

Article

Linking Arable Crop Occurrence with Site Conditions by the Use of Highly Resolved Spatial Data

Susanne Stein ^{*,†}, Horst-Henning Steinmann and Johannes Isselstein

Centre of Biodiversity and Sustainable Land Use, University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany; hsteinm@gwdg.de (H.-H.S.); jissels@gwdg.de (J.I.)

* Correspondence: susanne.stein@posteo.de

† Current Affiliation: Leibniz Centre for Agricultural Landscape Research (ZALF), Research Platform Data, Eberswalder Straße 84, 15374 Müncheberg, Germany.

Received: 27 February 2019; Accepted: 12 April 2019; Published: 18 April 2019



Abstract: Agricultural land use is influenced in different ways by local factors such as soil conditions, water supply, and socioeconomic structure. We investigated at regional and field scale how strong the relationship of arable crop patterns and specific local site conditions is. At field scale, a logistic regression analysis for the main crops and selected site variables detected, for each of the analyzed crops, its own specific character of crop–site relationship. Some crops have diverging site relations such as maize and wheat, while other crops show similar probabilities under comparable site conditions, e.g., oilseed rape and winter barley. At the regional scale, the spatial comparison of clustered variables and clustered crop pattern showed a slightly stronger relationship of crop combination and specific combinations of site variables compared to the view of the single crop–site relationship.

Keywords: arable land use pattern; maize; k-means; cluster analysis; Land Parcel Identification System; crop rotation

1. Introduction

In the last several decades, European arable farming has been characterized by modifications of cropping patterns and crop choice driven by enormous progress in plant breeding, plant protection, fertilization, and drainage techniques [1,2]. Additionally, market prices, farm subsidies, and political incentives such as support of bioenergy crops have influenced crop choice [3–5]. Recent studies have shown that a few cash crops are preferentially grown both in time and space while other crops are neglected [6,7]. In Northern Germany, maize and winter wheat are cropped on more than 50% of the arable area, and in many regions only one to three relevant crops are grown [7]. On the other hand, a decreasing importance of regional site conditions such as soil conditions, water supply, and climate for choosing a crop for a given site can be observed [8,9]. Thus, the relationship between site conditions and farmers crop choice (hereafter referred to as the crop–site relationship) seems to become weaker in modern farming.

One initial objective of the Common Agricultural Policy (CAP) is to increase productivity. This policy, therefore, has been a major driver of land use change for many decades [10]. The reform of 2003 introduced new rules of payments to farmers. Payments were decoupled from production to Single Farm Payment. At the same time, intervention prices for specific crops were maintained. National schemes on the promotion of renewable energy crops supported the intensive cultivation of crops for biomass production [11]. All this resulted in a continuation of intensive arable production in many historically intensively managed regions [12–14]. The latest reform of the CAP in 2013 implemented political instruments that are commonly named with the term “greening” [15] such as crop diversification. However, there is lack of knowledge ensuring that farmers have enough options to

diversify crop rotations. In a recent approach, it was shown on the basis of spatial data that some crop rotation patterns refer to site conditions, whereas others explicitly do not [16]. To our knowledge, there is no spatial explicit information on the extent to which the crop–site relationship still exists in recent landscapes. We present here a method to detect the relationship of crop cultivation and site conditions to improve the understanding and assessment of ecosystem services in the agricultural system.

With the presented methods, a binary logistic regression and a k-means clustering, we analyzed crop patterns in the landscape to understand to what extent crop choice still depends on site conditions. We first explored how intensive the individual relationship between the single crop and the single site variable is. Second, we localized regions of relationship between the clustered sets of site variables and the clustered crop patterns. Our study combines site variables and crop data of the year 2011 for the German federal state Niedersachsen (Lower Saxony), which includes an exceptional variety of agricultural systems. These characteristics made the region a good example for other arable regions and for the estimation of future trends in agricultural land use.

2. Materials and Methods

2.1. Research Area

Lower Saxony is characterized by various site conditions and a broad spectrum of agricultural land uses. The 2.6 million ha of farmland are cultivated by 41,730 farms with an average farm size of 61.8 ha [17]. During the last decade, maize (*Zea mays* L.) became the most dominant crop followed by winter wheat (*Triticum aestivum* L.) and oilseed rape (*Brassica napus* L.) (Figure 1). The northwestern part is dominated by marshy land with maritime climate, a high proportion of permanent grassland and extensive cattle breeding in the north and livestock breeding in the west. The cropping proportion of maize on arable land is above average for the Lower Saxonian acreage in this region. In the eastern part, sandy moraine soils with mixed farms are dominating. Arable farming characterizes the middle and south of Lower Saxony established on loessial soils in a hilly terrain influenced by subcontinental climate. The preferred crops under these conditions are sugar beet (*Beta vulgaris* subsp. *vulgaris*), oilseed rape, and winter wheat.

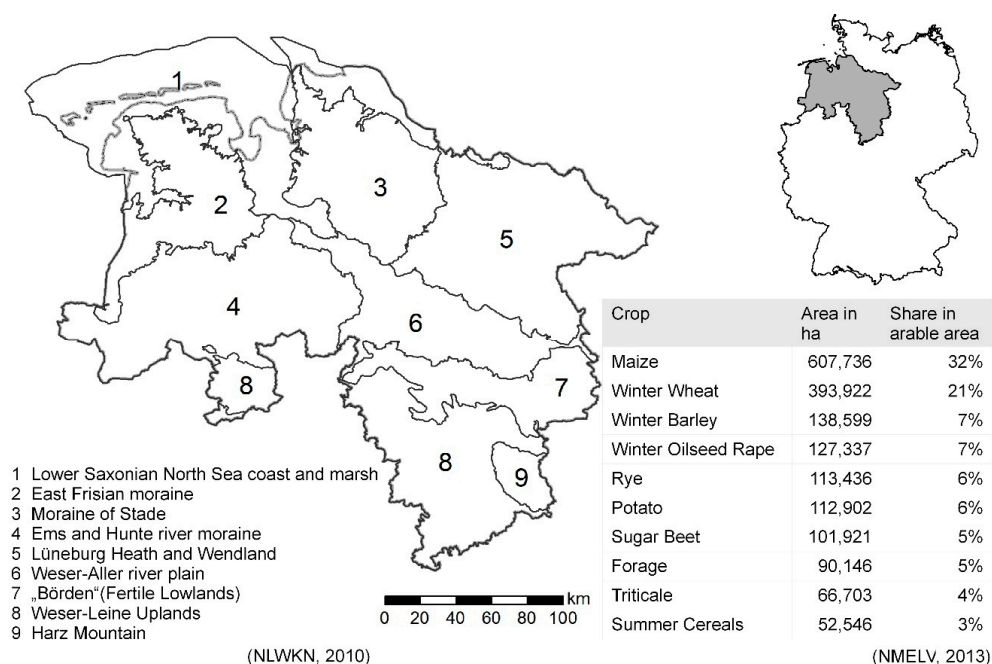


Figure 1. Natural area classification of the German federal state of Niedersachsen (Lower Saxony NUTS 1 region DE9 (European Nomenclature of Territorial Units for Statistics)) and the acreage of the ten main crops or crop groups in 2011, forage includes arable and temporary grassland and fodder legumes [17,18].

