FACTORS DETERMINING THE SPATIAL DISTRIBUTION OF GRAIN LEGUME CULTIVATION IN THE EU

Christine Oré Barrios, Eva Mäurer, Christian Lippert

christine.ore@uni-hohenheim.de

Institut für Landwirtschaftliche Betriebslehre, Universität Hohenheim,
Schwerzstraße 44, 70599 Stuttgart

2019

Vortrag anlässlich der 59. Jahrestagung der GEWISOLA
(Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.)
„Landwirtschaft und ländliche Räume im gesellschaftlichen Wandel“
Braunschweig, 25. bis 27. September 2019

Factors determining the spatial distribution of grain legume cultivation in the EU

Abstract

The spatial distribution of grain legumes can be explained by combining the analysis of traditional location factors with the concept of spatial dependence. We present an in-depth literature review and derive hypotheses about the conditions that make a certain location attractive for grain legume cultivation. These hypotheses are then tested via three different approaches: a fractional logit generalized linear regression model, a zero-one inflated beta regression model and spatial econometric models. We use secondary data for the EU-28 at the NUTS2 level. Location factors that contribute to the incidence of grain legume cultivation are the frequency of organic farming and irrigable area, a farther distance to the next main port, a prominent role of legumes in regional diets, and policy measures such as support coupled to protein crops and restrictions on the use of nitrogen fertilizers. Our results suggest that agglomeration effects also matter in the cultivation of dry pulses. Corresponding external economies of scale and positive spill-over effects could be further exploited by specifically supporting investments in regional supply chains, local extension services and training.

Keywords

grain legumes, fractional logit generalized linear model, zero-one inflated beta model, spatial econometrics

1. Introduction

Grain legumes are only niche crops in European agriculture despite their environmental and agronomic benefits and their wide use for food, feed, fuel and industrial purposes (EC-DG AGRI, 2011). Because of their capacity to fix air-borne nitrogen, legume crops contribute to soil fertility and increase the yield of subsequent crops in the rotation (PREISSEL ET AL., 2015). The reduced need for synthetic nitrogen fertilizers not only decreases overall farm costs, but also nitrate input into the groundwater and greenhouse gas emissions (BESTE and BOEDDINGHAUS, 2011). In addition to a “nitrogen effect”, the agronomic benefits of legumes include a “break crop effect” (CHALK, 1998 cited in PREISSEL ET AL., 2015) which refers to the benefits to soil organic matter and structure[1], the mobilization of otherwise unavailable nutrients through mechanisms such as dissolution by root exudates, and the reduced pest and disease pressure (SIDDIQUE ET AL., 2011). In 2016, grain legumes covered only 3.1% of EU-28 arable land (2.3% pulses, 0.8% soya) (EUROSTAT, 2016_a). On average over the period 1961-2011, 63% of the grain legume supply in the EU was imported (CERNAY ET AL., 2015 based on FAOSTAT, 2015). As a consequence, the high amount of imports of grain legumes for animal feed has made the EU livestock sector and hereof particularly the pig and poultry production (WATSON and STODDARD, 2017) vulnerable to price volatility and trade distortions on agricultural world markets (HÄUSLING, 2010). This protein deficit in the EU may also be discussed in terms of land use equality and sustainability goals. For instance, soya imports to the EU accounted for roughly 12% of the worldwide production of soya bean in 2013/14, and 15 million hectares of arable land outside the EU (WESTHOEK ET AL., 2011), the equivalent to 14% of the EU´s arable land (EUROSTAT, 2013_a). In some non-EU countries, the increased cultivation of grain legumes (mainly soya) has led to unsustainable farming on sensitive grassland and deforestation of rainforest, with negative effects such as soil erosion and the depletion of water resources and biodiversity (HÄUSLING, 2010). A lot of research has been

---

[1] Deep rooted legumes contribute to nutrient cycling and water infiltration into the soil

done to address these challenges in an attempt to rebalance the supply and demand of grain legumes in the EU (for examples, see EC-DG AGRI, 2014). Scientists have proposed several hypotheses to explain the current spatial location of grain legume production and its intensity, but to our knowledge have not tested them statistically so far. Moreover, no study has analyzed if agglomeration effects could play a role in the current distribution of grain legume production across the EU. To explain the production patterns of grain legumes we combine the econometric analysis of traditionally used location factors with the concept of spatial dependence. Grain legumes[2] include the aggregate dry pulses (from now on referred to as “pulses”) and soya. Soya is not only a protein crop, but also a predominant oilseed crop. We decided to analyze one pulse crop separate from the aggregate and chose field and broad beans (from now on referred to as “beans”) because beans are the second most common pulse crop after field peas in Europe and, unlike field peas, unevenly cultivated across the EU. The spatial distribution of pulses, beans and soya cultivation in the EU in 2016 is illustrated in Annex 1. In our analysis, we will use three different econometric approaches, a fractional logit generalized linear regression model, a zero-one inflated beta regression model and two spatial econometric models to account for spatial dependency and spatial heterogeneity.

2. Factors influencing legume production

Agricultural activity is usually influenced by regional climate and soil conditions. Hence, the partners of the EU project “GL-Pro”[3] suggested different degrees of suitability of common grain legumes for specific climate and soil characteristics, summarized in Table 1. The table shows that beans and soya have a low tolerance for drought. Soya is known to be very sensitive to water stress at three stages of growth: the beginning of bloom, the beginning of pod development and seed development (LFL, 2014). Soya is a tropical plant and therefore most tolerant to heat stress, whereas lower tolerance levels are found for the other common pulses, especially for field beans. We expect soya to be cultivated in regions with a relatively warm climate, probably on irrigated fields. Additionally, soya requires a high topsoil available water capacity (RP FREIBURG, 2017). Table 1 further indicates that all common grain legumes, except lupines, can grow very well on chalky soil, and that deep soils are important for beans and soya, but less important for other common pulses.

Table 1: Suitability of grain legumes for specific climate and soil characteristics

	Pea	Field Bean	Blue lupin	White lupin	Yellow lupin	Chick-pea	Common vetch	Lentil vetch	Soya
Tolerance for heat stress	+	-	+	+	+	++	n.s.	n.s.	+++
Tolerance for drought	+	-	++	+	++	++	++	++	-
Chalky soil (active CaCO3 > 2%)	++	++	--	--	--	++	++	++	++
Shallow soil	+	-	++	+	++	++	++	++	-

+++ very good, ++ good, + medium, - low, -- to be avoided, n.s. not specified

Source: Authors´ own illustration based on GL-PRO, 2005.

