The drivers of spatial cropping patterns in Burgundy

Valentin Bellassen, Jean-Sauveur Ay, Mohamed Hilal

Centre d'Economie et de Sociologie
appliquées à l'Agriculture et aux Espaces Ruraux

The drivers of spatial cropping patterns in Burgundy

Valentin Bellassen[1], Jean-Sauveur Ay[1], Mohamed Hilal[1]

Working Paper
2020/1

[1] CESAER, AgroSup Dijon, INRAE, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France.

Centre d'Economie et de Sociologie
appliquées à l'Agriculture et aux Espaces Ruraux

Abstract

This study quantifies the relative influence of pedo-climatic conditions – natural advantage – and market proximity in the determination of crop location in Burgundy. Econometric models are designed to study the share or presence of six different crops – barley, rapeseed, soy, proteaginous pea, hemp and lentils – at municipal scale. Market proximity is found to have a significant influence on the location of four of them, namely barley, rapeseed, proteaginous pea and hemp. Market proximity increases the share of barley in arable crops by more than 10 percentage points and its influence on the share of barley is as important as pedo-climatic conditions. The influence of malt plants is highest within 80 minutes from plant location. For hemp, the inclusion of market proximity variables in the model increases the phi correlation coefficient from 0.14 to 0.18 and strongly constrains the areas predicted as suitable for hemp. We also derive from these six test cases two likely feasibility conditions for the quantification of market proximity influence on crop location: markets must be crop-specific (eg. hemp factories, which process only hemp) and they must be present in two different locations to limit the multicollinearity and endogeneity problems associated with distance variables. Low crop abundance does not seem to hamper the implementation of our method. This method could be used to refine existing downscaling procedures in economic models of agricultural supply such as AROPAj or CAPRI.

Keywords:

Cropping patterns; market proximity; natural advantage; partial equilibrium models; agricultural sector; downscaling procedure;

1. Introduction

Asking why an economic activity occurs here rather than there is almost as old as modern economics. Ricardo's famous example of wine and sheep points to natural advantage – climate, soil, … – as a key explanation for the location of a given production. Von Thünen stressed the distance to the buyer as another key structuring variable. Despite the drastic reduction in transportation costs since von Thünen's time, these costs remain relatively high in the agricultural and food sectors (Kilkenny, 1998). In the case of ethanol production assessed in Kocoloski et al. (2011), a modest optimization of factory distance to feedstock – from uniform to optimized distribution of 8 factories in Illinois – translates into a 10% decrease in total production costs.

More recently, studies pointed out that increasing returns (Krugman, 1991) can lead to higher concentration of agricultural type (Chatellier and Gaigné, 2012; Daniel, 2005) or even crop type (Holmes and Lee, 2012) than warranted by natural advantage. Moreover, this concentration can outlive the natural advantage (Bleakley and Lin, 2010), locking a territory into its initial specialization (Magrini et al., 2016). Increasing returns explain why the agro-industrial context – proximity to a crop-processing factory, location within the collecting basin of a given crop collector, … – may be a key driver of crop location. Crop collectors can be reluctant to equip their silos for the specific needs of minor crops (Fleurat-Lessard, 2013). However, the empirical papers related to concentration of agricultural activity (Ben Arfa et al., 2009; Chatellier and Gaigné, 2012; Chevassus-Lozza and Daniel, 2006; Daniel, 2005; Holmes and Lee, 2012) do not investigate the drivers of this concentration.

In addition to increasing returns, the concentration of agricultural types generates negative environmental externalities. Nitrate-related water pollution caused by the concentration of animal productions (eg. Gaigné et al. (2012)) is probably the most famous example, but concentration is also dominant within the arable crops sector (CGDD, 2012). Crop scientists point to diversified crop rotations as a necessity in environmentally friendly crop systems such agro-ecology or organic farming (Doré et al., 2011; Malézieux, 2012; Meynard et al., 2016). At farm level, this diversification seemingly comes at little to no cost (Davis et al., 2012; Lechenet et al., 2014). Yet, farm-level studies assume that diversification crops can be sold at their market price and therefore neglect the collection and transportation costs associated with collecting a new crop in a territory where it was not collected before. Another reason to investigate the impact of market proximity on crop location is therefore to assess whether the increasing returns associated with these collection, transportation and processing costs represent a significant barrier to crop diversification and its associated environmental benefits.

Most of the existing literature on the drivers of the location of agricultural activity has focused on larger levels than crops such as agricultural practice (Geniaux et al., 2009; Schmidtner et al., 2011), type of crop collection organization (Triboulet et al., 2013), quality signs (Magrini et al., 2011) or land-use type (Chakir and Le Gallo, 2013). Triboulet et al. (2015) show that at least for cooperatives, food processing plants are correlated to the type of agricultural production in their neighbourhood. For the location of individual crops, most of existing studies focus on bio-physical determinants and global economic or policy conditions (Chakir, 2009; Hendricks et al., 2014; Kempen et al., 2011; Wu and Segerson, 1995), with a few recent exceptions (Garrett et al., 2013; Miao, 2013; Motamed et al., 2016).

The lack of market proximity variables for crop location in partial equilibrium models of the agricultural sector (Britz et al., 2011; e.g. Galko, 2007) can in turn overestimate the impact of public incentives to grow a specific crop (Alexander et al., 2014).

The objective of this study is to identify the main drivers of crop location in Burgundy. The main hypothesis being tested is that natural advantage – understood as pedo-climatic conditions – is not sufficient to explain where crops are located. To this end, several econometric models of the presence or share of six crop types within a municipality are estimated. This study adds two original features to the few existing studies on this topic (Garrett et al., 2013; Miao, 2013; Motamed et al., 2016): the issue is assessed for more than one crop in a given area – here the Burgundy region, and the decrease of market influence with increasing distance is endogenously estimated whereas existing studies exogenously prescribe the zone of influence of each intermediate market (eg. 50 km radius around an ethanol plant). The method developed here could be used to refine existing downscaling procedures in economic models of agricultural supply such as AROPAj (Chakir, 2009) or CAPRI (Kempen et al., 2011), although data availability on processing plant location would likely limit its application to the regional or national – rather than continental – scales.

2. Methods

2.1. Theoretical framework

Carpentier and Le Tort (2014) provide a simple theoretical background for the multinomial logit model, a classic in econometrics to derive acreage shares since Lichtenberg (1989) and Wu and Segerson (1995) (Equation 1).

Equation 1. $\ln\left(\frac{share_{i,j,t}}{share_{i,j0,t}}\right) = \beta_j X_{i,t} + \varepsilon_{i,j,t}$

where $share_{i,j,t}$ is the average area share of crop j in location i over period t, j0 is an arbitrarily set reference crop and $X_{i,t}$ are the independent variables.

The determinants $X_{i,t}$ of the relative profitability of crops – and hence their relative acreage share – for any given location i which are traditionally considered are fixed inputs such as farm endowment in land (both quantitatively and qualitatively through pedo-climatic conditions) and input and output prices (Garrett et al., 2013; Hendricks et al., 2014; Kempen et al., 2011; Wu and Segerson, 1995). But input and output prices are generally considered to vary only in time, not in space (Hendricks et al., 2014; Wu and Segerson, 1995).

Here we try to separately estimate the classical determinants of crop acreage share – fixed farm endowments – from determinants related to market proximity. In other words, we expect that proximity to intermediate markets – silos and processing plants buying raw agricultural products – generates some additional spatial variability in relative crop profitability, through variability in farm-gate output prices. We use a single time-period t for each crop j and "all other crops" as the reference crop j0 (see section 2.2.1). The standard multinomial logit model is therefore re-expressed as Equation 2.

Equation 2. $\ln\left(\frac{share_{i,j}}{1-share_{i,j}}\right) = \beta_j Pedoclim_i + \gamma_j Other\_controls_{i,j} + \sum_k \delta_{j,k} Market_{i,j,k} + \varepsilon_{i,j}$

where $share_{i,j}$ is the average area share of crop j in location i, $Pedoclim_i$ and $Other\_controls_{i,j}$ are the quasi-fixed exogenous inputs which in our case mainly come down to pedo-climatic conditions and $Market_{i,j}$ is the proximity of the intermediate markets for crop j. Because we have a specific interest in assessing how fast the effect of market proximity decreases with increasing distance, we use a generalized additive model (GAM) for $Market_{i,j}$ which estimates specific distance effects for a series of k distance intervals. The number of intervals is endogenously computed by the R software, balancing data fit and number of parameters (Wood, 2015). This flexibility is original compared to the similar studies which either use a linear approximation (Garrett et al., 2013) or prescribe a specific distance (Motamed et al., 2016).

For the econometric estimation of Equation 2, one could question the exogeneity of $Market_{i,j}$. One could indeed imagine that the location of a processing plant depends on the location of its feedstock as much as crop location depends on the existence of a local demand for it, creating simultaneity bias. As in Garrett et al. (2013), we neglect this possible bias, assuming that pedo-climatic variables explain most of the local systematic suitability for a given crop and thus limit the correlation between $Market_{i,j}$ and $\varepsilon_{i,j}$. Relaxing this assumption would be a priority for follow-ups to our study (see section 4.4).

2.2. Data

2.2.1. Crops of interest: two common crops and four diversification crops

To assess the feasibility of our method, six crops were tested: two common crops, barley and rapeseed, each totaling more than 15% of arable crop area on average, and four *diversification* crops, soy and minor oleaginous crops (more than half of it being soy), proteaginous crops (more than 90% of it being proteaginous pea), hemp and lentils, each totaling less than 2% of arable crop area on average. The pooling of minor oleaginous crops and of proteaginous crops is a constraint from the data on municipal crop shares provided by the observatory on rural development (RPG, 2013). The original source is the declaration of farmers to their local authority in order to receive Common Agricultural Policy subsidies. *Diversification* crops can, by definition, represent a marginal share of the Utilized Agricultural Area (UAA). Yet, market access is likely a more important barrier for them than for common crops. This is why the method is also assessed on rare crops.

The spatial extent of the analysis is the Burgundy region, in East central France. In 2013, the region totaled 0.9 million hectares of annual crops, 0.8 million hectares of grassland and 36 thousand hectares of permanent crops, mostly vineyards. Wheat, barley and rapeseed were by far the most frequent crops, summing up to three fourth of the regional arable land (SAA, 2015).

As a dependent variable, a three-year average municipal crop share is used. Three years correspond to the typical rotation length in Burgundy (Aouadi et al., 2015). Within the time series available (2006-2012), we choose the three years so that they best represent a "long-term average", that is as recent as possible if the annual area is stable and before the recent large increases in proteaginous pea, soy and lentils (Figure 1). This results in different periods: 2010-12 for rapeseed, barley and soy and 2007-09 for proteaginous pea, hemp and lentils.

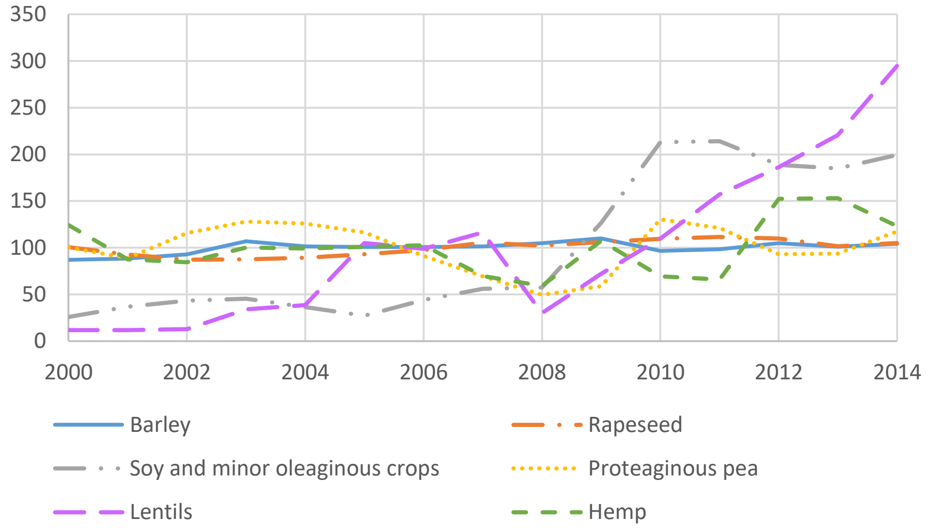


Figure 1. Cropped area of selected crops in Burgundy (100 = temporal average). Source: SAA (2015).

Common crops are more homogeneously distributed over Burgundy, as highlighted by their lower Gini coefficient (**Table 1**, supplementary material 8.1). They also tend to be more spatially correlated: although their Moran's I is higher than the Moran's I other diversification crops, although proteaginous pea and other oleaginous exhibit a relatively high degree of spatial auto-correlation.

Descriptive statistics of dependent variables

	Description	N	Min	Mean	Median	Max	% zeros	Gini coefficient	Moran's I
barl_1012	Average share of barley in municipal arable crops area over 2010 - 2012	2,017	0	0.19	0.19	0.78	2.58	0.30	0.63
rape_1012	Average share of rapeseed in municipal arable crops area over 2010 - 2012	2,017	0	0.16	0.18	0.41	15.02	0.36	0.72
soy_1012	Average share of other oleaginous in municipal arable crops area over 2010 - 2012	2,017	0	0.01	0	0.38	77.94	0.91	0.47
prot_0709	Average share of proteaginous crops in municipal arable crops area over 2007 - 2009	2,017	0	0.01	0	0.19	56.42	0.79	0.22
hemp_0709	Average share of hemp in municipal arable crops area over 2007 - 2009	2,017	0	0.0003	0	0.09	96.63	0.99	0.08
lent_0709	Average share of lentils in municipal arable crops area over 2007 - 2009	2,017	0	0.0003	0	0.06	94.10	0.98	0.13

Table 1. Descriptive statistics of crops of interest. Source: RPG (2013).

2.2.2. Potential drivers of crop location

Market proximity

The market for a given crop is defined here as the closest transformation factory of the raw crop. For barley for example, transformation factories consist mainly in malt plants in or within 50 km of Burgundy. *Factory proximity* is therefore defined as the transportation time by road from a municipality to the closest transformation factory. When a holding which owns transformation factory also has a collecting subsidiary, as second type of proximity is assessed: *collector proximity*, defined as the transportation time by road to the closest silo belonging to a firm which also owns a transformation factory. For barley for example, all malt plants are operated by *Soufflet*, a firm which also collects and sells raw crops and therefore operates silos.

