon. *Ecological Economics* 97:182-190.

Mapa. 2016. *Projeções do Agronegócio: Brasil 2015/2016 a 2025/2026. Projeções de longo prazo.* Mapa/Acs. Prod. Assessoria de Gestão Estratégica, Brasília, Brazil.

Masuda, T. and P.D. Goldsmith. 2009. World soybean production: area harvested, yield, and long-term projections. *International Food and Agribusiness Management Review* 12(4): 143-161.

Moreira, M.L. 2007. *Mato Grosso: solos e paisagens.* Entrelinhas, Cuiabá, Brazil.

Organisation for Economic Co-operation and Development (OECD). 2015. *Environmental performance reviews: Brazil 2015.* OECD Publishing, Paris, France.

Pocewicz, A., M. Nielsen-Pincus, C.S. Goldberg, M.H. Johnson, P. Morgan, J.E. Force, L.P. Waits and L. Vierling. 2008. Predicting land use change: comparison of models based on landowner surveys and historical land cover trends. *Landscape Ecology* 23(2): 195-210.

Rivero, S., O. Almeida, S. Ávila and W. Oliveira. 2009. Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento na Amazônia. *Nova Economia* 19(1): 41-66.

Simões, M., S. Pereira and R. Ferraz. 2014. *Geração de cenários de mudança de uso do solo na amazonia legal brasileira em função do agronegócio e da aplicação de políticas públicas. Anais do Simpósio Regional de Geoprocessamento e Sensoriamento Remoto*. Geonordeste 2014. Aracaju, Brazil, pp. 18-21.

Ustaoglu, E., C. Perpiña Castillo, C. Jacobs-Crisioni and C. Lavalle. 2016. Economic evaluation of agricultural land to assess land use changes. *Land Use Policy* 56: 125-146.

Veldkamp, A. and E.F. Lambin. 2001. Predicting land-use change. *Agriculture, Ecosystems & Environment* 85: 1-6.

Verburg, R., S. Rodrigues Filho, D. Lindoso, N. Debortoli, G. Litre and M. Bursztyn. 2014b. The impact of commodity price and conservation policy scenarios on deforestation and agricultural land use in a frontier area within the Amazon. *Land Use Policy* 37: 14-26.

Verburg, R., S. Rodrigues Filho, N. Debortoli, D. Lindoso, I. Nesheim and M. Bursztyn. 2014a. Evaluating sustainability options in an agricultural frontier of the Amazon using multi-criteria analysis. *Land Use Policy* 37: 27-39.

Weinhold, D. and E. Reis. 2008. Transportation costs and the spatial distribution of land use in the Brazilian Amazon. *Global Environmental Change* 18(1): 54-68.