2.2. Data Characteristics and Processing

Our analysis followed two complementary approaches to detect the characteristics and spatial distribution of specific crop–site relationship. In a first step, a logistic regression analysis was processed that combines crop information at the field scale for the ten most commonly used crops in Lower Saxony with site variables such as soil, precipitation, and livestock density to characterize the relationship between these and the crops at the field scale. This result is compared with the result from a k-means clustering process to localize spatial overlays of clustered crops and clustered site variables at the regional scale.

For the crop data at the field scale, the Land Parcel Identification System (LPIS) was used, which is a yearly updated database that supports the administration of direct payments for European farmers as part of the Integrated and Control System (IACS). It was established in all member states of the European Union in 1992 and developed concurrently with political reform measures [19]. In Germany, the data are managed by the German Federal States' institutions. The access is limited due to privacy protection reasons, and special permission is required for scientific use. For this study, information about the main agricultural land use type in 2011, the field size, and individual field identification numbers were provided for the state Lower Saxony. The dataset was attributed to a GIS-geometry that comprises the boundaries for all agricultural parcels (about 990,000 records in total) [20]. Due to a small amount of imprecise field identification, e.g., the assignment of one ID to more than one field, the IACS dataset had to be debugged for uncertainties. For the analysis, only arable fields were included. Hence, with a loss of 15% due to imprecise field identification and intersection loss, the basic dataset of the analysis consists of 444,009 agricultural parcels.

To analyze the crop–site relationship, it was necessary to find spatial variables that represent the site conditions of the investigated area in a suitable resolution and area-wide consistent availability. Official data from well-established public sources satisfied these requirements (Table 1). The variables were selected with the aim to represent the environmental site conditions in Lower Saxony. This northwestern part of Germany is characterized by high local densities of livestock husbandry and grassland farming ([21], Figure 2). Therefore, variables concerning animal production were included.

Table 1. Site variables with their classes, units, and source scale. Classification of the metric variables was implemented corresponding to the geometrical intervals.

Predictor Variable	Classes	Unit	Source
Arable farming potential	1–7 Classes: 'extremely low' to 'extremely high'		[22] 1:50,000
Soil texture (dominant surface textural class of the soil)	1 Peat soil 2 Coarse (>65% sand) 3 Medium (<65% sand) 4 Medium fine (<15% sand) 5 Fine (>35% clay)		[23] 1:1,000,000
Slope (dominant slope class)	1 Level (<8%) 2 Sloping (8–15%) 3 Moderately steep (>15%)		[23] 1:1,000,000
Multi-annual precipitation sum (1981–2010)	1 560–676 2 677–746 3 747–806 4 807–878 5 879–1202	mm*y ⁻¹	[24] 0.96 × 0.96 km
Relative biotope density	1 0.00–0.90 2 0.91–1.10 3 1.11–2.00 4 2.01–18.76	Observed Density/Potential Density	[25] LAU-2

Table 1. Cont.

Predictor Variable	Classes	Unit	Source
Grassland proportion	1 0.00–0.02	ha/ha agric. area	Based on IACS-data 2011 1 × 1 km
	2 0.03–0.06		
	3 0.07–0.17		
	4 0.18–0.44		
	5 0.45–1.00		
Cattle density	1 0.00–0.10	Livestock unit/ha (agricultural area)	[26]
	2 0.11–0.29		LAU-2
	3 0.30–0.65		
	4 0.66–1.32		
	5 1.33–2.93		
Pig/poultry density	1 0.00–0.02	Livestock unit/ha (agricultural area)	[26]
	2 0.03–0.09		LAU-2
	3 0.10–0.30		
	4 0.31–0.99		
	5 1.00–3.21		
Average farm size	1 0–40	ha (agricultural area)	[26]
	2 41–64		LAU-2
	3 65–104		
	4 105–172		
	5 172–311		

The data for cattle density, pig, and poultry density, and the average farm size were extracted from agricultural census data at LAU-2 (Local Administrative Unit) scale (Figure 2). The relative biotope index was developed by the Julius Kühn-Institute, the German Federal Research Centre for Cultivated Plants, to estimate the biotope features in agricultural landscapes. The value for the relative biotope density was calculated using the locally observed density of linear biotope habitats (field margins and hedgerows) and patch biotopes (small woods and grassland patches) per estimated minimum biotope density at LAU-2 scale. The latter was extrapolated from the intensity of plant protection in the corresponding landscape type—the higher the intensity of plant protection applications is, the higher the need for biotopes is [27]. The proportion of grassland refers to the area of grassland per arable area in a 1 × 1 km cell of a raster. The multi-annual precipitation sum (1981–2010, [24]) is available in a 0.96 × 0.96 km raster format. The temperature was not regarded due to the low variation of the thermal regime in the study region. For the soil texture and slope information, the data of the European Soil Database were used which are available in so-called Soil Typological Units [23]. The arable farming potential was derived by the Lower Saxonian State Office for Mining, Energy and Geology (LBEG) based on soil and climate parameters (e.g., soil texture, bulk density, humus content, soil structure, and water logging level) [28]. The higher the value of the arable farming potential is, the higher the natural locally potential for biomass production of the soil is. For the regression analysis, all metric variables were transformed from metric values into interval values to facilitate the comparison of the variables' potential (Table 1). The classification of the intervals was implemented by a geometrical interval algorithm that minimizes the sum of the squares of the number of elements per class to ensure approximately the same number of values in each range [29].

Due to the differences in format and spatial scales of the used datasets, they were processed in relation to a reference scale. For the logistic regression the reference scale was the field scale. For the cluster process the information content of the variable polygons was attributed to a 1 × 1 km grid according to their spatial location and proportion. Grid cells with less than 10% of arable area within the grid cell area, i.e., less than 10 ha of arable area, were not included in the analysis. The merging of the attributed information was performed with the Spatial Join tool in ArcGIS®. For the small patched polygons of the arable farming potential, the mean of all soil classes per quadrant was attributed. Furthermore, the grid surface permits the calculation of the crop area proportion (crop area per arable

area in a 1×1 km grid cell) as metric variables. The crop area per grid cell is the sum of all fields that had their centroid within one grid cell.