Factory proximity is the most intuitive proxy of intermediate market proximity whereas *collector proximity* adds the notion of *organizational proximity* into it: for equally distant territories, do factories preferentially buy feedstock from collecting subsidiaries of their own holding?

One could imagine other types of organizational proximities than belong to the same holding. Yet, the effect of collectors could not be systematically tested. The only data available was the location of their larger silos. And at regional scale, the distance to the closest larger silo of a given collector tends to be strongly correlated both with the distance to a type of agro-food plant and with the distance to the closest larger silo of a neighbouring collector. This is why, unless a restrictive hypothesis on organizational proximity can be formulated, namely belonging to the same holding, multicollinearity prevents the characterization of the effect of specific collectors.

The importance of processing varies between supply chains, and so may vary the importance of this intermediate market as a driver of crop location. The six crops assessed here however are seldom sold unprocessed.

Control variables

Classical – mostly pedo-climatic – drivers were listed based on literature review. Variables possibly related to self-consumption or local exchanges through animal feed, such as the number of livestock units per hectare of AAU or the location within a protected designation of origin, also appear in this list. Because these variables are likely endogenous, they are not included in the preferred models, but they are used to test the sensitivity of the results to their inclusion.

Type of independent variable	Description and reference
Market proximity ($Market_{i,j}$)	
Factory proximity	Transportation time by road to the nearest transformation factory. Location of factories is obtained from public statistics (INSEE, 2013) and correspond to non-artisanal factories officially active on 01/01/2012, with at least one employee[2]. When the official type of the factory is too wide to be paired with a given crop (eg. "refined oil" which encompasses all products from fatty liquids from nut oil to biodiesel), internet was used to determine whether the factory was likely to process the crop of interest. Transportation time is computed with Odomatrix (2013).
Collector proximity	Transportation time by road to the nearest major silo of a collector whose holding also owns a transformation factory. Major silos are defined as silos with a storage capacity higher than 5,000 m^3 belonging to a collector owning either more than two silos having a total storage capacity higher than 100,000 m^3 within 30 minutes of Burgundy. Their location is obtained from the national database on installations presenting an industrial risk (MEEM, 2014). Transportation time is computed with Odomatrix (2013).
Pedo-climatic variables ($Pedoclim_{i,j}$)	
Climate	Four climatic variables selected after soliciting expert knowledge on the most important pedo-climatic drivers of crop yield in Burgundy and after eliminating variables with more than 0.8 inter-correlation: number of freezing days, spring temperature, spring rainfall and spring relative humidity. Average values over 1991-2011 are retained from Météo-France (2013).

[2] This threshold allows to eliminate factories which have stopped processing but are not administratively closed.

Soil	Available water content is obtained from Ubertosi et al. (2009). Share of irrigated area is obtained from agricultural census (Agreste, 2010). All other soil variables were obtained from soil inventory data (Donesol, 2006), after applying a few quality check and gap-filling procedures.
Topography	Ruggedness, defined as the square sum of elevation differences between a pixel and its eight neighbors, was computed from elevation data (IGN, 2010).
Other control variables ($Other\_control_{i,j}$)	
Geographical indications	The list of municipalities for each geographical indication (PDO or PGI) was provided by INAO (2015).
Livestock	Livestock-related variables are derived from agricultural census (Agreste, 2010), using gap-filling procedures when necessary.

Table 2. Independent variables and their source. The exhaustive list of independent variables retained in the six preferred models – one model per crop – is provided in supplementary material 8.5).

Table 2 summarizes the independent variables and their origin.

2.3. Econometric specification

The econometric specification is fitted to the information richness – variability, number of zeroes – of the dependent variable:

- ✓ for barley, the crop with the richest information, the classical *share model* is applied;
- ✓ for rapeseed which is burdened by a non-negligible amount of zeroes, a selection model is designed;
- ✓ and for diversification crops, for which most of the information lies in whether the crop is farmed in a municipality, a coarser probit model is applied.

2.3.1. Barley

Distribution of the share of barley at municipal level is Gaussian with no specifically strong occurrence of zeros and covers a large range of values from 0 to 78% of municipal UAA (**Table 1** and supplementary material 8.1). Accordingly, the classical *share model* (Equation 2) can be applied with minimal changes[3].

Once the coefficients $\delta_{j,k}$ of the non-linear response to $Market_{i,j}$ have been estimated, the partial effect at the average of each independent variable X_k on $share_{i,j}$ is be computed from Equation 3.

Equation 3. $$share_{i,j} = \frac{e^{\beta_j X_i + \varepsilon_{i,j}}}{1+e^{\beta_j X_{i,j} + \varepsilon_{i,j}}} \Rightarrow \frac{\partial y}{\partial X_k}\Big|_{\overline{X}} = \frac{\frac{\partial \beta_j}{\partial X_k}(X_k) \times e^{\beta_j X_i + \varepsilon_{i,j}}}{\left(1+e^{\beta_j X_i + \varepsilon_{i,j}}\right)^2}$$

where $\overline{X}$ stands for "all independent variables except X_k taking their average value over the dataset".

[3] The Smithson and Verkuilen (2006) transformation is applied to the actual share to allow for null values: $share_{i,j} = \frac{p_{i,j}(n-1)+0.5}{n}$

where $p_{i,j}$ is the actual area share of crop j among all arable crops and n is the number of observations.

2.3.2. Rapeseed

Rapeseed share also exhibits a large variability from 0 to 41% of municipal UAA and its distribution would be Gaussian if it was not for the large proportion of municipalities – 15% – without rapeseed (see supplementary material 8.1). Therefore, we restrict the dataset by excluding municipalities without rapeseed. To correct for selection bias, we then use a generalization of the Heckman model (Wooldridge, 2013) for nonparametric settings (Das et al., 2003). The model thus has two stages: a probit selection equation (Equation 4) followed by the classical *share model* corrected (Equation 5). Unlike Wooldrige (2013), this model allows for a non-linear function of the inverse Mills ratio which dispenses from the exclusion condition.

Equation 4. $\boldsymbol{P\left(share_{i,j} > 0 | X_{i,j}\right) = \Phi(\beta_j X_{i,j})}$

where $share_{i,j}$ is the average area share of crop j – here rapeseed – among all arable crops in municipality i over 2010-2012, Φ is the normal cumulative distribution function and $X_{i,j}$ are the independent variables (Pedoclim, Other_control and Market).

Equation 5. $\boldsymbol{\ln\left(\frac{share_{i,j}}{1-share_{i,j}}\right) = \beta_j X'_{i,j} + s\left(\lambda_{i,j}\left(X_{i,j}\right)\right) + \varepsilon_{i,j}}$

where $share_{i,j}$ is the average area share of crop j – here rapeseed – among all arable crops in municipality i over 2010-2012, and $X'_{i,j}$ are the independent variables and $\lambda_{i,j}$ is the inverse Mills ratio of the predicted values from Equation 4.

When the partial effect at the average is computed here, both $Market_{i,j}$ and $\lambda_{i,j}$ vary, not only $Market_{i,j}$ as for barley. In principle, the standard errors of β_j should be corrected although in most cases the correction does not result in large changes (Wooldridge, 2013). Because there is no existing R package to combine generalized additive models and selection models, we do not correct standard errors. As discussed in section 3.3, this lack of correction does not change the conclusions.

2.3.3. Diversification crops

Diversification crops – soy and minor oleaginous crops, proteaginous crops, hemp and lentils – are not found in the large majority of municipalities with arable crops. They are present only in 6-44% of municipalities and their maximum share of arable crops area varies between 6 and 38% (**Table 1**, supplementary material 8.1). Most of the information therefore lies in the presence or absence of these crops in a given municipality. Accordingly, as in Allaire et al. (2015), simple probit models (Equation 6) are estimated by maximum likelihood.

Equation 6. $\boldsymbol{P\left(share_{i,j} = 1 | X_i\right) = \Phi(\beta_j X_i)}$

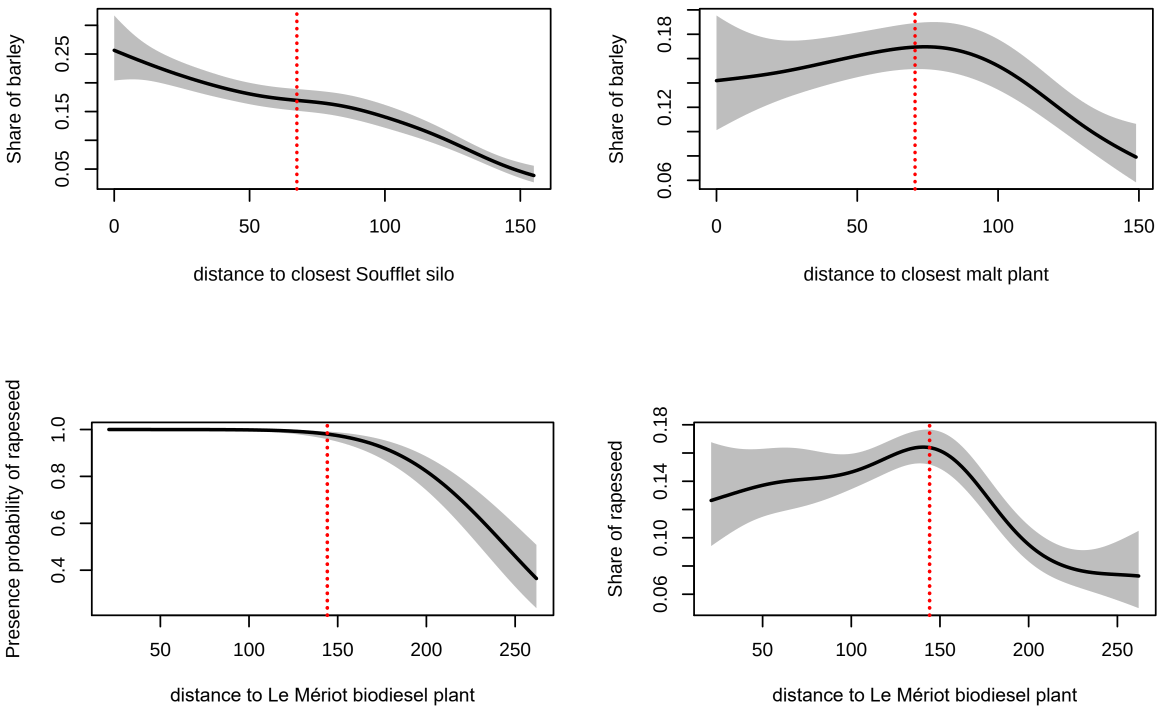
where $share_{i,j}$ is the average area share of crop j –other oleaginous, proteaginous, hemp and lentils respectively – among all arable crops in municipality i over 2010-12, 2007-09, 2007-09 and 2007-09 respectively, X_i are the independent variables and Φ is the cumulative distribution function of the standard normal law.

All estimations are computed with the R software, using in particular the following packages: *mgcv* (Wood, 2015) for generalized additive models, *stargazer* (Hlavac, 2015) for table display, and *ape* (Paradis, 2015) for Moran's I.

3. Results

3.1. Overview

The predicting ability of our econometric models for crop shares is around 0.35 (r^2) for both barley and rapeseed. The phi correlation coefficient of probit models – 0 if the model is not better than a random draw and 1 if the models perfectly predicts the observations – is however very variable: from 0.12 for lentils to 0.65 for *other oleaginous crops*. Among pedo-climatic independent variables, spring climatic conditions, rugosity and temporarily flooded areas are often significant. The significance and importance of market proximity variables varies between crops: they are significant for barley, rapeseed, proteaginous crops and hemp but not for other oleaginous crops and lentils. Interestingly, the shape of the relationship between market proximity and crop share or occurrence is similar across crops. The GAM always draws a concentric collection basin, often starting with a high plateau close to the factory and then decreasing for municipalities that are further than a specific threshold (Figure 2). This threshold varies from 50 minutes for feed factories in the cases of rapeseed and proteaginous to 150 minutes for biodiesel refineries (rapeseed), with intermediate values of 80 and 90 minutes for malt (barley) and hemp factories respectively.



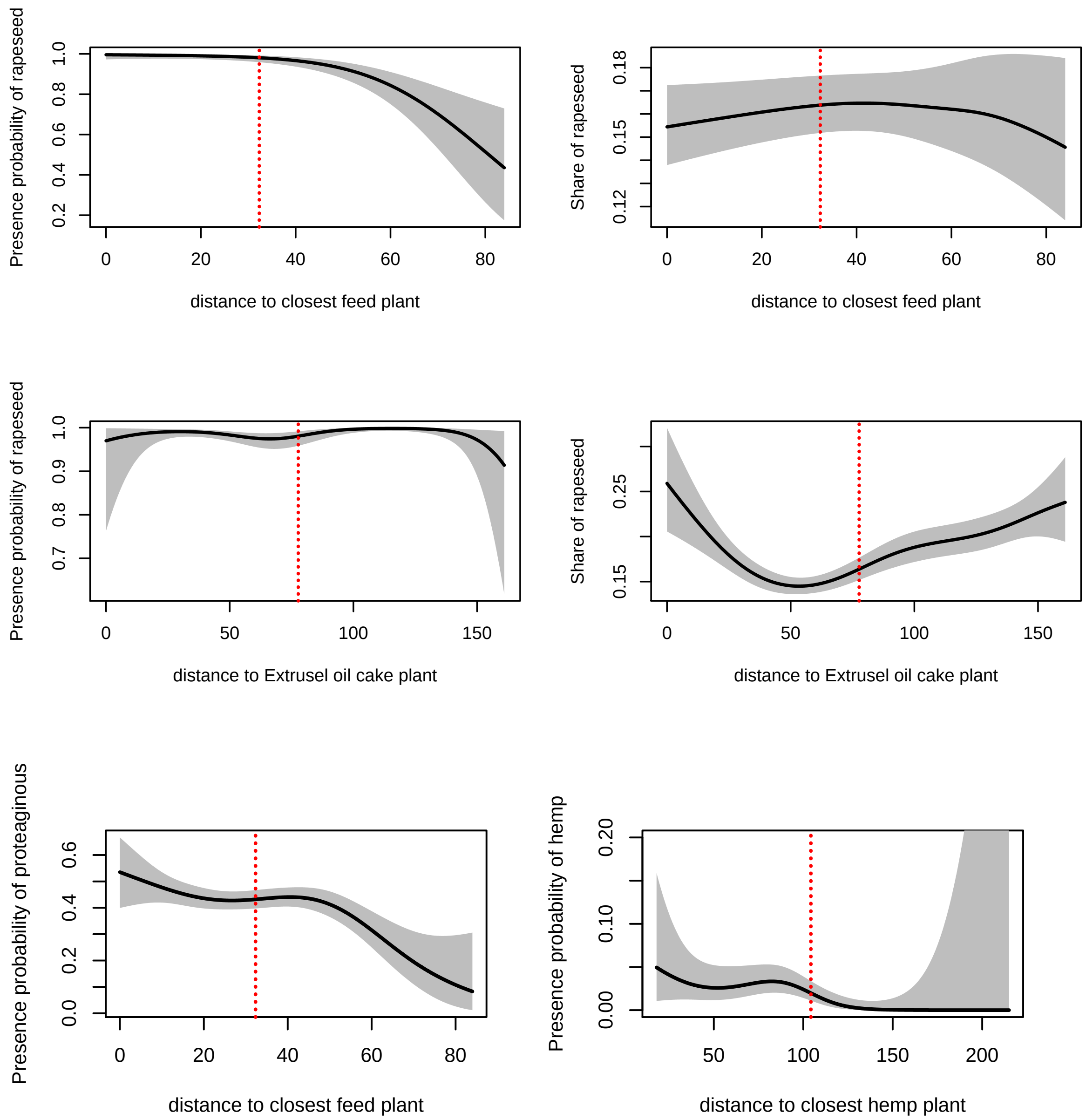


Figure 2. Partial effect at the average of distance (in minutes) to market

The red dotted line indicates the average value of the variable over our sample. The share is given as a proportion of municipal arable crops area. Shaded areas represent the 90% confidence interval under the Gaussian and homoscedasticity assumptions which only hold within 100 minutes of the silos.