Earlier policies targeted at boosting bioenergy within Europe have made rapeseed and sunflower relatively more profitable than grain legumes (ZANDER ET AL., 2016). In addition, farmers have reported that grain legumes are simply not profitable enough to compete with cereals, sugar beet and potatoes (RICHTHOFEN, 2006). Therefore, a higher incidence of one or more of the following crops may result in decreased grain legume production: rapeseed, sunflower, cereals, sugar beet and potatoes. However, in contrast to the above hypothesis, VOISIN ET AL. (2014) state that grain legumes are used as diversification crops in rotations based on cereals and oilseed rape. Consequently, an increase in cereals and rapeseed acreage may also have an opposite effect on grain legume acreage. Grain legume production may be more prevalent on mixed farms. They can grow their own grain legumes and use them as high

---

[2] as used by Eurostat

[3] European extension network for the development of grain legume production in the EU

protein feed instead of having to pay large transportation costs (WATSON ET AL., 2017) like specialized fattening farms, which purchase large quantities of animal feeds. On the other hand, a stronger specialization of agricultural holdings leads to a simplification and uniformity at field, farm, and regional level (LEMAIRE ET AL., 2015) and grain legume production may concentrate increasingly on specialized field crop farms instead. If the latter are in the same region as the specialized fattening farms the result would be a positive relationship between livestock density and the production of grain legumes; if they are located in other regions (or abroad) a negative correlation would result. Furthermore, a higher share of organic agriculture can have a positive effect on the production of grain legumes for two reasons. First, there will be an increased demand for GMO-free protein feed. Second, the prohibition of synthetic nitrogenous fertilizers in organic crop production makes crop rotations that comprise nitrogen fixing crops essential (BÖHM, 2009; WEHLING, 2009; BUES ET AL., 2013). Trade volumes tend to be lower in landlocked compared to coastal areas because overland transport costs tend to be higher than sea freight costs (RADELET and SACHS, 1998; LIMÃO and VENABLES, 2001). Thus, landlocked regions with a farther distance to main ports may import less protein feed from overseas with a positive effect on regional grain legume cultivation. By becoming member of an agricultural cooperative, farmers can achieve economies of scale in procurement, processing, marketing and distribution and can benefit from other joint activities such as information and knowledge exchange, research and promotion (COGECA, 2014). That way, agricultural cooperatives can address the often-mentioned obstacles to the widespread adoption of grain legumes such as the lack of marketing channels and poorly developed value chains for crops other than the “major crops” in many European countries (MEYNARD ET AL., 2013; RECKLING ET AL., 2016). Therefore, we suppose that legume production is more prevalent in regions where agricultural cooperatives focused on cereals, pulses and/or oilseed crops are already established. Further, we suggest that larger regional storage capacities may have a positive effect on grain legume production as they make farmers less vulnerable to price fluctuations on the world market. The current Common Agricultural Policy (CAP) offers several instruments that can support grain legume production in the EU, among them the possibility of Voluntary Coupled Support (VCS) to protein crops, including pulses, fodder legumes, rapeseed, sunflower seeds and soya beans (EUROPEAN COMMISSION, 2018_a). We expect a positive effect on grain legume production in those 16 EU Member States, which made use of VCS to protein crops in the year 2016 (EUROPEAN COMMISSION, 2015_a). In their farm-level economic analysis, PREISSEL ET AL. (2017) show that legume-supported cropping systems perform well where the use of nitrogen fertilizers is restricted. For that reason, we assume that grain legume production is more common in so called Nitrogen Vulnerable Zones (NVZs), where EU Member States have to establish Nitrate Action Programmes to reduce and prevent water pollution (EUROPEAN COMMISSION, 2018_b). ZANDER ET AL. (2016) show that the grain legume area in Mediterranean countries has declined less than in other European regions over the period 1961 to 2012. They refer to BOER ET AL. (2006) to suggest that this may be due to the prominent role of grain legumes in those countries´ regional diets. Finally, a larger share of grain legume area may coincide with a higher proportion of well-trained farmers because growing grain legumes requires more agronomic expertise than growing common cereals (WEHLING, 2009).

3. Econometric models to estimate regional grain legume cultivation

In our analysis, the dependent variables y are proportions. Hence, y is bounded between zero and one. An ordinary linear regression model is not appropriate for cases in which the dependent variable is restricted to the unit interval because it may generate fitted values that exceed its lower and upper bounds (FERRARI and CRIBARI-NETO, 2004). One common method to circumvent this problem is to model the log-odds ratio as a linear function (PAPKE and

WOOLDRIDGE, 1996). However, the inference based on the normality assumption can be misleading as proportions data typically display asymmetry (FERRARI and CRIBARI-NETO, 2004). In addition, heteroscedasticity is likely as the variance tends to decrease when the means get closer to one of the boundaries and the effect of independent variables tends to be non-linear (BUIS, 2010). Therefore, we focused on two different approaches suitable for proportions data that allow y to take on both zero and nonzero probability. The first approach assumes that regions with zero grain legume production occurred through the same process as regions with non-zero grain legume production. The second approach implies that there is something qualitatively different about regions that do not produce grain legumes at all and those regions that have some level of grain legume production (BUIS, 2010; BUIS, n.d.). The first approach is a generalized linear model for proportions data suggested by PAPKE and WOOLDRIDGE (1996). They assume an independent but not necessarily identically distributed sequence of observations $\{(x_i, y_i): i = 1,2, \dots, N\}$, where $0 \leq y_i \leq 1$ and N is the sample size. For all i,

$$(1) \quad \mathrm{E}(y_i|x_i) = G(x_i\beta)$$

where $G(.)$ is a known (link) function satisfying $0 < G(z) < 1$ for all $z \in \mathbb{R}$ (PAPKE and WOOLDRIDGE, 1996). In our analysis, we chose $G(.)$ to be the logistic function. The approach does not commit to specifying a particular distribution to estimate effects on the conditional mean $\mathrm{E}(y|x)$ but instead makes use of a fully robust and relatively efficient Bernoulli quasi-likelihood method. The Bernoulli, quasi-maximum likelihood estimator $\widehat{ß}$ is consistent and $\sqrt{N}$-asymptotically normal regardless of the true distribution of y_i (PAPKE and WOOLDRIDGE, 1996). One drawback of the approach could be that, besides the mean, no other quantities (i.e. the variance) are modelled (BUIS, 2010). The fractional logit generalized linear model was done using the Stata command “fracreg” and the model specifications “logit” and “robust” (Stata version 15.1). The second approach is a zero-one inflated beta (zoib) regression model. In contrast to the first approach, the zoib model does specify a distribution. In fact, it assumes a mixed continuous-discrete distribution with probability mass at zero or one (OSPINA and FERRARI, 2012). Thus, the zero-one inflated beta distribution comprises three parts: a probability that the dependent variable is zero, a probability that the dependent variable is one, and a beta distribution with the parametrization discussed in FERRARI and CRIBARI-NETO (2004) to describe the continuous component of the model (OSPINA and FERRARI, 2012; BUIS, n.d.). The beta densities can display many different shapes depending on the values of the mean of the response variable μ and the precision (or dispersion) parameter ϕ. It may be symmetric (when $\mu = 1/2$) or asymmetric (when $\mu \neq 1/2$) (FERRARI and CRIBARI-NETO, 2004). By means of maximum likelihood estimation the zoib regression model fits the zero-one inflated beta distribution to a distribution of a dependent variable (BUIS, n.d.). A detailed description of the model can be found in OSPINA and FERRARI (2012). The zoib regression model was run in Stata (Stata version 15.1) using the user-written command “zoib” (BUIS, 2012) as available from the Boston College Statistical Software Components (SSC) archive. The zoib results could be replicated by estimating a beta model for the continuous part between zero and one, and a logit model for the exact zero values (BUIS, 2011).