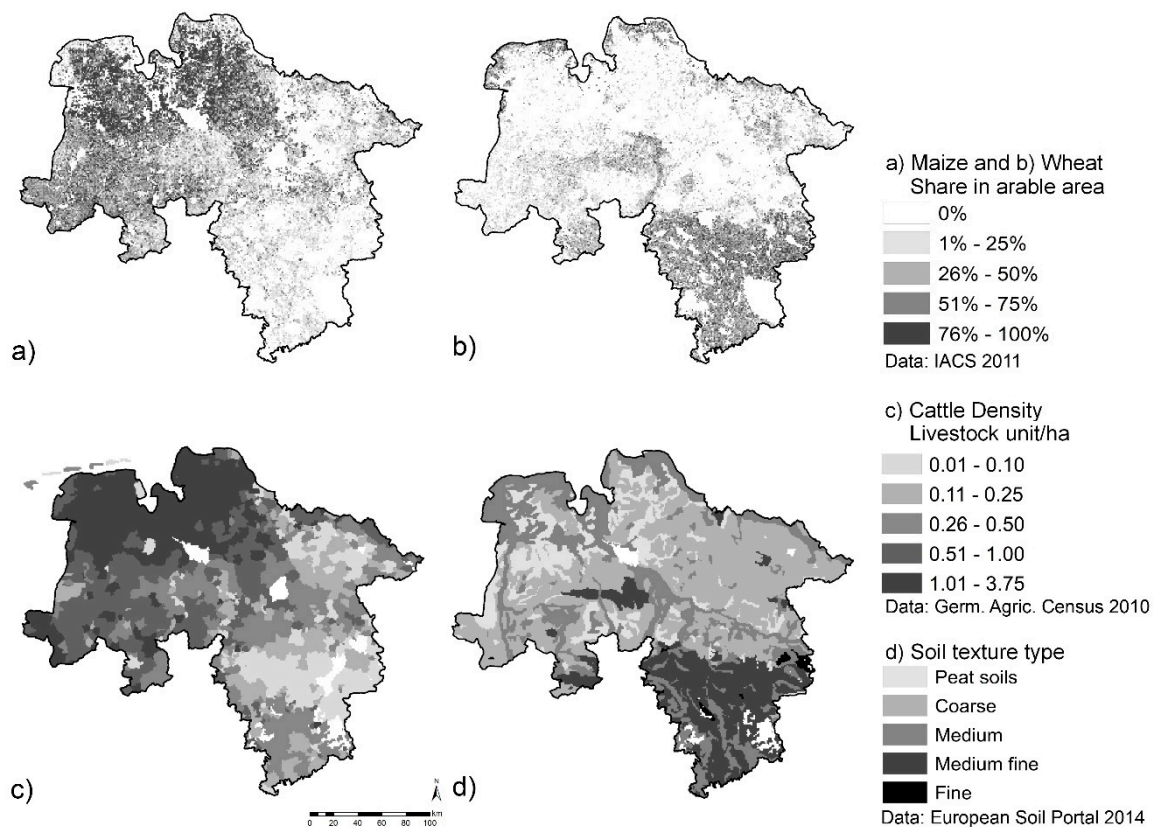


Figure 2. Exemplary mapping of the spatial distribution of two crops and two variables: (a) acreage of maize 2011; (b) acreage of winter wheat 2011; (c) cattle density per LAU-2 unit; (d) soil texture distribution.

2.3. Binary Logistic Regression (Field Scale)

Logistic regression is used instead of linear regression when the observed or measured response of interest is not continuous but binary to predict the likelihood of an event over the likelihood of nonoccurrence [30]. The cultivation of a crop on a specific field is such a binary event. Its likelihood under the occurrence of a specific site variable indicates the strength of its relationships to the cultivation site. If the site variable, e.g., cattle density, changes by one unit while all other variables stay stable, the likelihood of crop occurrence, e.g., maize, is increased or decreased by the resulting value of the regression equation. This value is larger or smaller than zero and can be larger than one. The two variables, arable farming potential and soil texture, have an ordinal scale and not a metric scale like all the other variables. Due to this, all characteristics of these two variables were analyzed separately (Table 3). The first characteristic, peat soil for soil texture and very low arable farming potential, had the role of the reference value, the same role that zero had for the other variables.

The nine main crops of Lower Saxony were chosen for analysis plus one group containing all spring cereals. For each of the ten crop categories, a binomial regression equation with a binary response variable, $y \in \{0, 1\}$, was defined to determine the probability of occurrence for each crop separately [31,32]. The regression analysis was performed using the software CRAN-R version 3.1.0 [33]. It uses a logarithmic function calculating the logit (π_i) for the ratio of the probability (P_{ij}) that a field (i) is cultivated with a specific crop (j) or not ($1 - P_{ij}$). Written in a logit equation as suggested by Fahrmeir et al. [34],

$$\pi_i = P(y_i = 1) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (1)$$

contains the linear predictor

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}. \quad (2)$$

The predictor (π_i) represents the logarithmic odds (log odds), while the coefficient (β_k) for this variable (x_{ik}) is the expected change in these log odds. While holding the corresponding predictor variables constant, a one unit increase of the predictor variable causes a change in the probability corresponding to the coefficient value for having the subject crop [29,35].

The likelihood ratio test with a null model for each crop resulted in a rejection of the null hypothesis for all crops. That means that the observed crop occurrence is more likely under the presented model than under the null model.

In contrast to the other variables, arable farming potential and soil texture are handled as factor variables. The coefficient of the first category acts as a reference category with a value of zero.

We inspected the correlation effects between the site variables to identify the rate of correlation between the variables, e.g., between cattle density and biotope density and between soil texture and arable farming potential (Table 2). These effects are immanent for variables that characterize ecological and spatial phenomena [36]. A high correlation of the variables is an expected effect and is therefore not considered in the equation. This decision is forced by the objective of the regression analysis, which is not used as a predicting model but as a method to characterize the relationship between the crops and the site conditions.

Table 2. Correlation matrix of the site variables used in the logistic regression model.

	A. F. Pot. ¹	Soil Texture	Slope	Precipit.	Biotope I ²	Farm Size	CattleD ³	PigPoulD ⁴	GrassL ⁵
A. F. Pot.	1								
Soil texture	0.617	1							
Slope	0.145	0.267	1						
Precipit.	-0.125	-0.093	0.117	1					
Biotope I	-0.503	-0.548	-0.227	0.350	1				
Farm Size	0.162	0.161	0.084	-0.421	-0.367	1			
CattleD	-0.439	-0.437	-0.190	0.501	0.665	-0.435	1		
PigPoulD	-0.207	-0.248	-0.161	0.248	0.227	-0.358	0.221	1	
GrassL	-0.242	-0.144	0.006	0.235	0.332	-0.154	0.388	-0.132	1

¹ arable farming potential; ² biotope index; ³ cattle density; ⁴ pig/poultry density; ⁵ grassland proportion.

The values of the correlation matrix indicate already the joint appearance of single variables (Table 2). The interaction of site variables creates regional patterns that are investigated in the next section employing a cluster analysis.

2.4. Cluster Analysis (Regional Scale)

A non-hierarchical k-means clustering with the Hartigan & Wong algorithm [37] was used to detect regional patterns of similarities for the site variables and for crops [38,39]. This was realized with the software CRAN-R version 3.1.0 [33,40]. The k-means clustering is a common method for identifying spatial units at the landscape scale [41–43]. It was used in this paper to identify spatial units with consistent properties. The crop clusters and the site clusters were then compared in their spatial concordance.

The optimal number of classes, k, was found by comparing results of multiple runs with different number of classes and visualizing the grade of clustering in a map [44]. The uncertainty of the initial random partition was adjusted by choosing the most frequent version of partition in ten runs. In a previous step, a z-transformation of all variable values standardized the very different scales to improve the comparability of the results. The cluster analysis generated five site clusters (S1, S2, S3, S4, and S5) and five crop clusters (C1, C2, C3, C4, and C5).

