Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Neighbourhood Effects in Farm Diversification

A Dutch Case Study

Willemijn Vroege

ETH Zurich

Agricultural Economics and Policy

&

Wageningen University and Research

Wageningen Economic Research

Master thesis

Neighbourhood Effects in Farm Diversification

February 2017

The Hague, Netherlands

Supervisors:

Wim Heijman, WUR

Nico Polman, WER

Robert Finger, ETH Zurich

Submitted by:

Willemijn Vroege

ETH: 12-753-000

WUR: 94050519080

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eigenständigkeitserklärung

Die unterzeichnete Eigenständigkeitserklärung ist Bestandteil jeder während des Studiums verfassten Semester-, Bachelor- und Master-Arbeit oder anderen Abschlussarbeit (auch der jeweils elektronischen Version).

Die Dozentinnen und Dozenten können auch für andere bei ihnen verfasste schriftliche Arbeiten eine Eigenständigkeitserklärung verlangen.

Ich bestätige, die vorliegende Arbeit selbständig und in eigenen Worten verfasst zu haben. Davon ausgenommen sind sprachliche und inhaltliche Korrekturvorschläge durch die Betreuer und Betreuerinnen der Arbeit.

Titel der Arbeit (in Druckschrift):

Neighbourhood Effects in Farm Diversification

Verfasst von (in Druckschrift):
Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich.

Name(n):	Vorname(n):
Vroege	Willemijn

Ich bestätige mit meiner Unterschrift:
- Ich habe keine im Merkblatt „Zitier-Knigge" beschriebene Form des Plagiats begangen.
- Ich habe alle Methoden, Daten und Arbeitsabläufe wahrheitsgetreu dokumentiert.
- Ich habe keine Daten manipuliert.
- Ich habe alle Personen erwähnt, welche die Arbeit wesentlich unterstützt haben.

Ich nehme zur Kenntnis, dass die Arbeit mit elektronischen Hilfsmitteln auf Plagiate überprüft werden kann.

Ort, Datum	Unterschrift(en)
Amersfoort, 15.02.2017	

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich. Durch die Unterschriften bürgen sie gemeinsam für den gesamten Inhalt dieser schriftlichen Arbeit.

Table of Contents

List of Figures

List of Tables

Neighbourhood Effects in Farm Diversification

Abstract

For the support of further developments in farm diversification, it is important to understand the characteristics of diversified farms. This research investigates neighbourhood effects between Dutch farms to better understand the emergence of diversified agriculture. The question arises whether the clustered pattern of emergence of diversification can be fully explained by spatial heterogeneity or that neighbourhood effects are relevant as well. Positive neighbourhood effects may result from social learning and from the existence of common pool resources within communities. With a spatial autoregressive probit model estimating model parameters with Bayesian techniques incorporating Markov Chain Monte Carlo (MCMC) sampling, we estimate the importance of neighbourhood effects in five different diversification strategies. We identified neighbourhood effects within diversification strategies, interactions between different diversification activities and influences of different farm types in the neighbourhood.