3.2. Barley

Barley is one of the rare crops for which the effect could be tested because all malt plants within 50 km of Burgundy belong to the Soufflet group (see section 2.3.1). Both the OLS and the GAM models provide reasonable fits with an r^2 close to 0.35. Two meteorological variables – spring rainfall and freezing days, two pedological variables – soil pH and submerged area, and the two market proximity variables are significant.

Both the proximity to malt plants - *factory proximity* - and silos owned by Soufflet - *collector proximity* - significantly indeed increase the share of barley (**Table 3**). Furthermore, adding the distance to Soufflet silo into the model eliminates a positive bias in residuals over central-west Burgundy and reduces the spatial correlation in residuals (see supplementary material 8.4.1). This shows that complex vertical links in the supply chain can strongly affect crop location: out of a partial r2 of 0.05 for market variables, the Soufflet variable has a partial r^2 of 0.04 (**Table 4**).

This partial effect at the average of increasing distance to malt plants and Soufflet silos is intuitive: it monotonously decreases barley share. More specifically, the closest municipalities to Soufflet silos have an extra quarter of their arable crops area planted with barley compared to the most distant ones. The effect of malt plants is not linear: distance to malt plant has no effect up to the sample average – around 80 minutes, and then it decreases the share of barley from 17 % down to 5%. A similar pattern is observed when *collector proximity* is removed from the model.

More generally, the share of variance explained by market proximity variables is comparable to the share explained by pedo-climatic ones (**Table 4**). The latter seem consistent with an antagonism between barley and maize: barley is less frequent in municipalities with higher seasonally flooded areas and higher spring temperature, two factors which are favorable to maize.

Although rather small – 0.13 – and diminished by the introduction of the Soufflet variable, the Moran's I of residuals remains significant (see supplementary material 8.4.1). Accordingly, some omitted variables – such discussions among neighbours, being advised by the same local advisor, … – probably still generate spatial autocorrelation in the errors and inconsistency in the estimators.

barl_1012/(1-barl_1012) = f(BX)

	GAM	OLS
Intercept	-7.98 (3.08)**	-1.77 (2.89)
rugosity	-0.02 (0.01)	-0.02 (0.01)
soil pH	0.45 (0.04)***	0.41 (0.04)***
proportion of municipality occasionally under water	-1.09 (0.21)***	-1.03 (0.21)***
spring rainfall	-0.01 (0.00)**	-0.02 (0.00)**
spring relative humidity	0.04 (0.03)	-0.02 (0.03)
spring temperature	-0.03 (0.08)	-0.08 (0.07)
number of freezing days	0.04 (0.01)***	0.05 (0.01)***
available water content	-0.00 (0.00)	-0.00 (0.00)
distance to closest Soufflet silo	4.17 (5.23)***	-0.01 (0.00)***
distance to closest malt plant	3.07 (3.85)***	-0.00 (0.00)***
R^2	0.34	0.33
Num. obs.	2017	2017

***p < 0.001, **p < 0.01, *p < 0.05. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables.

Table 3. Estimation results for barley

	Partial r2
Pedo-climatic variables	0.07
Market proximity variables	0.05
distance to closest malt plant	0.01
distance to closest Soufflet silo	0.04

Table 4. Partial r2 in barley model

3.3. Rapeseed

The GAM selection model provide a reasonable fit ($r^2 = 42$). Eight pedo-climatic variables are significant either in the selection model or the share model (second step of the selection model).

Three types of plants significantly impact the rapeseed location: biodiesel plants, oil cake plants and animal food plants (**Table 5**). The biodiesel and oil cake plants have a similar impact on the probability of finding rapeseed in a municipality and on the share of rapeseed knowing that the municipality has some (Figure 2). Further than 150 minutes from the biodiesel plant, the probability of finding rapeseed drops rapidly and its share decreases from 16 % to 8 % of arable crops area. The effect of the oil cake plant is stronger but saturates more rapidly: rapeseed share decreases from 25% to 15% of arable crops area within 50 minutes of the plant.

Proximity to animal food plant however has a minor influence on rapeseed: it slightly increases the probability of finding rapeseed and has no influence on rapeseed share in municipality where rapeseed is present. Indeed, these plants are likely second processors producing feed mixes out of processed cakes.

The density of animals also has a negative impact on rapeseed share, in particular in municipalities belonging to a poultry protected geographical indication (see supplementary material 8.3.2). Two elements may explain this pattern: rapeseed needs to be processed before being used as animal feed and PGIs sometimes include local sourcing of animal feed in their requirements. Accordingly, farmers with animals may prefer to grow crops that can be directly fed to them – such as maize, especially when they have an incentive to use local feed.

Non-random selection, P(rape_1012=1) = f(BX) & share_trans(rape_1012) = g(B'X')

	Selection model	Share model	OLS
Intercept	-8.78 (4.83)	-0.59 (1.51)	-0.32 (1.48)
rugosity	-0.13 (0.03)***	0.01 (0.01)	-0.02 (0.01)*
spring rainfall	-0.04 (0.01)***	-0.03 (0.00)***	-0.03 (0.00)***
spring relative humidity	0.12 (0.06)*	-0.00 (0.02)	-0.01 (0.02)
share of municipal UUA which is irrigated	-0.03 (0.01)**	-0.00 (0.00)	-0.00 (0.00)
silt content	0.00 (0.00)**	0.00 (0.00)***	0.00 (0.00)***
soil pH	0.68 (0.17)***	-0.03 (0.05)	-0.00 (0.04)
proportion of municipality occasionally under water	-0.03 (0.52)	-0.66 (0.14)***	-0.78 (0.14)***
soil carbonate	-0.72 (0.43)	0.52 (0.10)***	0.74 (0.09)***
distance to Le Mériot biodiesel plant	1.00 (1.00)***	4.39 (4.83)***	-0.00 (0.00)**
distance to closest feed plant	1.87 (2.36)***	1.00 (1.00)	0.00 (0.00)
distance to Extrusel oil cake plant	3.99 (4.57)***	3.98 (4.57)***	0.00 (0.00)**
inverse Mills ratio		3.04 (3.82)***	
R2	0.64	0.42	0.35
Num. obs.	2017	1714	1714

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables.

Table 5. Estimation results for rapeseed

3.4. Other oleaginous crops

For other oleaginous crops than rapeseed or sunflower, that is again a category dominated by soy, the econometric model does not grasp fine-scale spatial heterogeneity. Despite prediction accuracy of 89%, models only manage to separate Burgundy into two geographically distinct zones: an Eastern patch of municipalities where soy is always predicted to occur and the rest of the region where it is always predicted not to occur (supplementary material 8.4.3). This pattern is obtained both with and without including market proximity variables in the models. In this context, it seems difficult to provide robust insights on the quantitative effects of market proximity.

3.5. Proteaginous crops

Similarly to *other oleaginous crops*, explaining the occurrence of proteaginous crops is challenging: the Phi correlation coefficient is moderate, even when endogenous variables such as animal density or technical orientation of the municipality are added to the model (supplementary material 8.3.4). Incorrectly predicted municipalities again tend to appear in clusters (supplementary material 8.4.3).

This last point suggests spatial auto-correlation and caution against over-interpreting the estimated parameters. That being said, the proximity of feed plants has a modest but significant effect on the occurrence of proteaginous crops (**Table 6**, Figure 2).

P(prot_0709=1) = f(BX)

	GAM probit	**OLS probit**	**OLS probit**
Intercept	10.68^{***}	11.14^{***}	10.21^{***}
	(1.08)	(1.09)	(1.03)
rugosity	-0.12^{***}	-0.12^{***}	-0.12^{***}
	(0.01)	(0.01)	(0.01)
spring rainfall	-0.03^{***}	-0.03^{***}	-0.04^{***}
	(0.01)	(0.01)	(0.01)
spring temperature	-0.49^{***}	-0.50^{***}	-0.44^{***}
	(0.07)	(0.07)	(0.06)
number of freezing days	-0.03^{**}	-0.02^{**}	-0.02^{*}
	(0.01)	(0.01)	(0.01)
proportion of municipality occasionally under water	-1.42^{***}	-1.47^{***}	-1.53^{***}
	(0.27)	(0.27)	(0.27)
distance to closest feed plant	3.25^{**}	-0.01^{**}	
	(3.87)	(0.00)	
Phi	0.31	0.30	0.30
Num. obs.	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

Table 6. Estimation results for proteaginous crops

3.6. Hemp

Explaining the occurrence of hemp crops is difficult: the Phi correlation coefficient is low. Nevetheless, the proximity of transformation plants within 50 km of Burgundy – namely "Eurochanvre" and "Chanvrière de l'Aube" – significantly increases the probability to find hemp (**Table 7**, Figure 2). The introduction of these market proximity variables increases the phi correlation coefficient from 0.14 to 0.18. Visually, it also stringently constrains the areas where hemp would be predicted to occur based on pedo-climatic variables alone (see supplementary material 8.4.5).

Probit, P(hemp_0709=1) = f(BX)

	GAM probit	**OLS probit**	**OLS probit**
Intercept	-7.80^{***}	-6.08^{**}	-8.53^{***}
	(2.06)	(2.05)	(1.94)
spring rainfall	-0.03^{**}	-0.03^{**}	-0.03^{***}
	(0.01)	(0.01)	(0.01)
spring temperature	0.40^{***}	0.38^{**}	0.46^{***}
	(0.12)	(0.12)	(0.11)
clay content	0.00^{*}	0.00^{*}	0.00^{**}
	(0.00)	(0.00)	(0.00)
silt content	0.00^{*}	0.00^{*}	0.00^{***}
	(0.00)	(0.00)	(0.00)
distance to closest hemp plant	3.04^{*}	-0.01^{***}	
	(3.51)	(0.00)	
Phi	0.18	0.18	0.14
Num. obs.	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

Table 7. Estimation results for hemp

3.7. Lentils

Explaining the occurrence of lentils crops is even more difficult than hemp: the Phi correlation coefficient is lower when the proximity to vegetable processing plants is included (**Table 8**). The proximity to vegetable processing and preserving plants has a small and counter-intuitive impact on lentils when it is the only market proximity variable in the model. To test the robustness of this finding, the proximity to Cocebi silos is added to the model. Cocebi is the only local collector specialized in organic products. Legumes are necessary in organic crop rotations to replenish soil nitrogen, and Cocebi specifically promotes lentils for this purpose. The addition of the proximity to Cocebi does not change the sign of the effect of vegetable processing plants, but it renders it insignificant. This may mean the proximity to lentils processing plants is not paramount, but even more likely, that the INSEE category – *vegetable processing and preserving* – is not specific enough to capture a market proximity effect.

Probit, P(lent_0709=1) = f(BX)

	GAM probit	GAM probit w/o Cocebi	OLS probit
Intercept	-1.19***	-1.02***	-1.11***
	(0.12)	(0.11)	(0.10)
rugosity	-0.09***	-0.09***	-0.07***
	(0.02)	(0.02)	(0.02)
proportion of municipality occasionally under water	-1.58**	-2.10***	-2.17***
	(0.56)	(0.56)	(0.54)
distance to closest vegetable processing factory	1.62	1.83**	
	(1.86)	(1.97)	
distance to closest Cocebi silo	1.91***		
	(1.99)		
Phi	0.16	0.12	0.13
Num. obs.	2017	2017	2017

***p < 0.001, **p < 0.01, *p < 0.05. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

Table 8. Estimation results for lentils

4. Discussion

4.1. Transportation costs are still relevant in agriculture

Despite the fact that it is commonly overlooked in empirical agricultural economics, we have shown that market proximity is one of the key drivers of crop location for many crops, be they common such as barley or rapeseed or rare such as hemp. In the cases of barley, the partial r^2 of market proximity variables is close to the partial r^2 of pedo-climatic drivers, which constitute here something close to the Ricardian natural advantages. The inclusion of market proximity variables also allows to significantly constrain suitable crop locations and reduce the spatial autocorrelation of residuals for hemp, and to a lesser extent for barley.

This result has an important political implication when regulators push for crop diversification or for the re-introduction of "virtuous" crops such as legumes. Whereas public policies cannot influence soil type and climate, they may have some leverage on factory location. Supporting the establishment of a feed factory will likely increase the amount of proteaginous pea within 50 minutes of this factory. To the contrary, inciting the production of legumes within 80 minutes of a malt factory will meet stronger obstacles than further away: the higher share of barley within this perimeter likely reflects a higher farm-gate price for barley. Unfortunately, due to the simultaneity bias in our models, we cannot specify exactly by how much the probability of finding the crop of interest will be increased (see section 4.4).