3. Results

3.1. Site Dependency at Field Scale

The intensity of the crop–site relationship is reflected in the coefficient value of the logistic regression analysis (Table 3). In general, the probability of crop appearance in the dataset depends more strongly on soil variables than on other site variables. Arable farming potential and soil texture show a high likelihood of determining the occurrence or non-occurrence of a crop but vary in their direction of relationship.

There are linear relations between crop and site variables in different directions, e.g., the increase in farming potential increases the probability of wheat but decreases the probability of forage cropping. Oilseed rape is an example of non-linear relations. It was cropped on fields with a middle and high arable farming potential with a much higher likelihood than on fields with an extremely high farming potential. The log odd results of sugar beet prove that soil variables can differ in their direction of influence and explain different aspects of the crop–soil relationship. The ambivalent relationship of sugar beet cropping and soil texture is determined by historical production quotas rather than by soil conditions. The variables farm size, pig/poultry density, grassland density, and biotope index have in general a low influence on probability. Each of the analyzed crops has its own specific character of site dependencies. Some crops such as maize and wheat have diverging site relations, while other crops, e.g., oilseed rape and winter barley, show similar probabilities under comparable site conditions. This result will be examined further in the next section by identifying regions with convergent characteristics.

Table 3. The log odds values describe the likelihood of crop occurrence when the variable value changes by one unit, while all other variable stay stable. The positive/negative sign shows the direction of relationship; *ref.* is the reference category of the ordinal variables.

Variables	SBeet	WO Rape	Triticale	Potato	Rye	WBarley	WWheat	SCereal	Forage	Maize
Arab. Farm. Pot.										
Extremely Low	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Very Low	−0.082	−0.142	−0.141	0.419	−0.359	−0.143	0.140	0.112	0.086	−0.097
Low	0.330	0.040	0.081	0.613	0.430	0.364	−0.116	0.133	−0.311	−0.187
Middle	0.729	0.484	−0.090	0.489	0.172	0.665	0.468	0.112	−0.564	−0.408
High	0.611	0.480	−0.508	−0.285	−0.530	0.547	0.831	0.283	−0.397	−0.726
Very High	1.025	0.440	−0.638	−0.014	−0.831	0.585	0.775	−0.122	−0.676	−0.693
Extremely High	1.136	−0.457	−1.198	−0.388	−1.796	0.354	0.763	−0.443	−1.000	−0.710
Soil Texture										
Peat soil	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Coarse	0.727	0.445	0.137	−0.106	0.498	0.493	0.120	0.007	−0.015	−0.203
Medium	0.285	0.960	−0.075	−0.659	−0.160	0.511	1.077	0.026	0.023	−0.348
Medium Fine	0.480	1.043	−0.600	−1.312	−1.019	0.651	1.186	−0.837	−0.181	−0.549
Fine	0.225	0.861	−0.117	−2.576	−0.093	0.454	1.170	−0.111	−0.158	−0.114
Slope	−0.040	0.230	−0.146	−0.513	−0.269	0.254	0.159	−0.330	0.130	−0.493
Precipitation	−0.198	0.019	−0.213	−0.113	−0.285	0.018	0.021	0.092	0.078	0.093
Biotope Index	−0.278	−0.165	0.036	−0.003	0.205	−0.047	−0.240	−0.067	−0.037	0.173
Farm size	0.067	−0.026	−0.213	0.094	0.141	−0.304	−0.055	−0.060	−0.031	0.043
Cattle Density	−0.498	−0.323	−0.201	−0.145	0.391	−0.176	−0.034	−0.145	0.091	−0.176
Pig/Poultry Density	−0.215	0.125	−0.033	−0.209	0.141	0.167	0.202	−0.209	−0.008	0.167
Grassland/a. area	−0.192	−0.230	0.056	0.084	0.058	−0.008	0.002	0.084	0.221	−0.008

3.2. Statistical Clustering and Spatial Projection

The nature of the relationship between site variables and the grown crop is examined in the regression analysis. With two statistical clustering processes—one for the site variables and one for the crop data—the characterization of the crop–site relationship will be transferred into a spatial projection to visualize overlapping spatial patterns. The k-means clustering of the site variables formed five continuous regions, which are characterized by their mean value in the defined clusters (Table 4).

Table 4. Mean values per cluster of the k-means clustering for site variables (S1, S2, S3, S4, and S5—corresponding map in Figure 3a). Values are z-standardized and represent how strong the standard deviation differs from the mean value ($\mu = 0.000$). A small value shows no significant difference from the mean value. The positive and negative values represent the direction of deviation from the mean value in that cluster.

	S1	S2	S3	S4	S5	Mean	SD	Unit
A. F. Pot.	−0.520	−0.290	−0.254	0.530	1.648	3.63	1.14	middle
Soil texture	−0.545	−0.390	−0.453	1.017	1.298	2.52	0.94	medium
Slope	−0.278	−0.279	−0.269	3.415	−0.279	1.09	0.39	(<8%)
Precipit.	0.422	−0.638	0.276	0.414	−0.246	774.42	75.96	mm
Biotope I	1.030	−0.363	−0.159	−0.607	−0.703	1.68	1.19	oD/pD
Farm Size	−0.415	0.321	−0.612	0.205	0.318	69.59	29.77	ha
CattleD	1.362	−0.511	0.122	−0.680	−0.665	0.64	0.53	LU/ha Agric. A.
PigPoulD	−0.285	−0.244	1.861	−0.423	−0.306	0.38	0.54	LU/ha Agric. A.
GrassL	0.408	−0.356	−0.564	−0.314	−0.504	0.21	0.22	ha/ha Agric. A.

S1 is characterized by a low farming potential and sandy soils, which correlate with a higher than average cattle density, biotope density, and grassland proportion. A quite different pattern of site conditions and crops characterizes S2: less humid climate and larger farm sizes. S3 has strong relations to farms that are smaller than average, with a specialization in pig and poultry farming. S4 and S5 have many similar characteristics but are distinguishable by the steeper slope and higher precipitation of the fifth cluster. The k-means clustering of the regional crop area proportion also resulted in five clusters (C1, C2, C3, C4, and C5). Each of these clusters has a characteristic composition of dominant crops (Table 5): The regional pattern of site conditions in C1 is related to a much higher than average maize proportion of the crop clustering process. C2 is the only cluster that is not dominated by maize or wheat but by a mixture of other crops, mainly rye and potato. C3 is characterized by a mixture of maize, triticale, and forage cropping. A composition of oilseed rape, winter wheat, and winter barley is the distinct feature of C4. The most obvious characteristic of C5 is a winter wheat proportion that is three times higher than the mean in Lower Saxony.

The transfer in a spatial projection of the clustering results reveals relationships between the site variables and the crop clustering on the one hand and distinctive differences on the other (Figure 3). Significant congruencies can be proved for S2 and C2, the potato–rye cluster. The second and third highest proportions of quadrants with spatial congruence were observed for S5 with C5 and for S1 with C1. The other two crop clusters have less than 50% spatial congruence with the site clusters.