Keywords

farm diversification; spatial regression; neighbourhood effects; the Netherlands

Introduction

Farm diversification is one of the European Union's priorities for rural development. Due to the results of many studies (e.g. Benjamin, 1994; Barbieri and Mahoney, 2009; Heringa et al., 2013), it has been recognized that agricultural diversification is "contributing to income diversification of agricultural holding[s]" and supports the development of rural areas by "the creation and development of new economic activity in the form of new farms, […] the provision of services to agriculture and forestry and [the emergence of] activities related to health care, social integration and tourism" (Augère-Granier, 2016).
For the support of further developments in farm diversification, it is important to understand the characteristics of diversified farms. This research investigates neighbourhood effects between Dutch farms to better understand the emergence of diversified agriculture. Neighbourhood effects are constructed mechanisms within neighbourhoods which influence the behaviour of the individuals who live in them (Small and Newman, 2001). Multiple studies have highlighted the importance of physical spatial heterogeneity for the emergence of farm diversification (Ilbery, 1991; Meert et al., 2005; Jongeneel et al., 2008; Pfeifer et al., 2009; Barbieri and Mahoney, 2009; Lange et al., 2013; Meraner et al., 2015; Hassink et al, 2016). However, as diversified farms are found to be clustered (Meraner et al., 2015), the question arises whether this pattern of diversification can be fully explained by spatial heterogeneity or that neighbourhood effects are also relevant for the adoption of this farm strategy (Meraner et al., 2015). Clusters imply that spatially closer observations are more similar than spatially further observations[1]. In the case of clustering, spatial independence between observations is no longer fulfilled, leading to spatial autocorrelation. Autocorrelation of diversification activities can occur due to heterogeneity in locations as suggested in earlier research on farm diversification. Three locational variables have been included in this research. Firstly, soil properties are found to influence diversification decisions significantly (Pfeifer et al., 2009). Secondly, it was found that at sites closer to attractive landscapes, farm diversification is observed more often (Walford, 2001; Pfeifer et al., 2009; Lange et al., 2013; Hassink et al., 2016), Thirdly, the closeness of a farm to an urban centre has some influence, particularly on on-farm sale activities (Ilbery, 1991; Meraner et al., 2015; Hassink et al., 2016). Yet, the relationship depends on the definition of an urban centre and on the type of diversification activity (Barbieri and Mahoney, 2009; Meraner et al., 2015).
Neighbourhood effects lead to spatial dependence of decision-making. These effects can be either positive (also referred to as spatial spill-overs and clustering) or negative. With a visual analysis, Meraner et al. (2015) found positive neighbourhood effects on diversification. Positive neighbourhood effects may result from social learning (Munshi, 2004) and from the existence of common pool resources within communities (Ostrom, 2002). The concept of social learning describes interactions between farmers which influence the adoption of new technologies (Case, 1992). Communication with a diversified neighbour is an additional source of information for a farmer (Munshi, 2004), thus reducing information costs of a new farm strategy. Rutten and Boekema (2007) argue that firms in a spatial cluster create new resources by exchanging tacit knowledge. Ostrom (2002) refers to these resources produced through private activities of actors in a region as 'common pool resources'. When the neighbourhood produces a common pool resource, a certain activity might be more valuable than in other regions, which makes the adoption of this activity more favourable (Ostrom, 2002). This effect might be enlarged by local farm collaborations (Fischer and Ypma, 2012). Negative neighbourhood effects have been found to be relevant for developments in agriculture (Storm et al., 2015) and are therefore assumed to also exist for diversification. It is for example reported that if there is more competition, the supply of touristic accommodations is diversifying (Van der Meulen et al., 2014). Moreover, unattractive neighbourhood characteristics are found to influence agricultural decision-making (Lapple and Kelley, 2015). For example, being situated next to farm types which reduce the attractiveness of the neighbourhood, like intensive

[1] Tobler's first law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

livestock farms, may reduce the utility from agritourism. Social acceptance in the neighbourhood of an agricultural practice can as well contribute to neighbourhood effects, both positively and negatively (Home et al., 2013; Wollni et al., 2014).

Neighbourhood effects have to our knowledge not yet been included in the spatial analysis of diversified agriculture. Spatial diversification models (Walford, 2001; Pfeifer et al., 2009; Meraner et al., 2015; Hassink et al., 2016) have included factors explaining spatial heterogeneity (soil type, closeness to national parks, urbanization), but not neighbourhood effects. In this research, neighbourhood effects on diversification decisions are estimated by quantifying the importance of neighbourhood characteristics for the decision to start a diversification activity. For a case study of farms in the Netherlands, a spatial autoregressive probit regression is used.

The paper is organized as follows. First, we give some theoretical background including a definition of farm diversification, a description of the types of farm diversification relevant in the Netherlands, a brief explanation of neighbourhood effects expected for diversification activities in the Netherlands, and an overview of earlier studies on neighbourhood effects in agricultural literature. This is followed by the methodology used in this research. Hereafter, the data and independent variables are introduced. Next, results are presented and discussed, and conclusions are drawn.

Theoretical background

Farm diversification

In recent literature (Polman et al., 2010; Meraner et al., 2015; Hassink et al., 2016) three terms are used to describe slightly different concepts relating to the broader than traditional use of a farm's resources, i. e. pluriactivity, multifunctionality and diversification. Diversification refers to the farm expanding its activities on the farm itself (Meraner et al., (2015). Ilbery (1991) argued that diversification can be defined as "the reallocation and recombination of farm resources away from its original farming activity to generate another form of income". Therewith, it can be separated from pluriactivity, which refers to farm households combining agricultural and non-agricultural activities, with a focus on family members carrying out on- and off-farm labour (Polman et al., 2010; Hassink et al., 2016) and on multifunctionality which is "a characteristic of the agricultural system in a certain rural area or region, and not necessarily of an individual farm". In this way, pluriactivity describes the internalisation of external effects of farms (Polman et al., 2010). Ilbery (1991) suggested to classify different types of diversification into "agricultural" and "structural" diversification. Agricultural diversification includes activities with a focus away from agricultural production (defined as the growing of unconventional crops, the holding of unconventional animals and organic agriculture). Structural diversification includes tourism, adding value to farm enterprises, and passive diversification, referring also to the leasing of buildings and land. Van der Ploeg and Roep (2003) add another form of diversification to Ilbery's (1991) definition. These authors classify diversified agriculture into deepening (which refers to Ilbery's (1991) concept of agricultural diversification), broadening (which can be compared with Ilbery's (1991) concept of structural diversification) and additionally regrounding. Regrounding activities include pluriactivity as well as leasing of buildings and land. This research focuses on diversification activities which can be classified into Ilbery's (1991) concept of structural diversification and can further be delimited to Van der Ploeg and Roep's concept of broadening activities. Therefore, farm diversification is here defined as "farms actively carrying out broadening activities".