This is not to say that market proximity is paramount for all crops though. We fail to identify any significant relationship between possible markets for lentils or soy and their presence in a municipality. In these two cases however, our econometric models are limited by lack of specificity in the data: soy only represents 50% of the crop category studied and the markets for lentils could not be singled out from the large *vegetable processing factory* category.

4.2. Applicability conditions for the method: multi-centrism and market specificity

4.2.1. Multi-centrism

The effect of market proximity is less significant where markets for a given crops are concentrated in a single area. This is the case for rapeseed since there are only one biodiesel and one oil cake plant within 50 km of Burgundy, the study area. In the case of soy where no market proximity effect is identified, the dependent variable itself is spatially concentrated. This could naturally be a true finding: market proximity may not be a key driver for these crops. Yet, the conjunction of these observations makes us think that multi-centrism may be necessary for the applicability of our method. Multi-centrism does not increase the variability in observation *stricto sensu*, but it reduces multicollinearity among distance variables and endogeneity due to spatial correlation with between distance variables and unobserved variables with non-random spatial distribution patterns.

4.2.2. Market specificity

When the influence of market proximity is characterized, the independent variable is almost always crop-specific: malt plants for barley, biodiesel plants for rapeseed, hemp-specific textile factories for hemp. To the contrary, when independent variables characterizing market proximity are not specific to the studied crop – oil cake plants for soy, vegetable processing plants for lentils – their effect is not robust or inexistent. As for multi-centrism, this could be a chance finding. Yet, these observations point to market specificity as another applicability condition for our method.

4.2.3. Crop abundance is not a necessity

The study of diversity crops generates two mains challenges. Firstly, the econometric specification and its interpretation is less straightforward due to the necessary switch to generalized linear models. Secondly, the availability of crop-specific market proximity variables is diminished: while malt plants are distinguished in a yearly public systematic survey, the location of biodiesel plants and hemp-specific factories had to be obtained from sectoral reports which are not regularly updated (CETIOM, 2008). In the case of proteaginous crops and lentils, our inability to identify crop-specific processing plants does not guarantee that such plants do not exist. Yet, these challenges can be overcome and insights on the influence of market proximity can be obtained at the regional scale, as demonstrated by the example of hemp.

Multi-centrism, market specificity and crop abundance are of course somewhat related: the probability to find data on multiple crop-specific factories is more likely for a common crop than for a diversity crop.

4.3. Comparison with existing studies

The comparison of our results to other studies is challenging. The influence of market proximity in agriculture has seldom been the object of quantitative assessment and the concentration of agricultural activities tends to be studies either at coarser spatial scales than the municipality or through larger activity groups than crop types (eg. arable crops vs livestock rearing).

All comparable studies focus on a single crop/factory combination – namely maize/ethanol plant, and they do not focus on how far away from the plant an effect can be identified. Miao (2013) sets the county as the limit of the collecting basin and finds an effect. Given the average size and shape of Iowan counties, this translates into 15-30 km. Closer to our study, Motamed et al. (2016) defines the collecting basin with an arbitrary radius of 100 km around ethanol plans. It also finds an effect of the proximity to ethanol plants on maize acreage and through a sensitivity analysis, one understands that this effect decreases if the radius is set higher than 100 km. This radius is comparable to our 50-150 minutes range for collecting basins (see 3.1) which translates into a range of 53-160 km as the crow flies.

Nevertheless, we note that at regional scale Chevassus-Lozza and Daniel (2006) finds that barley is one of the most concentrated crop types, together with potatoes and fresh vegetables. This finding is consistent with the importance of the proximity to large malt plants and their suppliers that we demonstrated in this study.

In relation to the weak relationship between animal food plants and some of their supplies – rapeseed and proteaginous pea here, Charrier et al. (2013) explains that while animal food plants optimize their supply chain, including transportation costs, they also tend to be close to their customers, namely livestock farmers. While this could in principle explain our results, it does not seem to be a satisfactory explanation in our particular case: the correlation between distance to animal food plants and livestock density is significant (0.33, p-value < 3e-16) but positive.

4.4. Endogeneity and spatial correlation

The major limit of our study is that market proximity variables – that is the location of firms in our case – may be endogenous (Combes et al., 2008; Krugman, 1991). Indeed, one can easily imagine a simultaneity bias as firms locate their factories close to their feedstock as much as farmers produce crops which are consumed by close-by factories. The few comparable studies to our in terms of spatial and crop type resolution use either time-lagged or instrumental variables to treat the problem. Garrett et al. (2013) uses supply chain variables that are four years older than the crop abundance variable. Miao (2013) also uses time-lagged variables. While the time lag may reduce correlation with residuals, nothing indicates that this reduction is substantial since it does not eliminate the simultaneity bias. Motamed et al. (2016) use the railroad network as an instrument for ethanol plant location. The arguments and the tests for this instrument are convincing but they find a much higher effect for the instrumented variable than for the endogenous variable, which is not compatible with the suspected positive reversed causation. In our case, we could not think of an appropriate instrument for plant location.

Most likely, the only proper way to control for this problem is to use the panel structure of crop abundance data. But this would be a study in itself, requiring the identification of exogenous price shocks in the time series. And the interpretation of results would be more delicate as the dependent variable moves from crop abundance to its derivative.

Another limit is the existence of a spatial correlation in residuals, as identified by the small but significant Moran's I for barley and rapeseed. However, the software application of a spatial error model to the GAM and selection model used in this study would not be straightforward. As for a spatial autoregression model, it would assume neighbouring effects for which we find little ground. In general, spatial models complicate the interpretation of effects (Gibbons and Overman, 2012; Triboulet and Pérès, 2015). And in the particular case of land use models, the introduction of a spatial lag often does not strongly changes the estimates (Chakir and Le Gallo, 2013; Garrett et al., 2013). For these two reasons, we decided not to introduce a spatial lag in the models used here.

5. Conclusion

Market proximity is found to have a significant influence on the location of four crops out of the six studied, namely barley, rapeseed, proteaginous pea and hemp. The GAM always generates a shape of the proximity/crop occurrence relationship which is consistent with a collecting basin. In most cases, crop occurrence plateaus within 50-150 minutes of the relevant factory and then decreases sharply. Market proximity increases the share of barley in arable crops by more than 10 percentage points and its influence on the share of barley is as important as pedo-climatic conditions. The influence of malt plants is highest within 80 minutes from plant location. For hemp, the inclusion of market proximity variables in the model increases the phi correlation coefficient from 0.14 to 0.18 and strongly constrains the areas predicted as suitable for hemp. In a nutshell, we conclude that in many cases, the influence of market proximity on crop location cannot be neglected. Notwithstanding the decrease in transportation costs since 1950 and the associated globalization, the transportation costs put forward in von Thünen's framework remain relevant and complementary to Ricardo's natural advantages in explaining crop location.

From a methodological standpoint, we also derive from these test cases two likely feasibility conditions for the quantification of market proximity influence on crop location: markets must be crop-specific (eg. hemp factories, which process only hemp) and they must be present in two different locations to limit the multicollinearity and endogeneity problems associated with distance variables. Low crop abundance however does not seem to hamper the implementation of our method.

Yet, as in the three comparable studies we identified in the literature, endogeneity of market proximity variables likely biases these results. The use of the panel structure of crop abundance data is probably the most promising avenue to overcome this limit. Another interesting follow-up to this study would be to broaden the spatial extent of the analysis. Extension at national scale can seriously be considered for the crops meeting the feasibility conditions identified above. Above that, for example at European scale, data collection on market proximity variables is a daunting challenge.

6. Acknowledgements

The authors thank Pierre Cantelaube, Christine Le Bas and Marjorie Uberstosi for their help on accessing to and interpreting correctly the raw data. We are also grateful to Martin Lechenet and Nicolas Munier-Jollain for their expert judgements on the technical itineraries of crops in Burgundy. Finally, we thank Cécile Détang-Dessendre for her useful comments on the manuscripts.

This work benefited from the support of the Prosys PSDR project financed by INRA and region Bourgogne.

7. References

Agreste, 2010. Recensement général agricole (RGA) [WWW Document]. URL https://stats.agriculture.gouv.fr/cartostat/#v=map2;i=stru1.saumoy10;l=fr

Alexander, P., Moran, D., Smith, P., Hastings, A., Wang, S., Sünnenberg, G., Lovett, A., Tallis, M.J., Casella, E., Taylor, G., Finch, J., Cisowska, I., 2014. Estimating UK perennial energy crop supply using farm-scale models with spatially disaggregated data. GCB Bioenergy 6, 142–155. https://doi.org/10.1111/gcbb.12121

Allaire, G., Cahuzac, E., Maigné, E., Poméon, T., 2015. Localisation de l'agriculture biologique et accès aux marches. Revue d'Études en Agriculture et Environnement 96, 277–312. https://doi.org/10.4074/S1966960715012035

Aouadi, N., Aubertot, J.N., Caneill, J., Munier-Jolain, N., 2015. Analyzing the impact of the farming context and environmental factors on cropping systems: A regional case study in Burgundy. European Journal of Agronomy 66, 21–29. https://doi.org/10.1016/j.eja.2015.02.006

Ben Arfa, N., Rodriguez, C., Daniel, K., 2009. Dynamiques spatiales de la production agricole en France. Revue d'Économie Régionale & Urbaine novembre, 807. https://doi.org/10.3917/reru.094.0807

Bleakley, H., Lin, J., 2010. Portage: Path Dependence and Increasing Returns in U.S. History (NBER Working Paper No. 16314). National Bureau of Economic Research, Inc.

Britz, W., Verburg, P.H., Leip, A., 2011. Modelling of land cover and agricultural change in Europe: Combining the CLUE and CAPRI-Spat approaches. Agriculture, Ecosystems & Environment, Scaling methods in integrated assessment of agricultural systems 142, 40–50. https://doi.org/10.1016/j.agee.2010.03.008

Carpentier, A., Letort, E., 2014. Multicrop Production Models with Multinomial Logit Acreage Shares. Environ Resource Econ 59, 537–559. https://doi.org/10.1007/s10640-013-9748-6

CETIOM, 2008. Usines et filières courtes. Centre technique interprofessionnel des oléagineux et du chanvre.

CGDD, 2012. Diversification des cultures dans l'agriculture française - Etat des lieux et dispositifs d'accompagnement (No. 67), Etudes & documents. Commissariat général au développement durable.

Chakir, R., 2009. Spatial Downscaling of Agricultural Land-Use Data: An Econometric Approach Using Cross Entropy. Land Economics 85, 238–251.

Chakir, R., Le Gallo, J., 2013. Predicting land use allocation in France: A spatial panel data analysis. Ecological Economics 92, 114–125. https://doi.org/10.1016/j.ecolecon.2012.04.009

Charrier, F., Magrini, M.-B., Charlier, A., Fares, M. 'hand, Le Bail, M., Messéan, A., Meynard, J.-M., 2013. Alimentation animale et organisation des filières : une comparaison pois protéagineux-lin oléagineux pour comprendre les facteurs freinant ou favorisant les cultures de diversification. OCL 20, D407. https://doi.org/10.1051/ocl/2013011

Chatellier, V., Gaigné, C., 2012. Les logiques économiques de la spécialisation productive du territoire agricole français. Innovations Agronomiques 22, 185–203.

Chevassus-Lozza, E., Daniel, K., 2006. Market openness and geographical concentration of agricultural and agro-food activities: the challenges for French regions. Canadian Journal of Regional Science 29, 21–42.

Combes, P.-P., Mayer, T., Thisse, J.F., 2008. Economic geography: the integration of regions and nations. Princeton University Press, Princeton ; Oxford.

Daniel, K., 2005. Éléments sur la géographie de l'agriculture aux États-Unis et dans l'Union Européenne : les productions agricoles se concentrent-elles ? Revue d'Économie Régionale & Urbaine octobre, 533. https://doi.org/10.3917/reru.054.0533

Das, M., Newey, W.K., Vella, F., 2003. Nonparametric Estimation of Sample Selection Models. The Review of Economic Studies 70, 33–58.

Davis, A.S., Hill, J.D., Chase, C.A., Johanns, A.M., Liebman, M., 2012. Increasing Cropping System Diversity Balances Productivity, Profitability and Environmental Health. PLoS ONE 7, e47149. https://doi.org/10.1371/journal.pone.0047149

Donesol, 2006. Base de données Donesol © INRA, Unité INFOSOL, Orléans, Agrosup Dijon, 1998, 2006, 2005.

Doré, T., Makowski, D., Malézieux, E., Munier-Jolain, N., Tchamitchian, M., Tittonell, P., 2011. Facing up to the paradigm of ecological intensification in agronomy: Revisiting methods, concepts and knowledge. European Journal of Agronomy 34, 197–210. https://doi.org/10.1016/j.eja.2011.02.006

Fleurat-Lessard, F., 2013. Les freins à la diversification des productions d'oléo-protéagineux et de céréales au niveau des entreprises de collecte, stockage et mise en marché. OCL 20, D406. https://doi.org/10.1051/ocl/2013010

Gaigné, C., Le Gallo, J., Larue, S., Schmitt, B., 2012. Does Regulation of Manure Land Application Work Against Agglomeration Economies? Theory and Evidence from the French Hog Sector. American Journal of Agricultural Economics 94, 116–132.

Galko, E., 2007. Modélisation de l'offre agricole européenne face à de nouveaux enjeux : réformes politiques, effet de serre et changement climatique. AgroParisTech, Paris, France.

Garrett, R.D., Lambin, E.F., Naylor, R.L., 2013. Land institutions and supply chain configurations as determinants of soybean planted area and yields in Brazil. Land Use Policy 31, 385–396. https://doi.org/10.1016/j.landusepol.2012.08.002

Geniaux, G., Lambert, M., Bellon, S., 2009. Analyse de la diffusion spatiale de l'agriculture biologique en région Provence- Alpes-Côte d'Azur (Paca) : construction d'une méthodologie d'observation et de prospective. Innovations Agronomiques 4, 417–426.