Table 5. Mean values of the k-means clustering of crop data (corresponding map in Figure 3b). The values represent mean ratios of the crop area per arable area of the related quadrant. Values in bold are significantly higher than the mean value of the certain crop and are considered characteristic crops for the cluster type.

	C1	C2	C3	C4	C5	Mean	SD	Unit
SBeet	0.002	0.052	0.013	0.098	0.090	0.05	0.11	ha/ha Arab. A.
Potato	0.015	0.184	0.060	0.026	0.015	0.06	0.13	ha/ha Arab. A.
WO Rape	0.005	0.034	0.028	0.222	0.064	0.06	0.13	ha/ha Arab. A.
SCereal	0.018	0.094	0.040	0.030	0.021	0.04	0.10	ha/ha Arab. A.
Maize	0.816	0.120	0.463	0.092	0.070	0.34	0.31	ha/ha Arab. A.
Triticale	0.018	0.066	0.062	0.032	0.008	0.04	0.09	ha/ha Arab. A.
Rye	0.033	0.218	0.073	0.026	0.009	0.07	0.14	ha/ha Arab. A.
Forage	0.042	0.062	0.090	0.034	0.024	0.05	0.11	ha/ha Arab. A.
WWheat	0.021	0.044	0.074	0.228	0.621	0.21	0.25	ha/ha Arab. A.
WBarley	0.020	0.055	0.072	0.177	0.054	0.07	0.12	ha/ha Arab. A.
All others	0.008	0.071	0.025	0.035	0.022	0.03	0.08	ha/ha Arab. A.

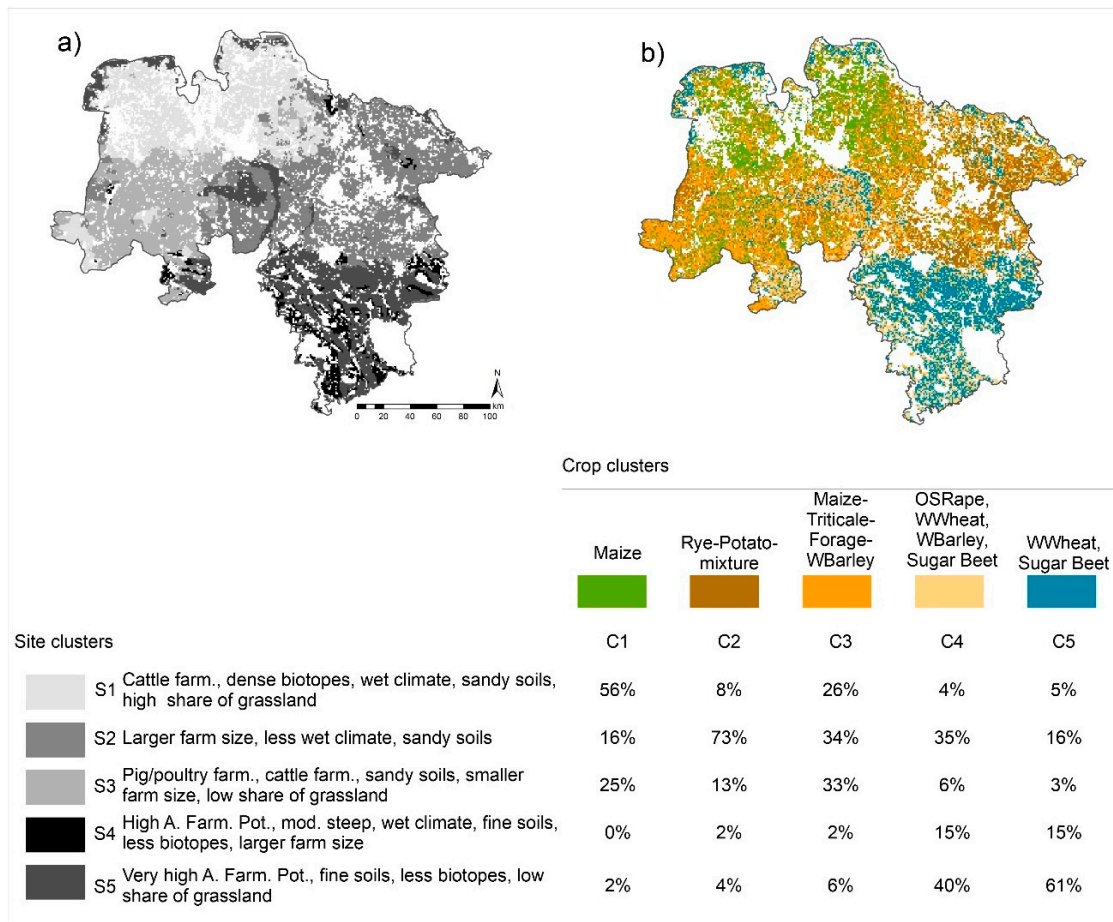


Figure 3. Spatial projection of the statistical k-means clustering results and the proportion of congruent areas in percent: (a) site clustering (S1–S5) and description, (b) crop clustering (C1–C5). Only quadrants ≥ 10 ha of arable area are included.

4. Discussion

4.1. General Discussion

Agricultural crops do not grow randomly at a specific site. Their spatial occurrence reflects the sum of farmers’ decisions as a product of site conditions and the political and economic framework. In recent decades, farmers, breeders, and the plant protection industry have focused on a few profitable crops. This was also a result of the market price development and the European agricultural policy and culture of yield-based subsidies. However, sustainable cropping systems rely on diverse cropping systems, among other factors [45,46]. In our study, we detected the strongest relationship of site variables, namely soil texture and arable farming potential, with crops at the most productive areas and the least productive areas. Crops like sugar beet, oilseed rape, and winter wheat are characterized by a high probability to be cropped on sites with a high arable farming potential. The spatial congruence of site clusters (e.g., S5) with crop clusters (e.g., C5) confirmed the regression result referring to the relationship of very high farming potential and the combined cropping of sugar beet and winter wheat. This was supplemented reversely by the significant absence of single crops on soils with high farming potential, such as rye and forage. Zimmermann and Britz concluded from their study of the use of agri-environmental measures by farmers in the EU that those measures were most likely found on less productive sites during 2000–2009 [47]. The recent CAP 2014–2020 includes agri-environmental measures such as crop diversification as obligatory requirements for initial pillar payments. Recent studies concerning the impact assessment of the CAP 2014–2020 show contrary

results: a limited environmental impact of the new greening rules [48] and strong effects on the farmland use in high-intensive agricultural regions [49].

The spring cereals and forage crops are characterized by a weak crop–site relationship as well as maize and winter wheat, which are the main arable crops, with acreage of 32% and 21% of the arable area, respectively [17]. The economical preference, the high tolerance for the combination with other crops, as well as the tolerance to short intervals in the rotation result in a dense cropping of maize and winter wheat in space and time [7,16]. Nevertheless, each of these two crops dominate regions that are characterized by contrasting conditions concerning the soil texture and arable farming potential, the slope, as well as the grassland and livestock density.