Diversification in Dutch agriculture

From 2008 until 2012, a Taskforce for multifunctional agriculture (Taskforce Multifunctionele Landbouw) was set up by the Dutch ministry of agriculture, nature and food quality (Ministerie voor Landbouw, Natuur en Voedselkwaliteit) to give temporary governmental support to the diversified agricultural sector (Fischer and Ypma, 2012).[2] The Taskforce distinguished six different broadening diversification activities in the Netherlands: on-farm sale, nature conservation, tourism and recreation, care farming, farm education and child care. The activities “on-farm processing” and “aquaculture” were not captured by the Taskforce but are nevertheless included in the Dutch agricultural census. The activities aquaculture, childcare and education are excluded from the further research because the total number of farms is very small (aquaculture) or the Taskforce did not highlight any spatial dependencies (education and child care). 19% of all Dutch farms are carrying out at least one of the five remaining diversification activities and are therefore classified as diversified farms in this research.

Farm characteristics influencing farm diversification in the Netherlands have been discussed intensively. Characteristics found to be related to diversification in Dutch agriculture include structural and locational farm characteristics. With respect to *structural farm characteristics*, Meraner et al. (2015) found that the age of the farmer has a significant influence on the decision to diversify. It appears that younger farmers tend to diversify more often than older farmers, since risk reduction is more important for them and they have a longer time horizon for their investments (Barbieri and Mahoney, 2009). It is also found that farms with more available workforce have a greater chance to diversify “as larger farms seek employment opportunities for family members on the farm” (Meraner et al., 2015; Weltin et al., 2017). Moreover, a farm’s specialisation is found to be relevant for the decision to diversify (Jongeneel et al., 2008; Meraner et al., 2015; Weltin et al., 2017). Both pasture-based and arable farms diversify less in on-farm sale and -processing activities than other farm types. These activities are probably too time-consuming for this type of farm (Jongeneel et al., 2008; Meraner et al., 2015). Pasture-based farms participate relatively often in nature conservation activities as these farms have more land available (Jongeneel et al., 2008; Meraner et al., 2015). Mixed farms are the most likely to diversify (Meraner et al., 2015).

Locational farm characteristics discussed in literature about diversification in the Netherlands include soil properties, closeness to nature and closeness to urban centres. Firstly, soil properties are found to influence diversification decisions significantly (Pfeifer et al., 2009; Meraner et al., 2015). Diversified activities are found less on farms located on peat soils, with the exception of nature-conservation activities which are more often found on peat soils (Meraner et al., 2015). Pfeifer et al. (2009) suggested that farms on peat soils are less productive, leading to an increasing need to seek for other sources of income. Secondly, it was found that at sites closer to nature, diversification is observed more often as it profits from a more attractive landscape (Pfeifer et al., 2009; Lange et al., 2013; Hassink et al., 2016). Thirdly, closeness to urban centres has some influence, particularly on on-farm sale activities which are more profitable when distances to markets are shorter (Meraner et al., 2015; Hassink et al., 2016). Yet, the relationship depends on the definition of an urban centre and on the type of diversification activity (Meraner et al., 2015).

We suspect that neighbourhood effects play a role in the emergence of diversification clusters in the Netherlands. A four-year impulse of the Taskforce highlighted the importance of local networks for specific diversification activities (Fischer and Ypma, 2012). The presence of multiple farms offering touristic activities enlarges a region’s attractiveness (Fischer and Ypma, 2012), resulting in the development of a common pool resource. Furthermore, Fischer and Ypma (2012) argue that within the activity “on-farm sale” new networks were formed which were then used to transfer knowledge (social learning) and that several regional parties may work together by selling under one regional brand (common pool resource). Fischer and Ypma (2012) as well as Van der Meulen et al. (2014) highlight that nature-conservation activities depend more on subsidies and less on entrepreneurship

[2] After 2012, the Dutch Agriculture and Horticulture Organisation (Land- en Tuinbouw Organisatie Nederland, LTO) took over the information distribution tasks of the Taskforce.