Gibbons, S., Overman, H.G., 2012. Mostly Pointless Spatial Econometrics? Journal of Regional Science 52, 172–191. https://doi.org/10.1111/j.1467-9787.2012.00760.x

Hendricks, N.P., Sinnathamby, S., Douglas-Mankin, K., Smith, A., Sumner, D.A., Earnhart, D.H., 2014. The environmental effects of crop price increases: Nitrogen losses in the U.S. Corn Belt. Journal of Environmental Economics and Management 68, 507–526. https://doi.org/10.1016/j.jeem.2014.09.002

Hilal, M., 2013. Odomatrix. INRA UMR1041 CESAER DIJON.

Hlavac, M., 2015. stargazer package for R.

Holmes, T.J., Lee, S., 2012. Economies of Density versus Natural Advantage: Crop Choice on the Back Forty. Review of Economics and Statistics 94, 1–19. https://doi.org/10.1162/REST_a_00149

IGN, 2010. BD Alti - Altitude en France à une résolution de 25 mètres.

INAO, 2015. Source INAO, traitements ODR. INRA.

INSEE, 2013. Système informatisé du répertoire national des entreprises et des établissements / SIRENE [WWW Document]. URL http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=fd-sidenomb2012&page=fichiers_detail/sidenomb2012/telechargement.htm

Kempen, M., Elbersen, B.S., Staritsky, I., Andersen, E., Heckelei, T., 2011. Spatial allocation of farming systems and farming indicators in Europe. Agriculture, Ecosystems & Environment, Scaling methods in integrated assessment of agricultural systems 142, 51–62. https://doi.org/10.1016/j.agee.2010.08.001

Kilkenny, M., 1998. Transport Costs, the New Economic Geography, and Rural Development. Growth and Change 29, 259–280. https://doi.org/10.1111/0017-4815.00087

Kocoloski, M., Michael Griffin, W., Scott Matthews, H., 2011. Impacts of facility size and location decisions on ethanol production cost. Energy Policy 39, 47–56. https://doi.org/10.1016/j.enpol.2010.09.003
Krugman, P., 1991. Increasing Returns and Economic Geography. Journal of Political Economy 99, 483–99.
Lechenet, M., Bretagnolle, V., Bockstaller, C., Boissinot, F., Petit, M.-S., Petit, S., Munier-Jolain, N.M., 2014. Reconciling Pesticide Reduction with Economic and Environmental Sustainability in Arable Farming. PLoS ONE 9, e97922. https://doi.org/10.1371/journal.pone.0097922
Lichtenberg, E., 1989. Land Quality, Irrigation Development, and Cropping Patterns in the Northern High Plains. American Journal of Agricultural Economics 71, 187. https://doi.org/10.2307/1241787
Magrini, M.-B., Anton, M., Cholez, C., Corre-Hellou, G., Duc, G., Jeuffroy, M.-H., Meynard, J.-M., Pelzer, E., Voisin, A.-S., Walrand, S., 2016. Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. Ecological Economics 126, 152–162. https://doi.org/10.1016/j.ecolecon.2016.03.024
Magrini, M.-B., Fares, M. 'hand, Filippi, M., 2011. La signalisation de la qualité chez les petites coopératives agricoles françaises. Revue d'Économie Régionale & Urbaine octobre, 705. https://doi.org/10.3917/reru.114.0705
Malézieux, E., 2012. Designing cropping systems from nature. Agron. Sustain. Dev. 32, 15–29. https://doi.org/10.1007/s13593-011-0027-z
MEEM, 2014. Base des installations classées [WWW Document]. URL http://www.installationsclassees.developpement-durable.gouv.fr/rechercheICForm.php
Météo-France, 2013. Source Météo-France, traitements ODR, version 12/2010. INRA, Toulouse, France.
Meynard, J.-M., Jeuffroy, M.-H., Le Bail, M., Lefevre, A., Magrini, M.-B., Michon, C., 2016. Designing coupled innovations for the sustainability transition of agrifood systems. Agricultural Systems. https://doi.org/10.1016/j.agsy.2016.08.002
Miao, R., 2013. Impact of Ethanol Plants on Local Land Use Change. Agricultural and Resource Economics Review 42, 291–309.
Motamed, M., McPhail, L., Williams, R., 2016. Corn Area Response to Local Ethanol Markets in the United States: A Grid Cell Level Analysis. American Journal of Agricultural Economics 98, 726–743. https://doi.org/10.1093/ajae/aav095
Paradis, E., 2015. ape package for R.
RPG, 2013. Réseau parcellaire graphique 2007-2012, source Agence de services et de paiement, traitements Observatoire du développement rural. INRA, Toulouse, France.
SAA, 2015. Ministère de l'agriculture, de l'agroalimentaire et de la forêt - Agreste - Statistique agricole annuelle 2000-2014 [WWW Document]. URL https://stats.agriculture.gouv.fr/disar/
Schmidtner, E., Lippert, C., Engler, B., Häring, A.M., Aurbacher, J., Dabbert, S., 2011. Spatial distribution of organic farming in Germany: does neighbourhood matter? Eur Rev Agric Econ jbr047. https://doi.org/10.1093/erae/jbr047
Smithson, M., Verkuilen, J., 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods 11, 54–71. https://doi.org/10.1037/1082-989X.11.1.54
Triboulet, P., Pérès, S., 2015. La répartition spatiale des industries agroalimentaires dans le secteur coopératif français. Économie rurale 49–69. https://doi.org/10.4000/economierurale.4619
Triboulet, P., Peres, S., Filippi, M., Chantelot, S., 2013. Empreinte spatiale de la coopération agricole française. Un éclairage par l'analyse des structures de localisation. Revue d'Économie Régionale & Urbaine avril, 363. https://doi.org/10.3917/reru.132.0363

Ubertosi, M., Vautier, A., Curmi, P., Lamy, C., Toussaint, H., Castel, T., Richard, Y., Philippe, A.-S., 2009. Évaluation de la réserve en eau à l'échelle des pédopaysages de Bourgogne, in: Teneur En Eau et Transferts En Milieux Poreux : Mesures et Statistiques À L'échelle Stationnelle. Aix-en-Provence, France, pp. 1–10.

Wood, S., 2015. mgcv package for R.

Wooldridge, J.M., 2013. Introductory econometrics: a modern approach, 5th ed. ed. South-Western Cengage Learning, Mason, OH.

Wu, J., Segerson, K., 1995. The Impact of Policies and Land Characteristics on Potential Groundwater Pollution in Wisconsin. American Journal of Agricultural Economics 77, 1033. https://doi.org/10.2307/1243826

8. Supplementary materials

8.1. Distribution of dependent variables

8.1.1. Maps

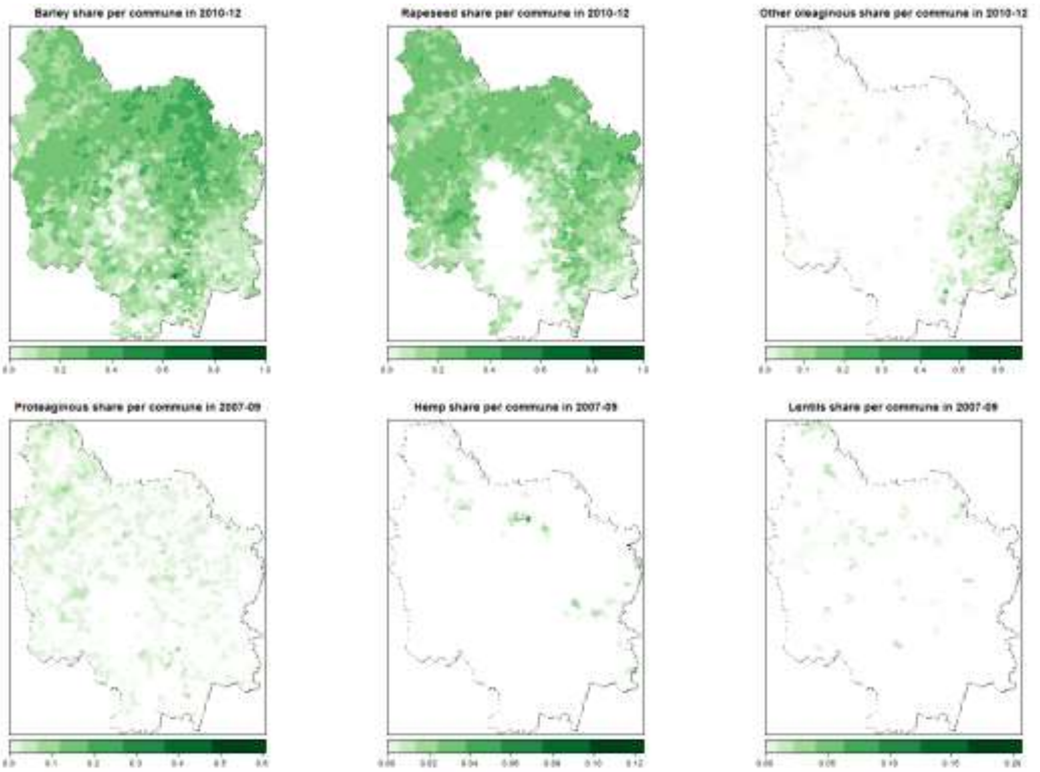
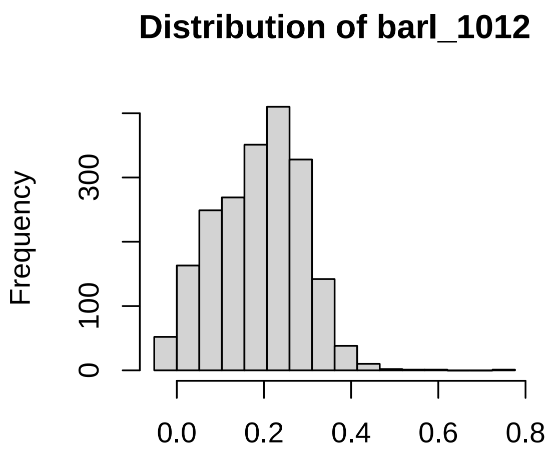
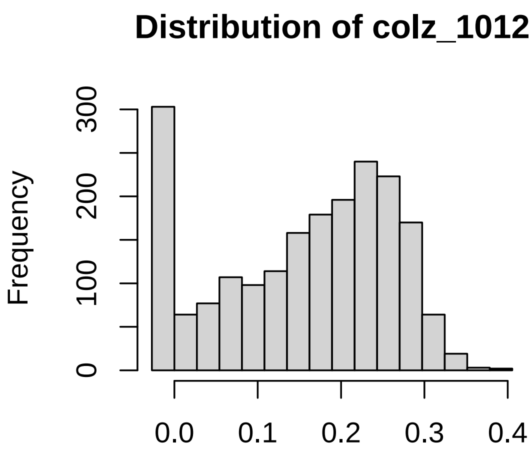


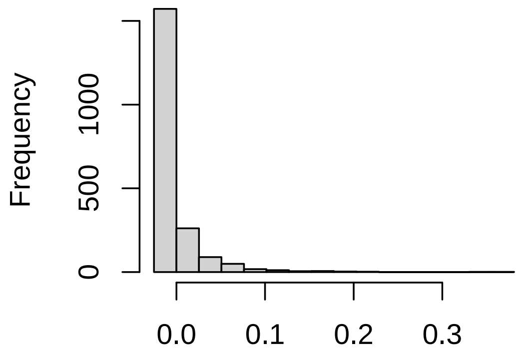
Figure 3. Maps of crops of interest. Source: RPG (2013).

8.1.2. Histograms

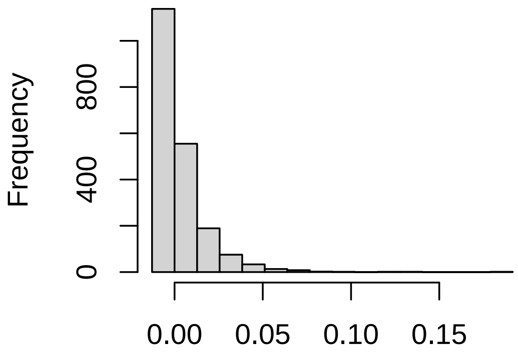




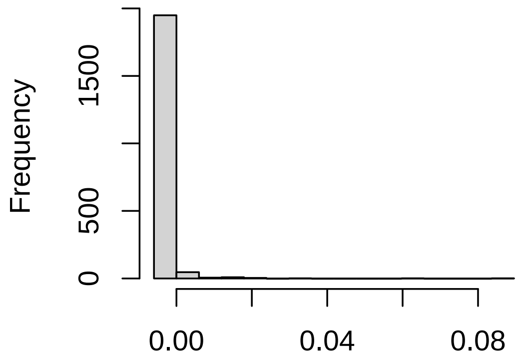
Distribution of soy_1012



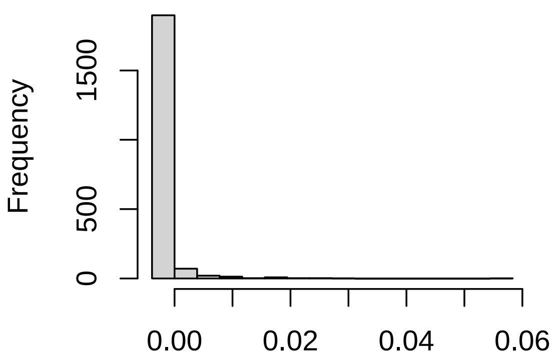
Distribution of prot_0709



Distribution of hemp_0709



Distribution of lent_0709

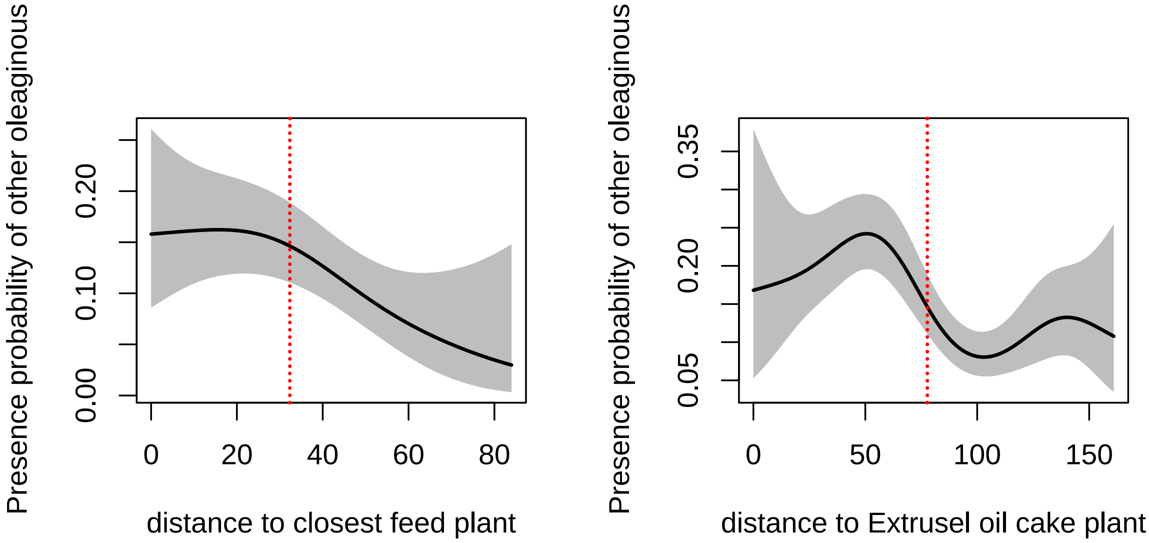


Centre d'Economie et de Sociologie
appliquées à l'Agriculture et aux Espaces Ruraux

8.2. Partial effects

This supplementary material displays the significant partial effects at the average for models which perform too poorly to be displayed in the main text.