The relationship of maize cropping and specific combinations of site conditions is strongly determined by the cultivation practice for this crop. Rotations with maize are characterized by very dense cropping up to permanent cropping on the one hand and maize as one part of very diverse rotations on the other hand [16]. These rotation phenomena are common in regions with different site characteristics and geography. This is further confirmed by the result that the spatial congruency of site clusters and the crop cluster with dense maize cultivation (Figure 3, C1) was clearly distinguishable from their relationship to the cluster of maize cultivation in combination with other crops (C3). Whether maize cropping is allocated to C1 or C3 apparently has consequences in terms of ecosystem effects. While the spatially dense maize cultivation can have negative impacts on ecosystem services, the maize cultivation within the more diverse system of C3 can have a positive impact [50]. As the identified areas with high maize acreage are only partly explainable by livestock farming, they may correspond with other factors such as biogas production that are not represented by the explanatory data. The area cultivated with maize increased in Northwestern Germany from 2005 to 2011 by 67% [17]. The widespread cultivation of maize is an effect of the expansion of biogas production after the implementation of the national renewable energy law [11,26].

4.2. Reflections on the Methods Used

For a realistic analysis of regional crop–site relationships, the use of crop information at field scale is essential [51,52]. The yearly updated database of the LPIS is a valuable data source for agronomical and environmental analysis. The LPIS data have a high spatial resolution that allows for a precise intersection with other spatial information and yields precise answers to field scale questions. Area-wide crop information on field scale could also be useful for the validation of crop growth models especially for areas with a large diversity of cropping systems [53,54] and for modeling procedures when information concerning cropping practices is needed [52,55,56]. The scientific use of LPIS data, e.g., for the prediction of the crop yield or for projecting changes in agricultural land use practice, is increasingly becoming important [55,57–60].

Two statistical methods were applied for the analysis of the crop–site relationship: logistic regression analysis and k-means clustering, visualized by a map projection. Both approaches concern different levels and aspects of the relationship. The level of spatial similarities between the crop clusters and the site clusters supplemented the results of the logistic regression analysis and elucidated in parts the fuzzy picture of direct relationships. This underpins the need to include cropping patterns instead of single crop information in modeling approaches.

Not all the chosen variables have the expected potential to explain crop–site dependencies. The low influence of farm size, pig/poultry density, grassland density, and biotope index on the probability of crop cultivation in comparison with the soil variables can be explained by their low tendency to form spatial patterns or clusters in Lower Saxony, which is reflected in the high standard deviation values. In our analysis, we focused on environmental variables instead of economic variables because most of the studies concerning the cropping-plan decision making process of farmers consider economical and sociological drivers [3,61]. However, we could show the still high potential of soil variables as drivers for decision making, which is also confirmed by a study of Peltonen-Sainio et al. [62]. This study

exposed also field size as a potent driver variable, which was not concerned in our study, because it is indirectly included in the biotope index.

The crop clustering process resulted in a much more scattered picture than the site cluster projection. The latter is based on variables with different spatial resolution, ranging from the smaller scaled LAU-2 data to 1 km² resolved raster data that gave a different degree of precision. However, the reason for the different degree of spatial clustering is not only caused by the spatial resolution of the data sources. While the site clusters are products of natural conditions, the crop clusters are a result of both site conditions and socio-economic factors, e.g., market prices and subsidies. This supports the flexibility of the farmers in crop choice and therefore the fragmentation of crop clusters, especially in the center of Lower Saxony (#3, 5, 6 referring to Figure 1), with medium arable farming potential, sandy soils, and a higher variation of farm types in this area than in other regions.

5. Conclusions

The relationship of site conditions and crop cultivation at the field scale is generally weak but detectable for some crops. One reason is that modern cropping practice enables the farmer to override the relationship of crop and site to a large extent. However, this does not apply to all crop–site relationships. In arable regions with productive soils, the crop–site relationship is stronger. This comes along with specialization of the farming systems to a few cash crops, mainly the most profitable crops like sugar beet and winter wheat. On the other hand, a stronger relationship of crop and site at the regional scale was also detected for clusters with less productive soils and the crop cluster with dominant maize cultivation. Economic reasons and policy-based incentives, such as support for bioenergy crops may have enforced this allocation. Farming practice and agricultural policy must face the chances but also the risks of this development.

In regions with less fertile soils and mixed farming structure, farmers' cultivation practices are much more diverse. These site clusters are not dominated by one crop cluster but by multiple crop clusters with a number of dominating crops. The likelihood of crop rotation diversification is higher in these multiform regions, but in rather monotonous regions diversification efforts are crucial.

Author Contributions: Conceptualization: S.S. and H.-H.S.; methodology: S.S.; formal analysis: S.S.; investigation: S.S.; resources: H.-H.S.; data curation: S.S.; writing—original draft preparation: S.S.; writing—review and editing: H.-H.S. and J.I.; visualization: S.S.; supervision: H.-H.S. and J.I.; project administration: H.-H.S.; funding acquisition: H.-H.S.

Funding: This research was funded by Fachagentur Nachwachsende Rohstoffe, a promoter of the German Federal Ministry of Food and Agriculture, grant number 12NR1092 FNR.

Acknowledgments: We are grateful to the Ministry for Human Nutrition, Agriculture, Consumer Protection and Rural Development of Niedersachsen (Lower Saxony), which provided administrative data. Developing indicators for land use and landscape was supported by the Ministry for Science and Culture of Niedersachsen within the network KLIF—climate impact and adaptation research in Niedersachsen. Data analysis was supported by the Federal Ministry of Food and Agriculture. We thank the anonymous reviewers for their careful reading of our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tilman, D.; Cassman, K.G.; Matson, P.A.; Naylor, R.; Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **2002**, *418*, 671–677. [[CrossRef](#)] [[PubMed](#)]
2. Van Zanten, B.T.; Verburg, P.H.; Espinosa, M.; Gomez-y-Paloma, S.; Galimberti, G.; Kantelhardt, J.; Kapfer, M.; Lefebvre, M.; Manrique, R.; Piorr, A.; et al. European agricultural landscapes, common agricultural policy and ecosystem services: A review. *Agron. Sustain. Dev.* **2014**, *34*, 309–325. [[CrossRef](#)]
3. Dury, J.; Garcia, F.; Reynaud, A.; Bergez, J.-E. Cropping-plan decision-making on irrigated crop farms: A spatio-temporal analysis. *Eur. J. Agron.* **2013**, *50*, 1–10. [[CrossRef](#)]