In addition, to account for spatial effects, we used spatial econometric models. The common version of the econometric models (see also ANSELIN, 1988; LESAGE, 1999) is provided by equations (2) and (3) as follows:

$$(2) \quad y = \rho W y + X\beta + u$$

$$(3) \quad u = \lambda W u + \varepsilon$$

with $\varepsilon \sim N(0, \sigma^2 I_N)$

where $y =$ the vector containing the observations for a dependent variable, each associated with a specific location $i (i = 1, \dots, N)$; $X =$ the design matrix containing in every row i the element 1 followed by a set of observations for the m explanatory variables; $W =$ a standardized spatial weight matrix; $I_N =$ the identity matrix; $u =$ the vector of spatially correlated residuals; $\varepsilon =$ the vector of normally distributed errors; $\beta =$ the vector containing the regression coefficients for the explanatory variables; $\rho =$ the spatial lag coefficient reflecting the importance of spatial dependence and $\lambda =$ the coefficient reflecting the spatial autocorrelation of the residuals u_i. We used a row-standardized inverse distance-based spatial neighborhood matrix W as we assumed spatial interactions to decrease with distance. However, the interactions between spatial units are not expected to be infinite, so a critical distance band was implied. The critical distance of 520 kilometers was chosen to ensure that every spatial unit i has at least one neighbor j. In this analysis we accounted for two versions of the spatial econometric model. First, the "spatial lag model" (where $\rho \neq 0$ and $\lambda = 0$) accounting for spatial dependence that may result from agglomeration effects. Second, the "spatial error model" (where $\rho = 0$ and $\lambda \neq 0$), which is more efficient than a common OLS model in the case of omitted spatially correlated explanatory variables. To support our model selection, we used the (robust) Lagrange Multiplier (LM) test as described in ANSELIN ET AL. (1996). The spatial models (equations 2 and 3) were estimated using the maximum likelihood method (cf. LESAGE and PACE, 2009). The spatial analysis was done in ArcGIS (version 10.0) and Stata (version 11.2) along with additional routines provided by JEANTY (2010_a-c) and (PISATI, n.d.).

4. Data and variable construction

For reasons of data availability the statistical analysis was conducted at the NUTS2[4] level (NUTS version 2013) excluding EU's overseas territories, some regions located off the coast, and combining the five London regions into Inner London and Outer London. Hence, the applied NUTS classification resulted in 27 countries and 262 NUTS2 regions. To account for varying NUTS-unit sizes, we defined all variables that relate to surface area as proportions of the total area or as densities. We consider sectoral agglomeration factors and five categories of independent variables: natural production factors, sector specific operational factors, infrastructure, political factors and socioeconomic factors (see also Annex 2). Precipitation and temperature data were extracted from FAO´s Global Agro-Ecological Zones data portal (GAEZ, 1961-1990). We obtained average climate values for each NUTS2 region using the QGIS tool "zonal statistics" (QGIS version 2.18.22). Quadratic terms of regional precipitation and temperature were also considered. Our soil data stems from the European Soil Database (LIEDEKERKE ET AL., 2006). We used the ArcGIS tool "zonal statistics" (ArcGIS version 10.5.1) to obtain the most common level of water capacity, base saturation and soil depth within each NUTS2 region. All variables describing specific soil characteristics were coded as dummy variables. A high base saturation served as a proxy variable for "chalky soil". The latest farm structure survey in the EU was carried out in the year 2013, from which the following data were derived: the UAA managed by mixed agricultural holdings (EUROSTAT, 2013_b) and organic agricultural holdings (EUROSTAT, 2013_c; DESTATIS, 2016_a), the irrigable UAA (EUROSTAT, 2013_d) and the number of livestock units (EUROSTAT, 2013_e; DESTATIS, 2016_b). Organic agricultural holdings include those under conversion to organic. We deducted pastures and meadows from the organic UAA in order to confine the area that farmers may use to cultivate grain legumes. Livestock includes all animals kept in holding, i.e. cattle, sheep, goats, pigs, poultry, equidae and other animals. It is expressed in livestock units (LSU). More recent (2016) data were available for the acreage of pulses, beans, soya, rapeseed and sunflower (EUROSTAT, 2016_a; DEFRA, 2016; INSTITUTUL NATIONAL DE STATISTICA, 2016; AMT FÜR STATISTIK BERLIN-BRANDENBURG, 2016; BAYERISCHES LANDESAMT FÜR STATISTIK, 2016; HESSISCHES STATISTISCHES LANDESAMT, 2016; IT.NRW,

---

[4] Nomenclature of Territorial Units for Statistics.

2016; LAIV, 2016; LANDESAMT FÜR STATISTIK NIEDERSACHSEN, 2016; STATISTISCHES AMT FÜR HAMBURG UND SCHLESWIG-HOLSTEIN, 2016; STATISTISCHES LANDESAMT BADEN-WÜRTTEMBERG, 2016; STATISTISCHES LANDESAMT RHEINLAND-PFALZ, 2016; STATISTISCHES LANDESAMT DES FREISTAATES SACHSEN, 2016; STATISTISCHES LANDESAMT SACHSEN-ANHALT, 2016; THÜRINGER LANDESAMT FÜR STATISTIK, 2016) and thus preferred. We obtained the distance between a region´s centroid and its nearest main port by locating each region's centroid using QGIS (version 2.18.22), then geographically locating the main ports handling dry bulk goods (EUROSTAT, 2016_b) using Google maps and finally calculating the distance using the QGIS tool "distance matrix". Information on agricultural cooperatives was taken from the 2011-2012 project "Support for Farmers' Cooperatives". Of the top 50 cooperatives identified in the respective 27 country reports (sources on request), those with a focus on cereals, pulses and/or oilseed crops were geographically located using Google maps. In the case of transnational cooperatives, only those locations within their country of origin were considered. In addition, if a cooperative had more than one location in the same region, they were counted as one. Data on Voluntary Coupled Support (VCS) to protein crops in 2016 and on the status of "Nitrate Vulnerable Zones" (NVZ) in 2015 (the latter in form of an ArcGIS map service) are published in EUROPEAN COMMISSION (2015_a and _b). Available NVZs were further processed in ArcGIS (version 10.5.1) in order to obtain the share of NVZ area in total surface area within each region. Some EU Member States chose to establish a Nitrate Action Programme on their entire territory instead of assigning specific subnational regions as NVZ. For those countries, zero values were assigned to the corresponding variable to indicate zero NVZ variation within their territory. Data on average daily (so-called "chronic") legume consumption (grain legumes and processed legumes) of adults aged between 18 to 64 years were taken from the Comprehensive European Food Consumption Database (EFSA, 2000-2015). Although food consumption data was available at national level only, we still considered it useful.

Annex 2 provides descriptive statistics for all dependent and independent variables as used in our analyses along with hypothesized directions of influence according to the theoretical considerations in Section 2.