8.2.1. Other oleaginous crops



8.2.2. Lentils

No significant market proximity effect.

8.3. Regression tests

8.3.1. Barley

OLS, barl_1012/(1-barl_1012) = f(BX)

	ols2	ols3	ols4	ols5	ols6	ols7	ols8	ols1
Intercept	-7.98 (3.08)**	-4.61 (3.07)	-8.84 (3.02)**	-8.19 (3.19)*	-1.43 (2.84)	2.71 (2.80)	-1.78 (0.03)***	-7.88 (3.24)*
rugosity	-0.02 (0.01)	-0.07 (0.01)***	-0.04 (0.01)**	-0.02 (0.02)	-0.01 (0.01)	-0.07 (0.01)***		-0.02 (0.02)
soil pH	0.45 (0.04)***	0.56 (0.04)***	0.43 (0.04)***	0.31 (0.08)***	0.47 (0.04)***	0.60 (0.04)***		0.32 (0.08)***
proportion of municipality occasionally under water	-1.09 (0.21)***	-1.20 (0.21)***	-1.16 (0.21)***	-1.06 (0.23)***	-1.13 (0.21)***	-1.30 (0.21)***		-1.08 (0.23)***
spring rainfall	-0.01 (0.00)**	-0.02 (0.00)***	-0.03 (0.01)***	-0.02 (0.01)**	-0.02 (0.00)***	-0.03 (0.00)***		-0.01 (0.01)*
spring relative humidity	0.04 (0.03)	0.08 (0.03)**	0.08 (0.03)**	0.06 (0.03)	-0.03 (0.03)	0.01 (0.03)		0.05 (0.04)
spring temperature	-0.03 (0.08)	-0.38 (0.07)***	-0.06 (0.08)	-0.03 (0.08)	-0.12 (0.08)	-0.47 (0.07)***		-0.01 (0.08)
number of freezing days	0.04 (0.01)***	0.01 (0.01)*	0.04 (0.01)***	0.03 (0.01)***	0.04 (0.01)***	0.02 (0.01)*		0.03 (0.01)***
available water content	-0.00 (0.00)	-0.00 (0.00)***	-0.00 (0.00)***	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)***		-0.00 (0.00)
distance to closest Soufflet silo	4.17 (5.23)***			3.90 (4.91)***	5.72 (6.94)***		7.39 (8.37)***	3.67 (4.65)***
distance to closest malt plant	3.07 (3.85)***	4.18 (5.17)***	1.00 (1.00)	3.38 (4.25)***			7.66 (8.52)***	3.22 (4.06)***
Capserval			5.86 (7.03)***					
share of municipal UUA which is irrigated				-0.01 (0.01)				-0.01 (0.01)
clay content				0.00 (0.00)				0.00 (0.00)
silt content				0.00 (0.00)				0.00 (0.00)
soil organic matter				-0.00 (0.00)				-0.00 (0.00)
soil carbonate				0.23 (0.18)				0.21 (0.18)

soil pierrosity				-0.18 (0.55)				-0.08 (0.55)
IGP beef								0.06 (0.13)
IGP dairy								0.01 (0.06)
IGP poultry								-0.14 (0.15)
number of granivore livestock unit per hectare								-0.22 (0.18)
number of herbivore livestock unit per hectare								0.01 (0.07)
AIC	6064.44	6177.79	6038.91	6068.01	6095.53		6276.02	6075.05
R2	0.34	0.31	0.35	0.35	0.33	0.29	0.27	0.34
Num. obs.	2017	2017	2017	2017	2017	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables.

8.3.2. Rapeseed

Non-random selection, P(rape_1012=1) = f(BX) & share_trans(rape_1012) = g(B'X')

	non_rand_sel1a	**non_rand_sel1b**	**non_rand_sel2a**	**non_rand_sel2b**	**non_rand_sel3a**	**non_rand_sel3b**	**ols**
Intercept	-18.34 (6.81)**	-0.73 (2.07)	-8.78 (4.83)	-0.59 (1.51)	-12.22 (5.01)*	-0.56 (1.50)	-0.32 (1.48)
rugosity	-0.15 (0.04)***	-0.00 (0.01)	-0.13 (0.03)***	0.01 (0.01)	-0.13 (0.03)***	0.00 (0.01)	-0.02 (0.01)*
IGP beef	-0.30 (0.23)	0.01 (0.10)			-0.34 (0.23)	0.05 (0.11)	
IGP poultry	0.19 (0.33)	-0.61 (0.09)***			0.09 (0.31)	-0.63 (0.08)***	
IGP dairy	0.38 (0.16)*	-0.11 (0.04)**			0.36 (0.15)*	-0.11 (0.04)**	
number of granivore livestock unit per hectare	0.19 (0.39)	-0.25 (0.12)*			0.19 (0.39)	-0.28 (0.12)*	

number of herbivore livestock unit per hectare	-0.26 (0.12)*	-0.20 (0.04)***			-0.30 (0.12)*	-0.21 (0.04)***	
spring rainfall	-0.04 (0.01)**	-0.01 (0.00)	-0.04 (0.01)***	-0.03 (0.00)***	-0.04 (0.01)***	-0.01 (0.00)	-0.03 (0.00)***
spring relative humidity	0.23 (0.08)**	-0.01 (0.02)	0.12 (0.06)*	-0.00 (0.02)	0.19 (0.06)**	-0.01 (0.02)	-0.01 (0.02)
spring temperature	0.16 (0.18)	0.01 (0.05)					
number of freezing days	0.02 (0.03)	0.00 (0.01)					
available water content	-0.00 (0.00)	-0.00 (0.00)					
share of municipal UUA which is irrigated	-0.03 (0.01)**	-0.00 (0.00)	-0.03 (0.01)**	-0.00 (0.00)	-0.03 (0.01)**	-0.00 (0.00)	-0.00 (0.00)
clay content	-0.00 (0.00)	0.00 (0.00)					
silt content	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)**	0.00 (0.00)***	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)***
soil organic matter	0.01 (0.01)	-0.00 (0.00)					
soil pH	0.63 (0.21)**	-0.11 (0.05)*	0.68 (0.17)***	-0.03 (0.05)	0.53 (0.18)**	-0.10 (0.04)*	-0.00 (0.04)
proportion of municipality occasionally under water	0.06 (0.57)	-0.67 (0.14)***	-0.03 (0.52)	-0.66 (0.14)***	-0.02 (0.53)	-0.65 (0.13)***	-0.78 (0.14)***
soil carbonate	-0.28 (0.50)	0.42 (0.11)***	-0.72 (0.43)	0.52 (0.10)***	-0.54 (0.45)	0.52 (0.09)***	0.74 (0.09)***
soil pierrosity	-0.61 (1.42)	0.56 (0.36)					
distance to Le Mériot biodiesel plant	1.27 (1.47)***	1.00 (1.00)	1.00 (1.00)***	4.39 (4.83)***	1.00 (1.00)***	1.27 (1.50)	-0.00 (0.00)**
distance to closest feed plant	1.89 (2.39)***	1.00 (1.00)*	1.87 (2.36)***	1.00 (1.00)	1.82 (2.31)***	1.00 (1.00)*	0.00 (0.00)
distance to Extrusel oil cake plant	4.15 (4.68)***	4.08 (4.66)***	3.99 (4.57)***	3.98 (4.57)***	4.04 (4.61)***	4.17 (4.71)***	0.00 (0.00)**
EDF: s(invmills1)		1.00 (1.00)***					
EDF: s(invmills2)				3.04 (3.82)***			
EDF: s(invmills3)						1.00 (1.00)***	
R2	0.66	0.44	0.64	0.42	0.65	0.44	0.35
Num. obs.	2017	1714	2017	1714	2017	1714	1714

***p < 0.001, **p < 0.01, *p < 0.05. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables.

8.3.3. Other oleaginous crops

Probit, P(soy_1012=1) = f(BX)

	prob1	prob2	prob3	prob4	ols
Intercept	-8.74 (5.17)	-12.73 (1.83)***	-8.21 (5.46)	-4.30 (4.68)	-12.92 (1.93)***
rugosity	-0.24 (0.03)***	-0.24 (0.02)***	-0.20 (0.03)***	-0.23 (0.03)***	-0.25 (0.02)***
spring rainfall	0.01 (0.01)		-0.01 (0.01)	0.01 (0.01)	
spring relative humidity	-0.08 (0.05)		-0.05 (0.06)	-0.16 (0.04)***	
spring temperature	0.59 (0.12)***	0.68 (0.10)***	0.47 (0.13)***	0.60 (0.12)***	0.70 (0.10)***
number of freezing days	0.05 (0.01)***	0.04 (0.01)***	0.07 (0.01)***	0.07 (0.01)***	0.06 (0.01)***
available water content	0.00 (0.00)		0.00 (0.00)	0.01 (0.00)*	
share of municipal UUA which is irrigated	-0.01 (0.01)		-0.01 (0.01)	-0.00 (0.01)	
clay content	-0.00 (0.00)		-0.00 (0.00)	-0.00 (0.00)	
silt content	0.00 (0.00)		-0.00 (0.00)	-0.00 (0.00)	
soil organic matter	0.01 (0.00)***	0.01 (0.00)**	0.01 (0.00)***	0.01 (0.00)***	0.01 (0.00)**
soil pH	0.42 (0.12)***	0.16 (0.07)*	0.37 (0.13)**	0.47 (0.11)***	0.16 (0.07)*
proportion of municipality occasionally under water	-0.50 (0.32)		-0.40 (0.34)	-0.56 (0.31)	
soil carbonate	-0.49 (0.27)		-0.41 (0.27)	-0.56 (0.26)*	
soil pierrosity	1.26 (0.96)		0.91 (0.97)	1.28 (0.91)	
distance to closest feed plant	1.94 (2.43)	1.90 (2.38)*	1.65 (2.06)		-0.01 (0.00)*

distance to Extrusel oil cake plant	3.16 (3.83)	4.39 (4.83)***	3.32 (4.01)		-0.00 (0.00)**
IGP beef			0.05 (0.19)		
IGP dairy			-0.21 (0.09)*		
IGP poultry			1.33 (0.22)***		
number of herbivore livestock unit per hectare			-0.31 (0.11)**		
number of granivore livestock unit per hectare			-0.13 (0.31)		
Phi	0.66	0.65	0.65	0.65	0.65
Num. obs.	2017	2017	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

8.3.4. Proteaginous crops

P(prot_0709=1) = f(BX)

	gam2	**prob2**	**prob3**	**gam3**	**prob1**	**gam1**
Intercept	10.68 (1.08)***	11.14 (1.09)***	10.21 (1.03)***	9.02 (1.36)***	6.36 (3.57)	5.73 (3.83)
rugosity	-0.12 (0.01)***	-0.12 (0.01)***	-0.12 (0.01)***	-0.10 (0.02)***	-0.14 (0.02)***	-0.10 (0.02)***
spring rainfall	-0.03 (0.01)***	-0.03 (0.01)***	-0.04 (0.01)***	-0.02 (0.01)**	-0.02 (0.01)***	-0.02 (0.01)**
spring temperature	-0.49 (0.07)***	-0.50 (0.07)***	-0.44 (0.06)***	-0.43 (0.08)***	-0.38 (0.08)***	-0.38 (0.10)***
number of freezing days	-0.03 (0.01)**	-0.02 (0.01)**	-0.02 (0.01)*	-0.02 (0.01)	-0.02 (0.01)*	-0.01 (0.01)
proportion of municipality occasionally under water	-1.42 (0.27)***	-1.47 (0.27)***	-1.53 (0.27)***	-1.13 (0.27)***	-1.29 (0.28)***	-1.27 (0.29)***

distance to closest feed plant	3.25 (3.87)**	-0.01 (0.00)**		2.94 (3.58)		2.78 (3.41)
IGP beef				0.12 (0.13)		0.09 (0.14)
IGP dairy				-0.05 (0.07)		-0.02 (0.07)
IGP poultry				0.33 (0.18)		0.38 (0.18)*
otex6aucun				-0.92 (0.49)		-0.96 (0.49)
otex6ffm				-0.49 (0.13)***		-0.47 (0.13)***
otex6grani				-0.47 (0.23)*		-0.49 (0.23)*
otex6herbi				-0.50 (0.12)***		-0.53 (0.12)***
otex6poly				-0.08 (0.08)		-0.09 (0.08)
number of granivore livestock unit per hectare				0.24 (0.23)		0.29 (0.24)
number of herbivore livestock unit per hectare				-0.24 (0.09)**		-0.25 (0.09)**
soil organic matter				-0.01 (0.00)**	-0.01 (0.00)	-0.01 (0.00)*
soil carbonate				0.53 (0.12)***	0.36 (0.20)	0.47 (0.21)*
soil pierrosity				0.30 (0.58)	0.86 (0.66)	0.57 (0.69)
spring relative humidity					0.02 (0.03)	0.03 (0.04)
available water content					-0.00 (0.00)	0.00 (0.00)
share of municipal UUA which is irrigated					-0.00 (0.01)	-0.00 (0.01)
clay content					0.00 (0.00)	0.00 (0.00)
silt content					0.00 (0.00)	-0.00 (0.00)
soil pH					0.12 (0.09)	0.00 (0.10)
Phi	0.31	0.30	0.30	0.37	0.37	0.37
Num. obs.	2017	2017	2017	2017	2017	2017