4. Aouadi, N.; Aubertot, J.N.; Caneill, J.; Munier-Jolain, N. Analyzing the impact of the farming context and environmental factors on cropping systems: A regional case study in Burgundy. *Eur. J. Agron.* **2015**, *66*, 21–29. [[CrossRef](#)]
5. Troost, C.; Walter, T.; Berger, T. Climate, energy and environmental policies in agriculture: Simulating likely farmer responses in Southwest Germany. *Land Use Policy* **2015**, *46*, 50–64. [[CrossRef](#)]
6. Bakker, M.M.; Hatna, E.; Kuhlman, T.; Mächer, C.A. Changing environmental characteristics of European cropland. *Agric. Syst.* **2011**, *104*, 522–532. [[CrossRef](#)]
7. Steinmann, H.-H.; Dobers, S. Spatio-temporal analysis of crop rotations and crop sequence patterns in Northern Germany: Potential implications on plant health and crop protection. *J. Plant Dis. Prot.* **2013**, *120*, 85–94. [[CrossRef](#)]
8. Antrop, M. Why landscapes of the past are important for the future. *Landsc. Urban Plan.* **2005**, *70*, 21–34. [[CrossRef](#)]
9. Bakker, M.M.; Sonneveld, M.P.W.; Brookhuis, B.; Kuhlman, T. Trends in soil–land-use relationships in the Netherlands between 1900 and 1990. *Agric. Ecosyst. Environ.* **2013**, *181*, 134–143. [[CrossRef](#)]
10. Viaggi, D.; Gomez, Y.; Paloma, S.; Mishra, A.; Raggi, M. The role of the EU Common Agricultural Policy: Assessing multiple effects in alternative policy scenarios. *Land Use Policy* **2013**, *31*, 99–101. [[CrossRef](#)]
11. EEG—Erneuerbare-Energien-Gesetz (Renewable Energies Act) of 21 July 2004 (Federal Law Gazette I p. 1918), last amended by Art. 1 Act of 7 November 2006 (Federal Law Gazette I p. 2550). Available online: https://www.gesetze-im-internet.de/eeg_2014/_14.html (accessed on 10 March 2019).
12. The Organisation for Economic Co-operation and Development (OECD). *Analysis of the 2003 CAP Reform*; OECD: Paris, France, 2004. Available online: <https://www.oecd.org/tad/32039793.pdf> (accessed on 20 March 2019).
13. Tzanopoulos, J.; Jones, P.J.; Mortimer, S.R. The implications of the 2003 Common Agricultural Policy reforms for land use and landscape quality in England. *Landsc. Urban Plan.* **2012**, *108*, 39–48. [[CrossRef](#)]
14. Trubins, R. Land-use change in southern Sweden: Before and after decoupling. *Land Use Policy* **2013**, *33*, 161–169. [[CrossRef](#)]
15. European Parliament and Council. Regulation (EU) No 1307/2013 Establishing Rules for Direct Payments to Farmers under Support Schemes within the Framework of the Common Agricultural Policy and Repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009. Official Journal of the European Union (L 347/608–669). Available online: <http://data.europa.eu/eli/reg/2013/1307/oj> (accessed on 16 March 2014).
16. Stein, S.; Steinmann, H.-H. Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems—A case study from Central Europe. *Eur. J. Agron.* **2018**, *92*, 30–40. [[CrossRef](#)]
17. Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz (NMELV). *Ergänzungen zur Broschüre: Die Niedersächsische Landwirtschaft in Zahlen 2011 (Stand: November 2013)*; NMELV: Hannover, Germany, 2013. Available online: http://www.ml.niedersachsen.de/download/83668/Die_niedersaechsische_Landwirtschaft_in_Zahlen_2011_-_Ergaenzung_11-2013.pdf (accessed on 4 May 2016).
18. Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN). *Naturräumliche Regionen in Niedersachsen*; NLWKN: Norden, Germany, 2010. Available online: https://www.nlwkn.niedersachsen.de/naturschutz/biotopschutz/naturraeumliche_regionen/naturraeumliche-regionen-niedersachsens-93476.html (accessed on 23 April 2014).
19. European Parliament and Council. Regulation (EC) No 1782/2003 Establishing Common Rules for Direct Support Schemes under the Common Agricultural Policy and Establishing Certain Support Schemes for Farmers. Official Journal of the European Union (L 270/1–69). Available online: <http://data.europa.eu/eli/reg/2003/1782/oj> (accessed on 23 March 2014).
20. SLA (Niedersächsisches Servicezentrum für Landentwicklung und Agrarförderung). *Digitale Feldblockkarte Niedersachsens (DFN)*; SLA: Hannover, Germany, 2011.
21. Niedersächsisches Ministerium für Ernährung, Landwirtschaft, Verbraucherschutz und Landesentwicklung. *Die niedersächsische Landwirtschaft in Zahlen*; NMELV: Hannover, Germany, 2011. Available online: https://www.ml.niedersachsen.de/startseite/aktuelles_veranstaltungen_veroeffentlichungen/die-niedersaechsische-landwirtschaft-in-zahlen-121348.html (accessed on 5 June 2017).

22. LBEG (Landesamt für Bergbau, Energie und Geologie). *Bodenübersichtskarte 1:50 000 (BÜK 50) von Niedersachsen, Standortbezogenes Natürliches Ackerbauliches Ertragspotenzial*; LBEG: Hannover, Germany, 1996. Available online: https://www.lbeg.niedersachsen.de/karten_daten_publicationen/karten_daten/boden/bodenkarten/bodenkundliche_uebersichtskarte_150000/bodenubersichtskarte-im-mastab-1--50-000-buek50-654.html (accessed on 12 March 2013).
23. ESDAC (European Soil Data Centre). *The European Soil Database v2 Raster Library 1kmx1km*; European Commission and the European Soil Bureau Network: Ispra, Italy, 2004. Available online: <https://esdac.jrc.ec.eu/content/european-soil-database-v2-raster-library-1kmx1km> (accessed on 3 May 2014).
24. DWD (Deutscher Wetterdienst). *Multi-Annual Precipitation Sum (1981–2010)*; DWD: Offenbach, Germany, 2014. Available online: <https://cdc.dwd.de/portal/> (accessed on 6 March 2014).
25. Julius Kühn-Institut (JKI). *Verzeichnis der Regionalen Kleinstrukturen des Landes Niedersachsen auf Gemeindebasis*; JKI: Kleinmachnow, Germany, 2004. Available online: <https://www.julius-kuehn.de/media/Institute/SF/RaeumlAnalyseModellierung/PDF/Niedersachsen.pdf> (accessed on 14 June 2014).
26. LSKN (Landesbetrieb für Statistik und Kommunikationstechnologie Niedersachsen). *Statistische Berichte Niedersachsen, Landwirtschaftszählung 2010, Brochure 1/A (Landuse) and 4 (Livestock)*; LSKN: Hannover, Germany, 2012. Available online: <https://www.statistik.niedersachsen.de/download/75470> (accessed on 12 May 2014).
27. Gutsche, V.; Enzian, S. Quantifizierung der Ausstattung einer Landschaft mit naturbetonten terrestrischen Biotopen auf der Basis digitaler topographischer Daten. *Nachrichtenbl. Deut. Pflanzenschutzd.* **2002**, *54*, 92–101.
28. Richter, U.; Eckelmann, W. Das Ertragspotential ackerbaulich genutzter Standorte in Niedersachsen—Beispiel einer Auswertungsmethode im Niedersächsischen Bodeninformationssystem NIBIS. *Geol. Jb.* **1993**, *F27*, 197–205.
29. ESRI. ArcGIS 9.2 Desktop Help: Geometrical Interval 2007. Available online: http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical_interval (accessed on 22 March 2015).
30. Tarpey, T. Generalized Linear Models 2012. (last update at: 29 May 2012). Available online: <http://www.wright.edu/~thaddeus.tarpey/ES714glm.pdf> (accessed on 10 June 2015).
31. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; Wiley: New York, NY, USA, 2000.
32. Menard, S. *Applied Logistic Regression Analysis*, 2nd ed.; Sage University Paper: Indore, India, 1995.
33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013. Available online: <http://www.R-project.org> (accessed on 11 February 2015).
34. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression: Models, Methods and Applications*; Springer: Berlin, Germany, 2013.
35. Kleinn, C.; Jovel, J.; Hilje, L. A model for assessing the effect of distance on disease spread in crop fields. *Crop Prot.* **1999**, *18*, 609–617. [[CrossRef](#)]
36. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 3rd ed.; Wiley: New York, NY, USA, 1998.
37. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser.* **1979**, *28*, 100–108. [[CrossRef](#)]
38. Hartigan, J.A. *Clustering Algorithms*; Wiley Series in Probability and Mathematical Statistics; John Wiley & Sons: Hoboken, NJ, USA, 1975.
39. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer: Berlin, Germany, 2001.
40. R Documentation. *K-Means Clustering*; Package ‘Stats’ Version 3.3.0; R Foundation for Statistical Computing: Vienna, Austria, 2015. Available online: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html> (accessed on 12 February 2015).
41. Schmidt, K.; Behrens, T.; Friedrich, K.; Scholten, T. A method to generate soilscares from soil maps. *J. Plant Nutr. Soil Sci.* **2010**, *173*, 163–172. [[CrossRef](#)]
42. Carvalho, M.J.; Melo-Goncalves, P.; Teixeira, J.C.; Rocha, A. Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. *Phys. Chem. Earth Parts A/B/C* **2016**, *94*, 22–28. [[CrossRef](#)]
43. Javadi, S.; Hashemy, S.M.; Mohammadi, K.; Howard, K.W.F.; Neshat, A. Classification of aquifer vulnerability using K-means cluster analysis. *J. Hydrol.* **2017**, *549*, 27–37. [[CrossRef](#)]
44. Morissette, L.; Chartier, S. The k-means clustering technique: General considerations and implementation in Mathematica. *Quant. Methods Psychol.* **2013**, *9*, 15–24. [[CrossRef](#)]
45. Smith, R.G.; Gross, K.L.; Robertson, G.P. Effects of Cropping Diversity on Agroecosystem Function: Crop Yield Response. *Ecosystems* **2008**, *11*, 355–366. [[CrossRef](#)]