5. Results and discussion

As the global Moran's *I* tests (cf. MORAN, 1948; ANSELIN, 1988) indicate a positive and highly significant autocorrelation for all dependent variables (see Annex 3) we reject the null-hypothesis of "no spatial dependence and / or spatial heterogeneity" and thus estimate spatial models. As also shown in Annex 3, the results of the (robust) Lagrange Multiplier test are heterogeneous. The spatial lag model is recommended as appropriate spatial model for pulses, while the spatial error model is suggested for beans and soya in order to take into account spatial autocorrelation. Table 2 presents the three model types for each dependent variable, the fractional logit generalized linear regression model (fracreg), the zero-one inflated beta regression model (zoib) and the spatial econometric model, together with their estimated (marginal) effects for selected predictor variables. Not all previously considered predictor variables were included in our models. First, a full model was fit on all explanatory variables. Then we excluded those that showed no significant effect on any of the three dependent variables analyzed. It stands out that the estimated marginal effects of the zoib model tend to be smaller (closer to zero) than the marginal effects of the fracreg model. This implies that, by assuming that regions with zero- and non-zero grain legume production have occurred through the same process (fracreg), the effect of explanatory variables tends to be overestimated. A qualitative difference of regions with zero grain legume production could be e.g. general unsuitability, lost knowledge of farmers about legume cultivation and their use, no extension services and training, or poorly established value chains for grain legumes. A high share of rapeseed seems to influence the cultivation of grain legumes positively. This could be due to the use of grain legumes as diversification crops in rotations based on cereals and oilseed rape. Regarding the share of sunflowers, our models present different results. Sunflowers seem to influence the share of soya positively, whereas the share of pulses is influenced negatively. Those regions with a higher share of sunflower may offer suitable growing conditions and marketing channels for other oilseed crops as well, e.g. soya. The results for pulses support our hypothesis that grain legumes were among the crops being replaced by state-aided bioenergy crops (i.e. sunflower). In case of beans, our results do not support the assumption, that mixed farms grow their own grain legumes and use them as high protein feed. A higher share of organic area has a significant positive effect on the share of arable land used for pulses and beans, but a negative effect on soya in the spatial model. The former results support the hypotheses, that grain legumes play an important role in organic crop rotation and GMO-free animal protein food production, the latter result however needs a different explanation. One could be a lack of specialized processing plants for post-harvest handling of organic soya beans in many regions of the EU. Livestock density is negatively related to the production of pulses and beans, which may be due to spatial dissociation of specialized livestock and crop production. In two models, a higher density of agricultural cooperatives focused on cereals, pulses and oilseed crops is slightly significantly associated with a smaller share of pulses and soya. It could be that those cooperatives use a larger share of arable land for the production of cereals and oilseed crops (other than soya) than they do for grain legumes. A larger distance between a region and its nearest main port seems to positively influence the share of soya, and negatively influence the share of pulses and beans in total arable land. These results support our assumption that a larger distance between a region and its nearest main port may decrease the relative cost advantage of feed imports and increase the regional production of high protein feed, especially soya, accordingly. The negative result for pulses and beans may be explained by the geographic location of the respective regions. Continental regions far from main ports exhibit hot summers and thus offer critical growing conditions for beans and other pulses (except chickpeas), which are only medium or little tolerant to heat stress (compare Table 1). As expected, voluntary coupled support to protein crops (including pulses, fodder legumes, rapeseed, sunflower seeds and soya beans) has a significant positive influence on the share of pulses in total arable land. However, the share of soya seems to be influenced negatively. We assume, that other protein crops turn out to be more competitive under coupled support payments and consequently replace soya to some extent. Next, our results show a slightly significant increase in the share of arable land used for beans with an increasing presence of NVZs in the respective region. This seems to support the findings of PREISSEL ET AL. (2017), which suggest that legume-supported cropping systems perform economically well where the use of nitrogen fertilizers is restricted. Increasing levels of food legume consumption seem to significantly increase the share of arable land used for pulses and beans in the respective region. Besides cultural reasons, this could be due to a growing emphasis of the food industry on regional production. The observed significant decrease in the production of soya where legume based diets are more prominent may be explained by the predominant use of soya for high protein feed instead for human consumption. Beans and soya are sensible to drought and thus cultivated in regions with sufficient water supply. According to the fracreg model, the estimated share of arable land used for beans is largest at a total annual precipitation of 972mm[5]. Soya on the other hand is associated with a higher share of irrigable UAA and a high available water capacity in the topsoil. Temperature has a significant effect on the production of pulses and soya. The estimated optimal temperature for pulses is between 3158 and 3546 degree sums per year and for soya between 3380 and 5386 degree sums per year.

---

[5] $T_{opt} = \frac{-0.0387}{2*(-0.0199)} * 1000 = 972{,}36\ mm/year$

Table 2: Estimated regression models of grain legume production (pulses, beans, soya)

	Share of arable land used for pulses in total arable land			Share of arable land used for beans in total arable land			Share of arable land used for soya in total arable land		
	fracreg model	zoib model	spatial model	fracreg model	zoib model	spatial model	fracreg model	zoib model	spatial model
Share of AA used for rapeseed in total AA	0.0129	0.0481 ***	0.0293	0.0272 **	0.0276 ***	0.0627 ***	-0.0360	0.0323 *	0.0321
Share of AA used for sunflower in total AA	-0.1007 ***	-0.0589 **	-0.0523 *	0.0064	-0.0152	0.0090	0.0722 ***	0.0656 **	-0.0498
Share of UAA managed by mixed farms in total UAA	-0.0004	-0.0020	-0.0103	-0.0111	-0.0103 *	-0.0066	-0.0204	0.0067	-0.0180
Share of organic UAA in total UAA	0.0427 ***	0.0352 ***	0.0412 **	0.0338 ***	0.0216 ***	0.0228 *	0.0078	0.0032	-0.1073 ***
Share of irrigable UAA in total UAA	-0.0244 ***	-0.0134 **	-0.0166 **	-0.0273 ***	-0.0108 ***	-0.0121 ***	0.0315 ***	0.0151 **	0.0730 ***
No. of livestock units per ha UAA	-0.0111 ***	-0.0047 ***	-0.0023 **	-0.0057 ***	-0.0017 ***	-0.0005	-0.0032	-0.0003	-0.0016
No. of cooperatives per 1000sqkm land area	-0.0038 *	-0.0018	-0.0030	-0.0002	-0.0001	0.0013	-0.0034	-0.0026 *	-0.0021
Distance to nearest main port (in 100km)	-0.0022 ***	-0.0015 **	-0.0013	-0.0046 ***	-0.0021 ***	-0.0020 **	0.0042 ***	0.0034 ***	0.0000
VCS to protein crops (in 100€/ha and year)	0.0064 ***	0.0048 ***	0.0039 ***	0.0012	0.0006	-0.0008	-0.0017	-0.0002	-0.0104 ***
Share of designated NVZs in total surface area	0.0011	0.0012	0.0001	0.0024	0.0018 *	0.0046 *	0.0032	-0.0014	0.0081
Chronic legume consumption of adults (in 100g/day)	0.1283 ***	0.0928 ***	0.0755 ***	0.0500 ***	0.0328 ***	0.0945 ***	-0.1001 ***	-0.0511 ***	0.0037
Total annual precipitation (in m)	-0.0011	0.0171	0.0009	0.0387 *	0.0219	0.0031	0.0156	0.0036	0.0011
Total annual precipitation (in m)^2	-0.0103	-0.0151	-0.0001	-0.0199	-0.0111	-0.0001	0.0097	0.0074	0.0003
Temperature sum of frost free days [a] (in 1000 dC)	0.013 **	0.0078	0.0120 **	0.0005	0.0009	-0.0031	0.0365 *	0.0049	0.0377 ***
Temperature sum of frost free days [a] (in 1000 dC)^2	-0.002 **	-0.0011 *	-0.0019 ***	0.0001	-0.0001	0.0007	-0.0054 **	-0.0005	-0.0035 **
Dummy high available water capacity (=1)	0.0004	-0.0028	0.0015	0.0021	0.0002	0.0014	0.0062 ***	0.0006	0.0008
Dummy high base saturation (=1)	0.0008	0.0007	-0.0003	0.0001	0.0021 ***	0.0008	0.0003	-0.0021 *	0.0038
Dummy deep soil (=1)	0.0032 *	0.0009	0.0044 **	0.0042 ***	-0.0002	0.0046 ***	0.0013	0.0023 **	0.0024
Constant	-5.19 ***	-5.59^{b} *** 77.42^{c} ***	-0.02	-7.6501 ***	-6.91^{b} *** 20.58^{c} ***	-0.02	-18.54 ***	-3.80^{b} ** 15.00^{c} ***	-0.093 **
$\boldsymbol{\rho}$			0.6418 ***						
$\boldsymbol{\lambda}$						0.6203 ***			0.8622 ***
No. of observations	203	203	203	200	200	200	227	227	227
Prob > χ^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

AA: Arable Area; UAA: Utilized Agricultural Area; No.: Number; VCS: Voluntary Coupled Support; NVZ: Nitrate Vulnerable Zones

[a] frost free days = days with average temperature > 10°C

[b] 0 < depvar < 1

[c] depvar = 0

***, ** and * indicate statistical significance at the 1, 5 and 10 % significance level, respectively.

Source: Authors´ own calculations based on different sources given in the text.