***p < 0.001, **p < 0.01, *p < 0.05. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

8.3.5. Hemp

Probit, P(hemp_0709=1) = f(BX)

	gam3	**prob3**	**prob4**	**gam2**	**prob2**	**prob1**	**gam1**
Intercept	-7.80 (2.06)***	-6.08 (2.05)**	-8.53 (1.94)***	-7.88 (2.87)**	-5.87 (2.62)*	-14.53 (7.91)	-22.38 (10.57)*
spring rainfall	-0.03 (0.01)**	-0.03 (0.01)**	-0.03 (0.01)***	0.01 (0.01)	-0.03 (0.01)*	-0.03 (0.01)*	0.02 (0.02)
spring temperature	0.40 (0.12)***	0.38 (0.12)**	0.46 (0.11)***	0.20 (0.16)	0.37 (0.14)**	0.48 (0.20)*	0.38 (0.25)
clay content	0.00 (0.00)*	0.00 (0.00)*	0.00 (0.00)**	0.00 (0.00)*	0.00 (0.00)*	0.00 (0.00)*	0.00 (0.00)
silt content	0.00 (0.00)*	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)*
distance to closest hemp plant	3.04 (3.51)*	-0.01 (0.00)***		3.14 (3.54)*	-0.01 (0.00)*		3.23 (3.66)**
Interval				4.33 (4.66)***	-0.00 (0.00)		4.32 (4.67)**
rugosity						-0.07 (0.04)	-0.03 (0.05)
spring relative humidity						0.06 (0.07)	0.13 (0.11)
number of freezing days						0.01 (0.02)	0.04 (0.03)
available water content						-0.00 (0.00)	-0.00 (0.00)
share of municipal UUA which is irrigated						-0.01 (0.02)	-0.00 (0.02)
soil organic matter						0.00 (0.01)	0.00 (0.01)
soil pH						0.07 (0.16)	-0.07 (0.22)

proportion of municipality occasionally under water						0.25 (0.46)	0.11 (0.54)
soil carbonate						-0.58 (0.38)	-0.47 (0.48)
soil pierrosity						1.44 (1.32)	0.27 (1.72)
number of herbivore livestock unit per hectare							-0.12 (0.24)
number of granivore livestock unit per hectare							-0.19 (0.48)
IGP dairy							-0.34 (0.19)
IGP beef							-157.10 (3947580.24)
IGP poultry							0.12 (0.39)
IGP_other							0.00 (0.00)
Phi	0.18	0.18	0.14	0.25	0.18	0.19	0.26
Num. obs.	2017	2017	2017	2017	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

8.3.6. Lentils

Probit, P(lent_0709=1) = f(BX)

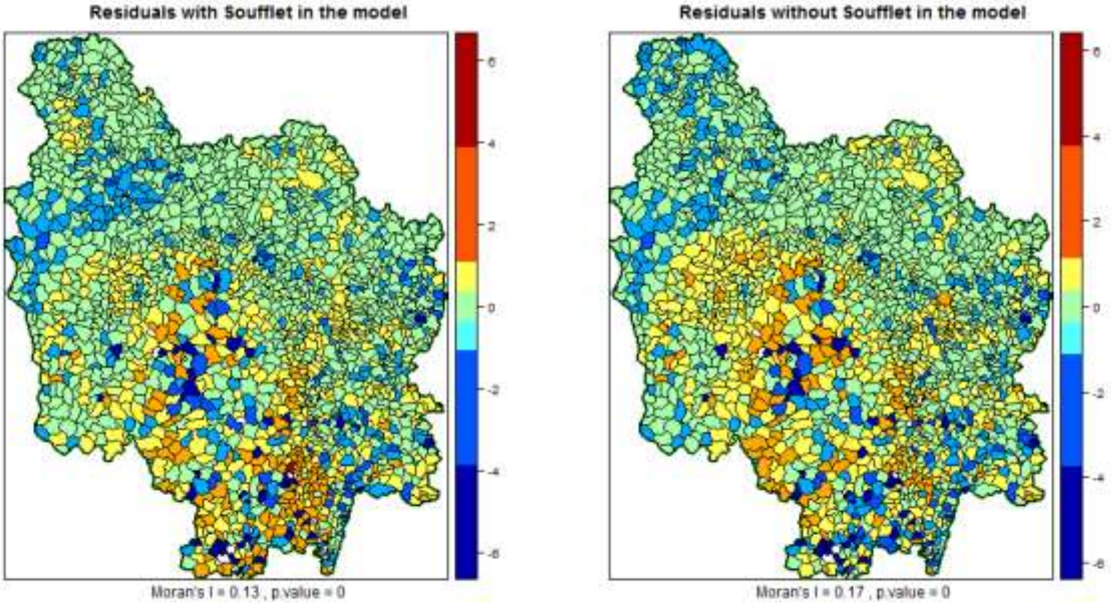
	gam3	gam2	prob4	prob2	gam4	gam5	prob3	prob1	gam1
Intercept	-1.19 (0.12)***	-1.02 (0.11)***	-1.11 (0.10)***	-1.28 (0.13)***	-0.83 (0.12)***	-3.08 (6.14)	-0.72 (0.18)***	-3.81 (6.07)	-7.73 (6.73)
rugosity	-0.09 (0.02)***	-0.09 (0.02)***	-0.07 (0.02)***	-0.08 (0.02)***	-0.07 (0.02)**	-0.11 (0.03)**	-0.08 (0.02)***	-0.11 (0.03)***	-0.08 (0.03)*
proportion of municipality occasionally under water	-1.58 (0.56)**	-2.10 (0.56)***	-2.17 (0.54)***	-2.18 (0.55)***	-1.74 (0.56)**	-1.41 (0.61)*	-1.75 (0.56)**	-1.36 (0.60)*	-1.20 (0.64)

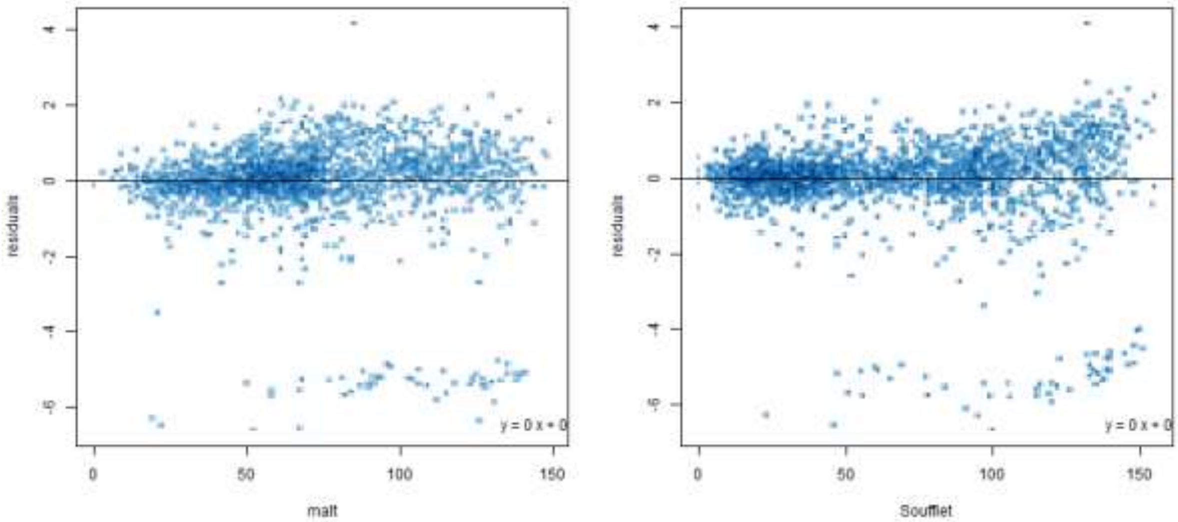
distance to closest vegetable processing factory	1.62 (1.86)	1.83 (1.97)**		0.00 (0.00)*	1.73 (1.93)***	1.00 (1.00)	0.00 (0.00)		1.00 (1.00)
distance to closest Cocebi silo	1.91 (1.99)***						-0.01 (0.00)***		1.70 (1.91)
number of herbivore livestock unit per hectare					-0.74 (0.14)***				-0.47 (0.18)**
spring rainfall						-0.01 (0.01)		-0.01 (0.01)	-0.01 (0.02)
spring relative humidity						0.05 (0.06)		0.07 (0.06)	0.09 (0.07)
spring temperature						-0.19 (0.14)		-0.21 (0.14)	-0.03 (0.16)
number of freezing days						0.00 (0.02)		-0.00 (0.02)	0.03 (0.02)
available water content						-0.00 (0.00)		-0.00 (0.00)	-0.00 (0.00)
share of municipal UUA which is irrigated						-0.00 (0.01)		-0.00 (0.01)	-0.01 (0.02)
clay content						0.00 (0.00)		0.00 (0.00)	-0.00 (0.00)
silt content						0.00 (0.00)		0.00 (0.00)	0.00 (0.00)
soil organic matter						-0.01 (0.00)		-0.01 (0.00)	-0.01 (0.01)
soil pH						0.20 (0.16)		0.12 (0.14)	0.09 (0.17)
soil carbonate						-0.03 (0.35)		0.11 (0.32)	-0.02 (0.36)
soil pierrosity						1.12 (1.01)		1.15 (1.02)	0.67 (1.05)
number of granivore livestock unit per hectare									-0.13 (0.42)
IGP dairy									0.10 (0.12)
IGP beef									-0.45 (0.42)
IGP poultry									0.18 (0.40)
IGP_other									0.00 (0.00)
Phi	0.16	0.12	0.13	0.13	0.19	0.21	0.15	0.20	0.23
Num. obs.	2017	2017	2017	2017	2017	2017	2017	2017	2017

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For splines, values refer to two estimates of the equivalent degree of freedom, not to the estimated coefficient and its standard error as for the parametric variables. The Phi correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw.

8.4. Residuals

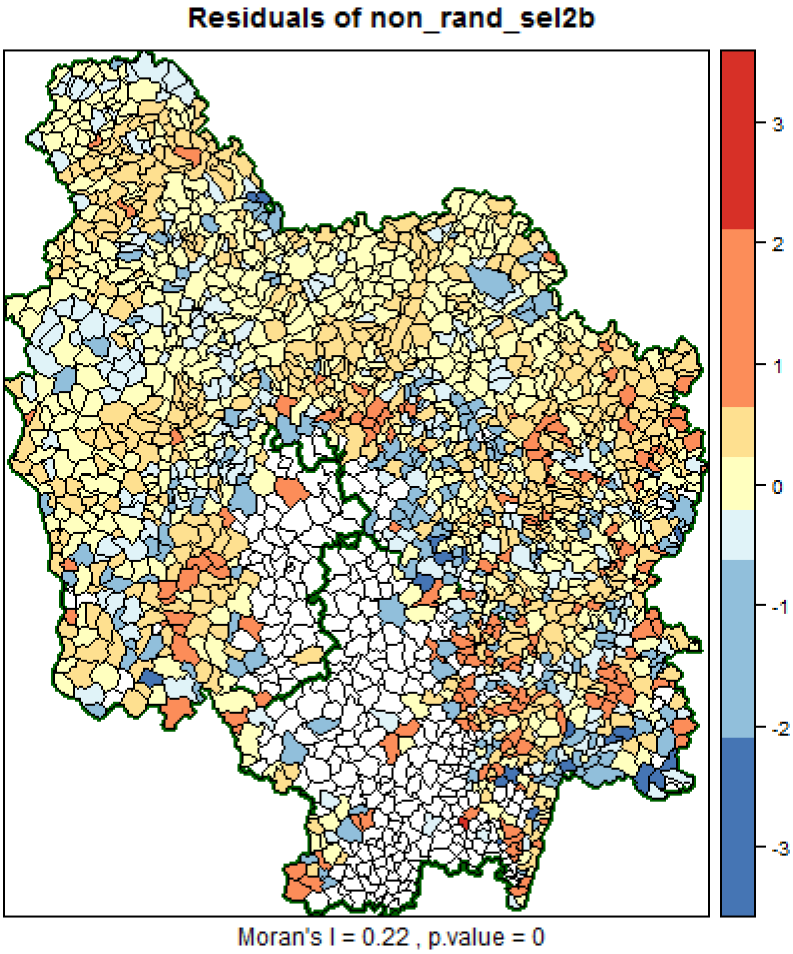
8.4.1. Barley



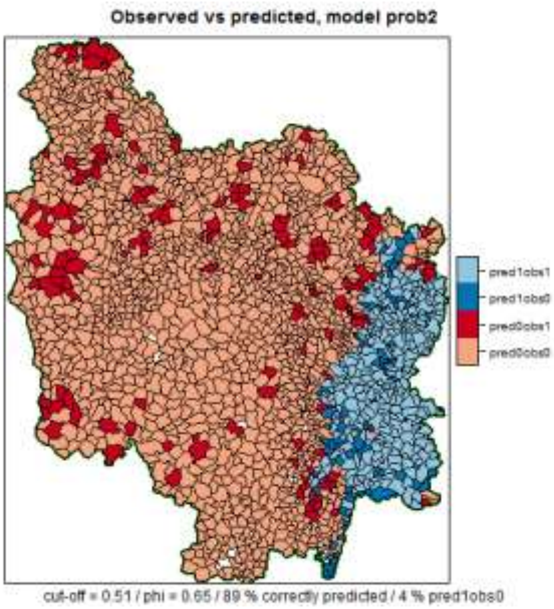


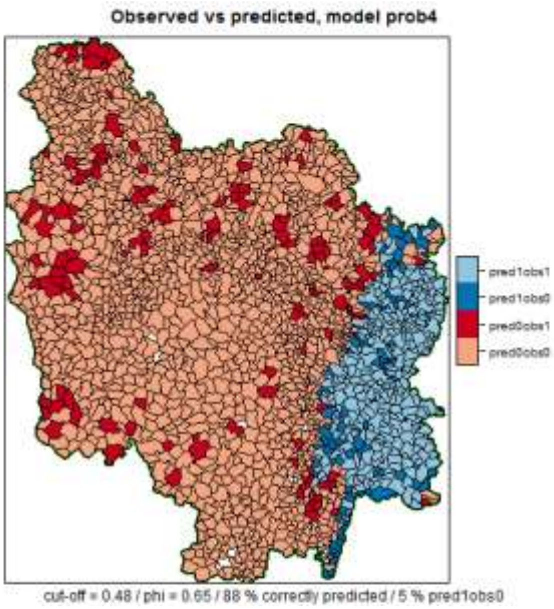
Note that there are large residuals, both positive and negative, in central and south-central Burgundy. These areas also have small arable crops area (often less than 100 ha per municipality) which likely creates more local heterogeneity (smaller denominator).