46. Storkey, J.; Bruce, T.J.A.; McMillan, V.E.; Neve, P. Chapter 12—The future of sustainable crop protection relies on increased diversity of cropping systems and landscapes. In *Agroecosystem Diversity*; Academic Press: Cambridge, MA, USA, 2019; pp. 199–209.
47. Zimmermann, A.; Britz, W. European farm's participation in agri-environmental measures. *Land Use Policy* **2016**, *50*, 214–218. [[CrossRef](#)]
48. Cortignani, R.; Dono, G. CAP's environmental policy and land use in arable farms: An impacts assessment of greening practices changes in Taly. *Sci. Total Environ.* **2019**, *647*, 516–524. [[CrossRef](#)] [[PubMed](#)]
49. Bertoni, D.; Aletti, G.; Ferrandi, G.; Micheletti, A.; Cavicchioli, D. Farmland use transition after the CAP greening: A preliminary analysis using Markov chains approach. *Land Use Policy* **2018**, *79*, 789–800. [[CrossRef](#)]
50. Albert, C.; Hermes, J.; Neuendorf, F.; von Haaren, C.; Rode, M. Assessing and Governing Ecosystem Services Trade-Offs in Agrarian Landscapes: The Case of Biogas. *Land* **2016**, *5*, 1. [[CrossRef](#)]
51. Leteinturier, B.; Herman, J.L.; de Longueville, F.; Quintin, L.; Oger, R. Adaptation of a crop sequence indicator based on a land parcel management system. *Agric. Ecosyst. Environ.* **2006**, *112*, 324–334. [[CrossRef](#)]
52. Schönhart, M.; Schmid, E.; Schneider, U.A. CropRota—A crop rotation model to support integrated land use assessments. *Eur. J. Agron.* **2011**, *34*, 263–277. [[CrossRef](#)]
53. Nendel, C.; Wieland, R.; Mirschel, W.; Specka, X.; Guddat, C.; Kersebaum, K.C. Simulating regional winter wheat yields using input data of different spatial resolution. *Field Crops Res.* **2013**, *145*, 67–77. [[CrossRef](#)]
54. Van Wart, J.; Kersebaum, K.C.; Peng, S.; Milner, M.; Cassman, K.G. A protocol for estimating crop yield potential at regional to national scales. *Field Crops Res.* **2013**, *143*, 34–43. [[CrossRef](#)]
55. Mitter, H.; Heumesser, C.; Schmid, E. Spatial modeling of robust crop production portfolios to assess agricultural vulnerability and adaptation to climate change. *Land Use Policy* **2015**, *46*, 75–90. [[CrossRef](#)]
56. Xiao, Y.; Mignolet, C.; Mari, J.-F.; Benoît, M. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France. *Comput. Electron. Agric.* **2014**, *102*, 51–63. [[CrossRef](#)]
57. Tychon, B.; Buffet, D.; Dehem, D.; Eerens, H.; Oger, R. The Belgium crop growth monitoring system. In Proceedings of the 2nd International Symposium: Modelling Cropping Systems, Florence, Italy, 16–18 July 2001.
58. Kandziora, M.; Burkhard, B.; Müller, F. Mapping provisioning ecosystem services at the local scale using data of varying spatial and temporal resolution. *Ecosyst. Ser.* **2013**, *4*, 47–59. [[CrossRef](#)]
59. Andersson, G.K.S.; Ekroos, J.; Stjernman, M.; Rundlöf, M.; Smith, H.G. Effects of farming intensity, crop rotation and landscape heterogeneity on field bean pollination. *Agric. Ecosyst. Environ.* **2014**, *184*, 145–148. [[CrossRef](#)]
60. Levavasseur, F.; Martin, P.; Bouty, C.; Barbottin, A.; Bretagnolle, V.; Théron, O.; Scheurer, O. RPG Explorer: A new tool to ease the analysis of agricultural landscape dynamics with the Land Parcel Identification System. *Comput. Electron. Agric.* **2016**, *127*, 541–552. [[CrossRef](#)]
61. Huber, R.; Bakker, M.; Balmann, A.; Berger, T.; Bithell, M.; Brown, C.; Gret-Regamey, A.; Xiong, H.; Le, Q.B.; Mack, G.; et al. Representation of decision-making in European agricultural agent-based models. *Agric. Syst.* **2018**, *167*, 143–160. [[CrossRef](#)]
62. Peltonen-Sainio, P.; Jauhiainen, L.; Sorvali, J.; Laurila, H.; Rajala, A. Field characteristics driving farm-scale decision-making on land allocation to primary crops in high latitude conditions. *Land Use Policy* **2018**, *71*, 49–59. [[CrossRef](#)]