Our results do not show a good suitability of all common grain legumes except lupines for chalky soil because only beans are positively related to a high base saturation, whereas soya is not. The share of soya, beans and pulses is positively associated with deep soils, even though some crops (e.g. lupines, chickpeas and vetches) of the aggregate pulses could also grow well under shallow soil conditions. The spatial lag coefficient for pulses (0.6418) is highly significant and may hint to relevant agglomeration effects. The share of arable land used for pulses in total arable land in one region hence could positively influence the share of pulses in neighboring regions. We assume the proximity to customers and appropriate marketing channels to be of particular importance in the case of pulses. Additionally, local networks and advisory services could be relevant for the spatial location of pulses. The spatial error coefficients for beans (0.6203) and soya (0.8622) are highly significant as well. This hints at one or more omitted explanatory variables correlated with different locations in space. The breeding and supply of grain legume seeds and the availability of specific supply chains could be such variables that are relevant and unevenly distributed in the EU.

6. Conclusions

Using three different methodological approaches, our results show that several factors may contribute to a high share of grain legumes in the EU. In the case of pulses, these factors are: the prevalence of organic farming and rapeseed production, the prominent role of grain legumes in regional diets, and voluntary coupled CAP support to protein crops. For beans, high annual precipitation, the prominent role of grain legumes in regional diets, restrictions on the use of nitrogen fertilizers (NVZ), organic farming and rapeseed production seem to be favorable. For soya, we found that a high topsoil water capacity, the possibility to irrigate, a large distance to the nearest main port as well as rapeseed and sunflower production may promote its production. The production of pulses is regionally agglomerated. Regions with a high share of pulses tend to be close to each other. Thus, external economies of scale and positive spillover effects such as knowledge diffusion may play an important role in fostering legume cultivation. If so, they could be further exploited by support of investments in regional supply chains for pulses, local extension services and training. In this context, it should be pointed out that agglomeration was also observed for other sustainable farming practices off the mainstream such as organic farming (SCHMIDTNER ET AL., 2012). This may be of interest when designing transition paths to sustainable legume-based farming systems. The spatial level at which our analysis was conducted cannot depict large variations within different regions. Some explanatory variables that were not tested due to a lack of data but which could be of interest in consecutive studies are the incidence of processing facilities and agricultural trading companies, contract farming, aquaculture, support to grain legume production through national agri-environment schemes, extension services, training programs, and the regional availability of (suitable) seeds. In addition, time series analysis would be a promising approach to deepen the understanding of causalities in the location and intensity of grain legume production. Meaningful explanatory variables when studying time series may be yield and price fluctuations of grain legumes, grain legume imports (especially soya) and price levels of mineral fertilizers. Finally, future research could focus on a methodological approach combining the advantages of the applied non-linear models (fracreg and zoib) with those of spatial econometrics.

References

AMT FÜR STATISTIK BERLIN-BRANDENBURG, 2016. Ernteberichterstattung über Feldfrüchte und Grünland im Land Brandenburg 2016. [Online] Available at: https://www.statistik-berlin-brandenburg.de [Accessed 28 02 2018].

ANSELIN, L., 1988. Spatial econometrics: methods and models. Dordrecht, the Netherlands; Boston; London: Kluwer Academic Publishers.

ANSELIN, L., BERA, A. K., FLORAX, R. and YOON, M. J., 1996. Simple diagnostic tests for spatial dependence. Regional science and Urban Economics 26: 77-104.

BAYERISCHES LANDESAMT FÜR STATISTIK, 2016. Landwirtschaftliche Betriebe insgesamt 2016 nach jeweiligen Flächen und Anbaukulturen nach regionaler Einheit. [Online] Available at: https://www.statistik.bayern.de [Accessed 20 02 2018].

BESTE, A. and BOEDDINGHAUS, R., 2011. Artenvielfalt statt Sojawahn - Der Eiweißmangel in der EU: Wie lässt sich das seit langem bestehende Problem lösen?. [Online] Available at: http://www.martin-haeusling.eu/images/attachments/Proteinbroschuere_Haeusling_05sep11_klein.pdf [Accessed 29 11 2017].

BOER, J., HELMS, M. and AIKING, H., 2006. Protein consumption and sustainability: Diet diversity in EU-15. Ecological Economics 59: 267-274.

Böhm, H., 2009. Körnerleguminosen – Stand des Wissens sowie zukünftiger Forschungsbedarf aus Sicht des Ökologischen Landbaus. Journal für Kulturpflanzen 61(9): 324–331.

BUES, A.; PREIßEL, S.; RECKLING, M.; ZANDER, P.; KUHLMAN, T.; TOPP, K.; WATSON, C.; LINDSTRÖM, K.; STODDARD, F. L.; MURPHY-BOKERN, D., 2013. The environmental role of protein crops in the new common agricultural policy. [Online] Available at: http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/495856/IPOL-AGRI_ET(2013)495856_EN.pdf [Accessed 23 09 2018].

BUIS, M., 2010. Analyzing Proportions. [Online] Available at: https://www.stata.com/meeting/germany10/germany10_buis.pdf [Accessed 30 08 2018].

BUIS, M., 2011. ZOIB procedure. [Online] Available at: https://www.stata.com/statalist/archive/2011-09/msg00839.html [Accessed 16 01 2019].

BUIS, M., 2012. 'ZOIB': module to fit a zero-one inflated beta distribution by maximum likelihood. s.l.:StataCorp LLC.

BUIS, M., n.d.. zoib. [Online] Available at: http://www.maartenbuis.nl/software/zoib.html [Accessed 23 10 2018].

CERNAY, C.; BEN-ARI, T.; PELZER, E.; MEYNARD, J.; MAKOWSKI, D., 2015. Estimating variability in grain legume yields across Europe and the Americas. Scientific Reports 5(11171): 1-11.

CHALK, P., 1998. Dynamics of biologically fixed N in legume-cereal rotations: A review. Australian Journal of Agricultural Research 49: 303-316.

COGECA, 2014. Development of agricultural cooperatives in the EU 2014. [Online] Available at: http://www.copa-cogeca.eu/Cogeca [Accessed 21 09 2018].

DEFRA, 2016. Structure of the agricultural industry in England and the UK at June. [Online] Available at: https://www.gov.uk/government/statistical-data-sets/structure-of-the-agricultural-industry-in-england-and-the-uk-at-june [Accessed 28 02 2018].