8.4.2. Rapeseed



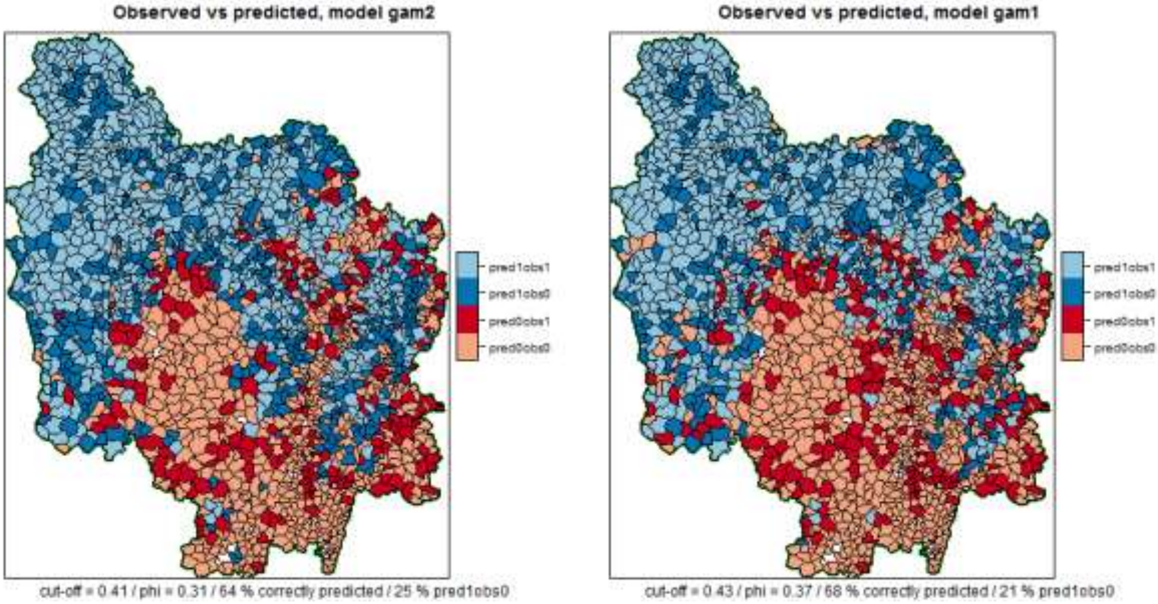
8.4.3. Other oleaginous crops





The *phi* correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw. *cut-off* is the probability cut-off which maximizes phi. Left pane: model with market proximity variable. Right pane: model with pedo-climatic drivers only.

8.4.4. Proteaginous crops



The *phi* correlation coefficient is a performance criteria which equals 1 for a perfect model and 0 for a random draw. *cut-off* is the probability cut-off which maximizes phi. Left pane: model without endogenous variables. Right pane: model with model with endogenous variables.

8.4.5. Hemp

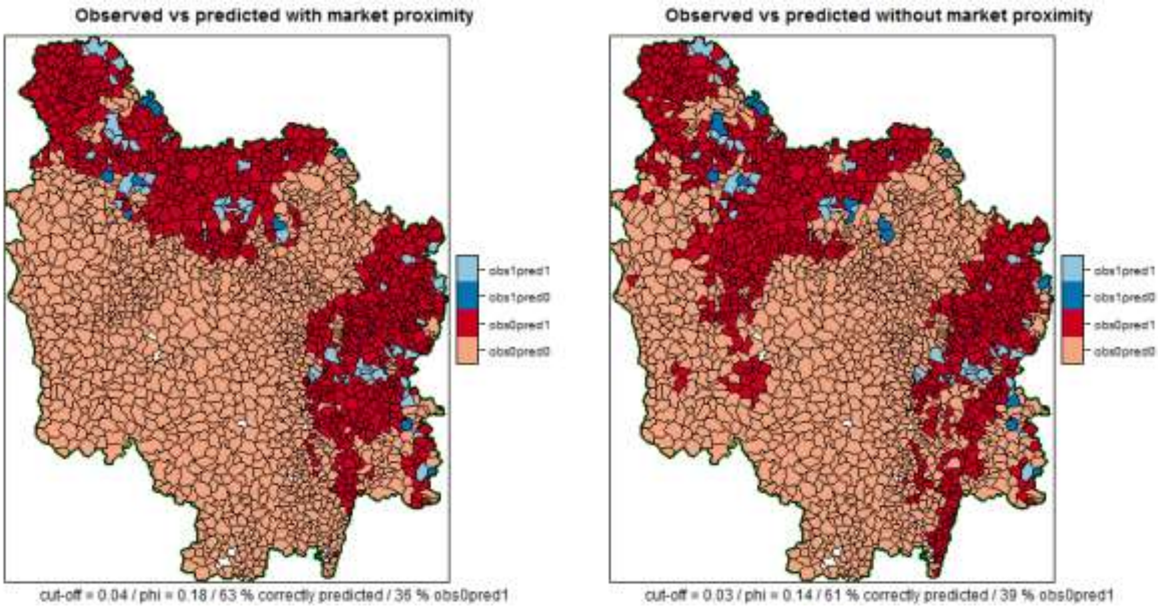


Figure 4. Confusion map for hemp probit models with and without market proximity variables

8.4.6. Lentils

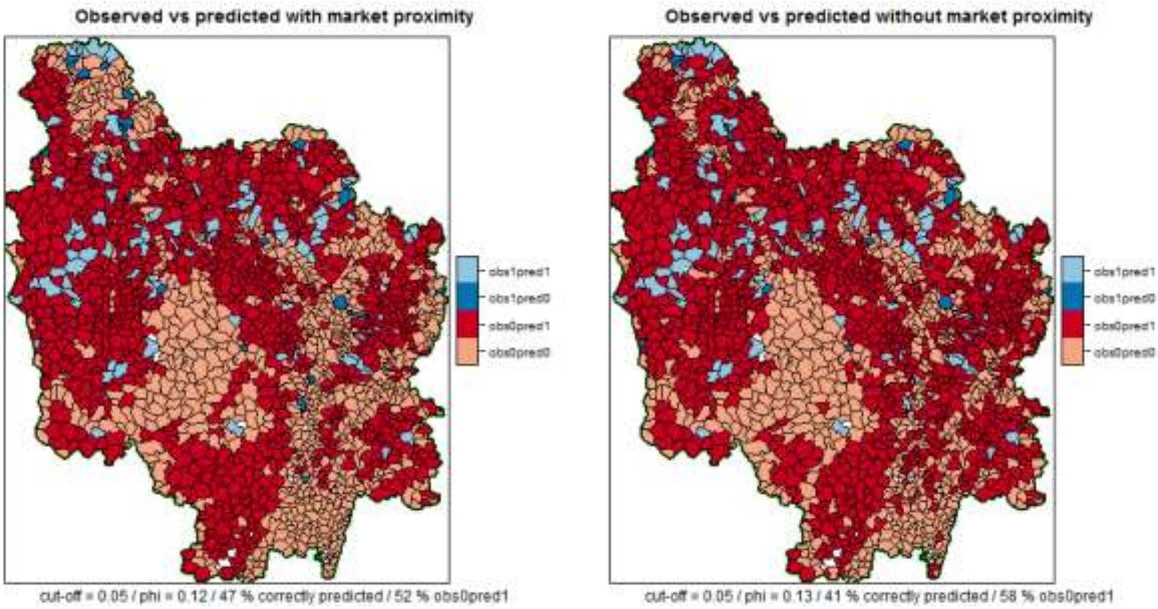


Figure 5. Confusion map for lentils probit models with and without market proximity variables

8.5. Summary statistics of independent variables

Descriptive statistics of independent variables

	Description	N	Min	Mean	Median	Max	% zeros
Pedo-climatic variables							
rugmean	Average ruggedness in municipality. Based on IGN_25m elevation data (BD alti), treated in QGIS to obtain ruggedness. Ruggedness according to Riley (1999) definition : sqrt(sum of square difference in elevation between center and the eight neighbouring pixels).	2,017	0.27	4.77	4.48	16.08	0
freeze_0111_0103	Average number of freezing days(min temperature below -4 °C ?) between 01/11 and 01/03 over 1991-2010 (crop years from November y-1 to October y). Source: Meteo France retreated by ODR.	2,017	31.05	44.13	44.43	71.76	0
temp_apr_jun	Average temperature between 01/04 and 30/06 over 1991-2010 (crop years from November y-1 to October y). Source: Meteo France retreated by ODR.	2,017	12.08	13.90	13.88	15.43	0
rain_apr_jun	Average liquid rainfall between 01/04 and 30/06 over 1991-2010 (crop years from November y-1 to October y). Source: Meteo France retreated by ODR.	2,017	51.83	71.63	71.67	115.20	0
humrel_apr_jun	Average relative humidity between 01/04 and 30/06 over 1991-2010 (crop years from November y-1 to October y). Source: Meteo France retreated by ODR.	2,017	70.30	74.60	75.01	77.96	0

argile_g_kg.1_30cm	Clay content in the first 30 cm. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	66.31	302.07	307.61	525.44	0
limon_g_kg.1_30cm	Silt content in the first 30 cm. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	73.20	449.54	474.90	659.81	0
mo_g_kg.1_30cm	Organic matter content in the first 30 cm. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	13.63	41.35	37.99	138.39	0
ph_eau_30cm	pH of soil diluted in water in the first 30cm. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	4.09	6.54	6.67	8.04	0
reg_submer	Proportion of municipality at least seasonally submerged by water. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	0	0.09	0.04	1	24.89
carbonate_eg1	Proportion of municipality with calceous coarse materials. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	0	0.37	0.28	1	17.85
pierro_surf	Proportion of coarse materials. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	0.02	0.12	0.11	0.36	0
awc2	Available water content between pF 1.5 and pF 4.2. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by M. Ubertosi.	2,017	18.70	89.10	89.29	172.81	0
awc	Available water content. Source: 1/250 000 soil map of Burgundy (Donesol) provided by AgroSup Dijon and INRA-Infosol and retreated by authors.	2,017	46.47	135.96	128.01	337.23	0
irrig	Share of municipality UAA which is irrigated. Source: RGA 2010 through Geoclip.	2,017	0	0.96	0	100	80.81

Market proximity variables							
IGP_dairy	Whether the municipality is classified within an IGP related to dairy production. 0 = no, 1 = yes. Data source: INAO, treated by ODR.	2,017	0	0.51	1	1	49.28
IGP_beef	Whether the municipality is classified within an IGP related to beef production. 0 = no, 1 = yes. Data source: INAO, treated by ODR.	2,017	0	0.14	0	1	85.67
IGP_poultry	Whether the municipality is classified within an IGP related to poultry production. 0 = no, 1 = yes. Data source: INAO, treated by ODR.	2,017	0	0.06	0	1	94.35
IGP_other	Whether the municipality is classified within an IGP related to other production. 0 = no, 1 = yes. Data source: INAO, treated by ODR.	2,017	0	0	0	0	100
otex6							
herbi_UGB_per_ha	Number of herbivorous (cattle, sheep, goats, horses, ...) UGB per hectare of utilized agricultural area in the municipality. Source: RGA 2010, filled when necessary (no data for statistical secret) sum at higher level data (canton, PRA, departement) corresponds to the actual higher level data.	2,017	0	0.57	0.44	3.46	12.59
grani_UGB_per_ha	Number of granivorous (swine, poultry) UGB per hectare of utilized agricultural area in the municipality. Source: RGA 2010, filled when necessary (no data for statistical secret) sum at higher level data (canton, PRA, departement) corresponds to the actual higher level data.	2,017	-0	0.07	0.03	2.29	14.43
Capserval	Travel time by road to the nearest large silo (> 5000 m3) of the corresponding 'major collector' (ie a collector with more than two large silos or more than 100 000 m3 of capacity). Source for silos location and capacity: ICPE database downloaded on 17/12/2014. Source for travel time: INRA, Odomatrix.	2,017	0	102.20	107	215	0.20

Interval	Travel time by road to the nearest large silo (> 5000 m3) of the corresponding 'major collector' (ie a collector with more than two large silos or more than 100 000 m3 of capacity). Source for silos location and capacity: ICPE database downloaded on 17/12/2014. Source for travel time: INRA, Odomatrix.	2,017	3	89.80	88	191	0
Soufflet	Travel time by road to the nearest large silo (> 5000 m3) of the corresponding 'major collector' (ie a collector with more than two large silos or more than 100 000 m3 of capacity). Source for silos location and capacity: ICPE database downloaded on 17/12/2014. Source for travel time: INRA, Odomatrix.	2,017	0	67.48	63	155	0.25
canned_leg	Travel time by road to the nearest canned legume factory with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	58.76	56	137	0.15
animal_food_farm	Travel time by road to the nearest farm animal food factory with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	32.33	31	84	0.55
malt	Travel time by road to the nearest malt plant with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	70.51	66	149	0.05
og71076	Travel time by road to this first relevant transformation factory (animal food - af - or canned legume - cl) with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	77.71	74	161	0.05
oi10231	Travel time by road to this first relevant transformation factory (animal food - af - or canned legume - cl) with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	21	144.14	146	262	0

cocebi	Travel time by road to this first relevant transformation factory (animal food - af - or canned legume - cl) with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	73.08	71	173	0.05
relevant_cl	Travel time by road to the nearest canned legume factory possibly using lentils with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	64.00	62	142	0.10
chanv	Travel time by road to the nearest hemp factory with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	18	104.27	98	215	0
animal_food	Travel time by road to the nearest animal food factory with at least one employee and not classified as artisan. Source for factory location: INSEE Etablissement database as of 01/01/2012. Source for travel time: INRA, Odomatrix.	2,017	0	30.55	30	84	0.69

Centre d'**E**conomie et de **S**ociologie
appliquées à l'**A**griculture et aux **E**spaces **R**uraux