DESTATIS, 2016_a. Landwirtschaftliche Betriebe und deren landwirtschaftlich genutzte Fläche (LF) nach Art der Bewirtschaftung. [Online] Available at: https://www-genesis.destatis.de/genesis/online [Accessed 11 09 2018].

DESTATIS, 2016_b. Landwirtschaftliche Betriebe mit Viehhaltung und Zahl der Tiere. [Online] Available at: https://www-genesis.destatis.de/genesis/online [Accessed 26 09 2018].

EC-DG AGRI, 2011. Oilseeds and protein crops in the EU. [Online] Available at: http://ec.europa.eu/agriculture/cereals/factsheet-oilseeds-protein-crops_en.pdf [Accessed 28 02 2019].

EC-DG AGRI, 2014. EIP-AGRI Focus Group - Protein Crops: final report. [Online] Available at: https://ec.europa.eu/eip/agriculture/en/publications/eip-agri-focus-group-protein-crops-final-report [Accessed 06 07 2018].

EFSA, 2000-2015. European Food Safet Authority - Chronic and acute food consumption statistics - Level 2. [Online] Available at: http://www.efsa.europa.eu/en/microstrategy/foodex-level-2 [Accessed 17 08 2018].

EUROPEAN COMMISSION, 2015_a. Voluntary coupled support – Sectors mostly supported: Notification of decisions taken by Member States by 1 August 2014. [Online] Available at: https://ec.europa.eu/agriculture/sites/agriculture/files/direct-support/direct-payments/docs/voluntary-coupled-support-note_en.pdf [Accessed 18 01 2019].

EUROPEAN COMMISSION, 2015_b. Nitrates / NVZ Reporting 2015. [Online] Available at: https://water.jrc.ec.europa.eu/arcgis/rest/services/Nitrates/NvzReporting2015/MapServer/0 [Accessed 10 12 2018].

EUROPEAN COMMISSION, 2018_a. Report on the development of plant proteins in the EU (COM(2018) 757 final). [Online] Available at: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/plants_and_plant_products/documents/report-plant-proteins-com2018-757-final_en.pdf [Accessed 27 11 2018].

EUROPEAN COMMISSION, 2018_b. Questions and answers: Nitrates directive implementation report 2012-2015. [Online] Available at: http://ec.europa.eu/environment/pdf/04_05_2018_QA_en.pdf [Accessed 08 02 2019].

EUROSTAT, 2013_a. [ef_lus_main] Main farm land use by NUTS 2 regions. [Online] Available at: https://ec.europa.eu/eurostat/data/database [Accessed 27 09 2018].

EUROSTAT, 2013_b. [ef_m_farmleg] Farm indicators by agricultural area, type of farm, standard output, legal form and NUTS 2 regions. [Online] Available at: https://ec.europa.eu/eurostat/data/database [Accessed 28 06 2018].

Eurostat, 2013_c. [ef_mporganic] Organic farming: number of farms, areas with different crops and heads of different types of animals by agricultural size of farm (UAA) and NUTS 2 regions. [Online] Available at: https://ec.europa.eu/eurostat/data/database [Accessed 20 06 2018].

Eurostat, 2013_d. [aei_ef_ir] Share of irrigable and irrigated areas in utilised agricultural area (UAA) by NUTS 2 regions. [Online] Available at: https://ec.europa.eu/eurostat/de/data/database [Accessed 26 09 2018].

Eurostat, 2013_e. [ef_olslsureg] Livestock: number of farms and heads of animals by livestock units (LSU) of farm and NUTS 2 regions. [Online] Available at: https://ec.europa.eu/eurostat/data/database [Accessed 18 06 2018].

Eurostat, 2016_a. [apro_cpshr] Crop production in EU standard humidity by NUTS 2 regions. [Online]
Available at: https://ec.europa.eu/eurostat/data/database [Accessed 08 08 2018].

Eurostat, 2016_b. [mar_go_qmc] Gross weight of goods handled in main ports by direction and type of cargo - quarterly data. [Online] Available at: https://ec.europa.eu/eurostat/de/data/database [Accessed 13 07 2018].

FAOSTAT, 2015. Statistics Database of the Food and Agriculture Organization of the United Nations, Rome. [Online] Available at: http://faostat3.fao.org/home/E [Accessed 02 02 2016].

Ferrari, S. and Cribari-Neto, F., 2004. Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics 31(7): 799-815.

GAEZ, 1961-1990. Global Agro-ecological Zones. [Online]
Available at: http://www.gaez.iiasa.ac.at/ [Accessed 26 06 2018].

GL-Pro, 2005. Ratgeber für den Anbau von Körnerleguminosen in Europa. [Online]
Available at: www.grainlegumes.com/gl-pro/ [Accessed 29 11 2017].

Häusling, M., 2011. The EU protein deficit: what solution for a long-standing problem?. [Online]
Available at: http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2011-0026&language=EN [Accessed 06 07 2018].

Hessisches Statistisches Landesamt, 2016. Landwirtschaftliche Betriebe insgesamt 2016 nach jeweiligen Flächen und Anbaukulturen nach regionaler Einheit. [Online] Available at: https://statistik.hessen.de [Accessed 21 02 2018].

Institutul National de Statistica, 2016. Area cultivated with main crops by ownership form, macroregions, development regions and counties. [Online] Available at: www.insse.ro/cms/ro [Accessed 16 11 2018].

IT.NRW, 2016. Ackerland der landwirtschaftlichen Betriebe 2016 nach Fruchtarten sowie kreisfreien Städten und Kreisen. [Online] Available at: https://webshop.it.nrw.de [Accessed 20 02 2018].

Jeanty, P., 2010_a. [Anketest] Stata module to perform diagnostics tests for spatial autocorrelation in the residuals from OLS, SAR, IV, and IV-SAR models. [Online] Available at: http://fmwww.bc.edu/RePEc/bocode/s [Accessed 31 03 2011].

Jeanty, P., 2010_b. [Splagvar] Stata module to generate spatially lagged variables, construct the Moran scatter plot, and calculate global Moran´s I statistics. [Online] Available at: http://fmwww.bc.edu/RePEc/bocode/s [Accessed 31 03 2011].

Jeanty, P., 2010_c. [Spmlreg] Stata module to estimate the spatial lag, the spatial error, the spatial durbin, and the general spatial models. [Online] Available at: http://fmwww.bc.edu/RePEc/bocode/s [Accessed 31 03 2011].

LAiV, 2016. Bodennutzung der Betriebe in Mecklenburg-Vorpommern. [Online]
Available at: https://www.laiv-mv.de/Statistik/ [Accessed 21 02 2018].

Landesamt für Statistik Niedersachsen, 2016. Bodennutzung und Ernte 2016. [Online]
Available at: https://www.statistik.niedersachsen.de [Accessed 20 02 2018].

Lemaire, G., Gastal, F., Franzluebbers, A. and Chabbi, A., 2015. Grassland–Cropping Rotations: An Avenue for Agricultural Diversification to Reconcile High Production with Environmental Quality. Environmental Management 56: 1065–1077.

LeSage, J., 1999. Spatial econometrics. [Online]
Available at: http://www.rri.wvu.edu/WebBook/LeSage/spatial/spatial.html [Accessed 02 02 2003].

LeSage, J. and Pace, R. K., 2009. Introduction to spatial econometrics. Boca Raton, USA: Chapman&Hall/ CRC.

LfL, 2014. Bayerische Landesanstalt für Landwirtschaft - Sojabohne - der Eiweißlieferant. [Online] Available at: https://www.lfl.bayern.de/mam/cms07/publikationen/daten/informationen/061490_sojabohne.pdf [Accessed 13 02 2018].

Liedekerke, M., Jones, A. and Panagos, P., 2006. ESDBv2 Raster Library - a set of rasters derived from the European Soil Database distribution v2.0 (CD-ROM, EUR 19945 EN), Brussels: European Commission and the European Soil Bureau Network.

Limão, N. and Venables, A., 2001. Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. The world bank economic review 15(3): 451–479.

Meynard, J.M.; Messéan, A.; Charlier, A.; Charrier, A.; Farès, M.; Le Bail, M.; Magrini, M.B.; Savini, I., 2013. Crop diversification: obstacles and levers - Study of farms and supply chains. [Online] Available at: https://www6.paris.inra.fr/depe/content/download/3736/35824/version/1/file/Version+Anglaise+Diversification-8pages.pdf [Accessed 21 09 2018].

Moran, P. H. P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society B 10: 243-251.

Ospina, R. and Ferrari, S., 2012. A general class of zero-or-one inflated beta regression models. Computational Statistics and Data Analysis 56: 1609-1623.

Papke, L. and Wooldridge, J., 1996. Econometric methods for fractional response variables with an application to 401(K) plan participation rates. Journal of Applied Econometrics 11: 619-632.

PISATI, M., n.d.. [Spatwmat] Spatial weights matrices for spatial data analysis. [Online] Available at: http://www.stata.com/stb/stb60 [Accessed 31 03 2011].

PREISSEL, S., RECKLING, M., BACHINGER, J. and ZANDER, P., 2017. Introducing legumes into European Cropping Systems: Farm-level Economic Effects. In: Legumes in Cropping Systems. Wallingford: CAB International: 209-225.

PREISSEL, S., RECKLING, M., SCHLÄFKE, N. and ZANDER, P., 2015. Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: A review. Field Crops Research 175: 64-79.

RADELET, S. and SACHS, J., 1998. Shipping costs, manufactured exports, and economic growth. [Online] Available at: http://documents.worldbank.org/curated/en/483661483605903571/Shipping-costs-manufactured-exports-and-economic-growth [Accessed 19 07 2018].

RECKLING, M.; BERGKVIST, G.; WATSON, C.; STODDARD, F.; ZANDER, P.; WALKER, R.; PRISTERI, A.; TONCEA, I.; BACHINGER, J., 2016. Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. Frontiers on Plant Science 7(669): 1-15.

RICHTHOFEN, J.-S. V., 2006. What do European farmers think about grain legumes?. Grain Legumes 45, 1st quarter: 14-15.

RP FREIBURG, 2017. Anbauanleitung für Sojabohnen 2017. [Online] Available at: https://rp.baden-wuerttemberg.de/rpf/Abt3/Documents/ [Accessed 25 01 2018].

SCHMIDTNER, E.; LIPPERT, C.; ENGLER, B.; HÄRING, A.; AURBACHER, J.; DABBERT, S., 2012. Spatial distribution of organic farming in Germany: does neighbourhood matter? European Review of Agricultural Economics 39(4): 661-683.

SIDDIQUE, K.; JOHANSEN, C.; TURNER, N.; JEUFFROY, M.-H.; HASHEM, A.; SAKAR, D.; GAN, Y.; ALGHAMDI, S., 2011. Innovations in agronomy for food legumes. A review. Agronomy for Sustainable Development 32: 45-64.

STATISTISCHES AMT FÜR HAMBURG UND SCHLESWIG-HOLSTEIN, 2016. Die Bodennutzung in Hamburg und Schleswig-Holstein 2016. [Online] Available at: https://www.statistik-nord.de [Accessed 20 12 2017].

STATISTISCHES LANDESAMT BADEN-WÜRTTEMBERG, 2016. Bodennutzung der landwirtschaftlichen Betriebe in Baden-Württemberg 2016 nach Regierungsbezirken. [Online] Available at: https://www.statistik-bw.de [Accessed 22 12 2017].

STATISTISCHES LANDESAMT DES FREISTAATES SACHSEN, 2016. Bodennutzung und Ernte im Freistaat Sachsen 2016. [Online] Available at: https://www.statistik.sachsen.de [Accessed 22 02 2018].

STATISTISCHES LANDESAMT RHEINLAND-PFALZ, 2016. Ackerland landwirtschaftlicher Betriebe 2016 nach Fruchtarten. [Online] Available at: https://www.statistik.rlp.de [Accessed 04 07 2018].

STATISTISCHES LANDESAMT SACHSEN-ANHALT, 2016. Agrarstrukturerhebung Teil1 - Bodennutzung 2016. [Online] Available at: https://www.statistik.sachsen-anhalt.de [Accessed 20 02 2018].

THÜRINGER LANDESAMT FÜR STATISTIK, 2016. Landwirtschaftliche Betriebe insgesamt 2016 nach jeweiligen Flächen und Anbaukulturen (0102.1 T). [Online] Available at: https://statistik.thueringen.de [Accessed 21 02 2018].

VOISIN, A.-S.; GUÉGUEN, J.; HUYGHE, C.; JEUFFROY, M.-H.; MAGRINI, M.-B.; MEYNARD, J.-M.; MOUGEL, C.; PELLERIN, S.; PELZER, E., 2014. Legumes for feed, food, biomaterials and bioenergy in Europe: a review. Agronomy for Sustainable Development 34: 361-380.

WATSON, C.; RECKLING, M.; PREISSEL, S.; BACHINGER, J.; BERGKVIST, G.; KUHLMAN, T.; LINDSTRÖM, K.; NEMECEK, T.; TOPP, C.; VANHATALO, A.; ZANDER, P.; MURPHY-BOKERN, D.; STODDARD, F.L., 2017. Grain Legume Production and Use in European Agricultural Systems. Advances in Agronomy 144: 235-303.

WATSON, C. and STODDARD, F., 2017. Introduction - Perspectives on Legume Production and Use in European Agriculture. In: Legumes in Cropping Systems. Wallingford: CAB International: 1-17.

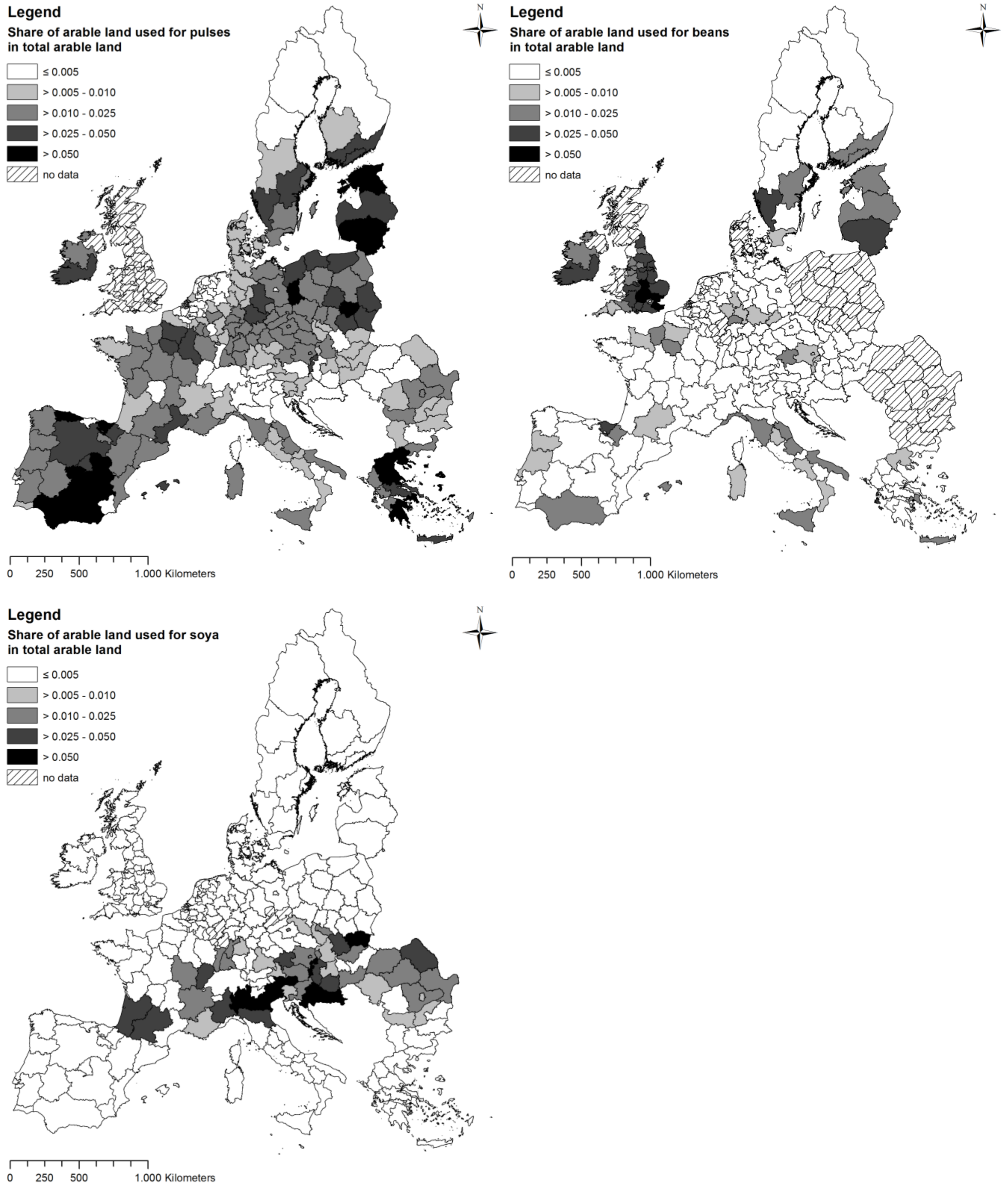
WEHLING, P., 2009. Anbau und Züchtung von Leguminosen in Deutschland – Sachstand und Perspektiven. Journal für Kulturpflanzen 61(9): 359-364.

WESTHOEK, H.; ROOD, T.; VAN DEN BERG, M.; NIJDAM, D.; REUDINK, M.; STEHFEST, E.; JANSE, J., 2011. The protein puzzle. [Online] Available at: https://www.pbl.nl/en/publications/2011/meat-dairy-and-fish-options-for-change-in-production-and-consumption [Accessed 24 02 2019].

ZANDER, P.; AMJATH-BABU, T. S.; PREISSEL, S.; RECKLING, M.; BUES, A.; SCHLÄFKE, N.; KUHLMANN, T.; BACHINGER, J.; UTHES, S.; STODDARD, F.; MURPHY-BOKERN, D.; WATSON, C., 2016. Grain legume decline and potential recovery in European agriculture: a review. Agronomy for Sustainable Development 36(26): 1-20.

Annex

Annex 1: Spatial distribution of pulses, beans and soya cultivation in the EU-28 in 2016



Source: Authors´ own illustration based on different sources given in the text.

Annex 2: Descriptive statistics for dependent variables and independent variables with hypothesized direction of influence

Dependent Variables	Year	Hypothesis	**N**	Mean	Std. Dev.	Median	Min	Max
Share of AA used for pulses in total AA	2016 [a]		223	0.02	0.02	0.01	0.00	0.11
Share of AA used for beans in total AA	2016 [a]		213	0.01	0.01	0.00	0.00	0.08
Share of AA used for soya in total AA	2016 [a]		253	0.01	0.03	0.00	0.00	0.35
Natural production factors								
Total annual precipitation (in m)	1961-1990	+/-	262	0.78	0.22	0.74	0.00	1.87
Temperature sum of frost free days [c] (in 1000 dC)	1961-1990	+/-	262	2.76	0.97	2.58	0.00	6.29
Dummy high available water capacity (=1)	2006	+/-	261	0.84	0.36	1.00	0.00	1.00
Dummy high base saturation (=1)	2006	+/-	261	0.65	0.48	1.00	0.00	1.00
Dummy deep soil (=1)	2006	+/-	261	0.60	0.49	1.00	0.00	1.00
Sector specific operational factors								
Share of AA used for rapeseed in total AA	2016 [a]	+/-	254	0.05	0.06	0.03	0.00	0.25
Share of AA used for sunflower in total AA	2016 [a]	-	256	0.02	0.05	0.00	0.00	0.29
Share of UAA managed by mixed farms in total UAA	2013	+/-	261	0.16	0.10	0.14	0.00	0.52
Share of organic UAA in total UAA	2013	+	261	0.03	0.05	0.02	0.00	0.33
Share of irrigable UAA in total UAA	2013	+	261	0.12	0.16	0.04	0.00	0.85
No. of livestock units per ha UAA	2013 [b]	+/-	262	0.88	0.99	0.60	0.00	7.60
Infrastructure								
Distance to nearest main port (in 100km)	2016	+	262	1.35	1.31	0.85	0.05	6.37
No. of agricultural cooperatives per 1000sqkm land area	2012	+	254	0.26	0.48	0.13	0.00	5.03
Political factors								
VCS to protein crops (in 100€/ha and year)	2016	+	262	0.60	0.77	0.18	0.00	4.17
Share of designated NVZ in total surface area	2015	+	255	0.23	0.32	0.05	0.00	1.01
Socio-economic factors								
Chronic legume consumption of adults aged 18-64 years (in 100g/day)	2000-2015	+	262	0.12	0.08	0.10	0.02	0.29

[a] data for Italy from year 2015

[b] data for Germany from year 2016

[c] frost free days = days with average temperature > 10°C

AA: Arable Area; UAA: Utilized Agricultural Area; dC: degree Celsius; No.: Number; VCS: Voluntary Coupled Support; NVZ: Nitrate Vulnerable Zones

+/- = the variable positively / negatively influences the share of pulses, beans or soya in total arable land in a region.

Source: Authors´ own calculations based on different sources given in the text.

Annex 3: Diagnostic tests for spatial autocorrelation of the dependent variables

	Share of arable land used for pulses in total arable land	Share of arable land used for beans in total arable land	Share of arable land used for soya in total arable land
Moran´s ***I***	0.33 ***	0.40 ***	0.14 ***
LM (spatial error)	36.83 ***	12.29 ***	11.28 ***
Robust LM (spatial error)	0.05	9.97 ***	3.66 *
LM (spatial lag)	54.98 ***	3.17 *	7.82 ***
Robust LM (spatial lag)	18.20 ***	0.85	0.19

***, ** and * indicate statistical significance at the 1, 5 and 10 % significance level, respectively.

LM=Lagrange Multiplier test

Source: Authors´ own calculations based on different sources given in the text.