than other activities. However, they emphasize that agricultural associations for nature-conservation activities might induce social learning within a neighbourhood. Also, for care farming neighbourhood effects especially caused by social learning might be relevant. Regional cooperation alliances for care farming exist and are increasingly embedded in the municipality (Fischer and Ypma, 2012; Van der Meulen et al., 2014). Generally, social acceptance and perceived responsibilities within the local community as described by Home et al. (2013) may contribute to neighbourhood effects as well.
The emergence of specific activities may not only induce the emergence of the same activity in the neighbourhood (clustering), but it is reasonable to assume that diversification activities are interdependent. The large percentage of diversified farms (18.5%) active in more than one diversification activity supports this assumption. This aspect is referred to as *diversification interactions* in this research. For example, neighbours active in tourism might increase a farmer's utility of starting with on-farm sale, as potential buyers might be attracted to the region by touristic offers. And neighbours with a farm shop may increase the farmer's utility of on-farm production, when its products can be sold in the neighbour's shop. Moreover, structural characteristics of neighbours may influence the farmer's decision-making (Storm et al., 2015; Lapple and Kelley, 2015).
Table 1 gives an overview of the expected direction of total effects of the structural, locational and interaction variables on the included diversification activities as discussed above. Total effects include direct effects of the farm's own characteristics as well as the indirect effects of the characteristics in the neighbourhood. Following the results of Wollni and Andersson (2014) and Lapple and Kelley (2015), we assume that the direct and indirect effects have the same direction.

Table 1. Hypothesised directions of impacts of independent variables.

	Nature conservation	**Tourism**	**On-farm sale**	**On-farm processing**	**Care**
Structural farm characteristics					
Age	-	-	-	-	-
Farm size	-	+	+	-	+
Pasture farm	+	-	-	-	-
Arable farm	+	-	-	+	-
Horticulture farm	-	-	+	+	-
Perennial farm	0	0	0	0	0
Intensive farm	-	-	-	-	-
Mixed farm	+	-	+	+	-
Locational farm characteristics					
Urbanization	-	+	+	-	+
Landscape attractiveness	+	+	+	+	+
Clay	-	+	+	0	+
Sands	-	+	+	-	+
Peat	+	-	-	-	-
Loam	+	+	+	+	+
Diversification interactions					
Nature conservation	+	-	-	-	0
Tourism	0	+	+	+	0
On-farm sale	0	+	+	+	0
On-farm processing	0	+	+	+	0
Care	0	0	+	+	+

Hypotheses are based on findings from Jongeneel et al. (2008), Pfeifer et al. (2009), Fischer and Ypma (2012), Meraner et al. (2015), and Hassink et al. (2016). + indicate positive, - negative and 0 no expetations of neighbourhood effects.

Quantifying neighbourhood effects

An overview of the most important studies on neighbourhood effects in agricultural decision-making is provided in Table 2. Holloway et al. (2002), Nyblom et al. (2003), Lewis et al. (2011), Wollni and Andersson (2014), Storm et al. (2015) and Lapple and Kelley (2015) regarded single farms, whereas Schmidtner et al. (2012) examined effects on an aggregated level. Another important difference between these studies on neighbourhood effects in agriculture (Holloway et al., 2002; Nyblom et al., 2003; Lewis et al., 2011; Schmidtner et al., 2012; Wollni and Andersson, 2014; Storm et al., 2015; Lapple and Kelley, 2015) is the difference in definition of what a neighbour is. Only Schmidtner et al. (2014) do not work with farms as points in space, but with polygons containing the aggregated information. Polygons having a common border with the polygon examined are defined as neighbours. With no further spatial information available than in which village a farm is located, Holloway et al. (2002) define neighbours as all the other farms in the same village. All other authors define neighbouring farms as farms within a certain radius from the farm. However, in contrast to the other studies introduced, Storm et al. (2015) work with a radius that depends on the distance from the farm to the furthest field in its municipality, and Nyblom et al. (2003) work with a radius which depends on the distance to the nearest neighbour. The highlighted studies (Holloway et al., 2002; Nyblom et al., 2003; Lewis et al., 2011; Schmidtner et al., 2012; Wollni and Andersson, 2014; Storm et al., 2015; Lapple and Kelley, 2015) use different models to estimate neighbourhood effects. Except from Nyblom et al. (2003), who intended to make methodological contributions to the field of neighbourhood effects in agriculture, all mentioned authors (Lewis et al., 2011; Schmidtner et al., 2012; Wollni and Andersson, 2014; Storm et al., 2015; Lapple and Kelley, 2015) use models (also) described by LeSage and Pace (2009). As a follow-up work of Anselin's (1988) [one of the first developers of the field] influential text on spatial economics, it introduces several spatial regression models estimating spatial functional relationships.

Table 2. Overview of literature regarding neighbourhood effects in agricultural decision-making.

Author	**Subject**	**Estimation of spatial functional relationship**	**Country**	**Aggregation level**	**Definition of a neighbour**
Holloway et al. (2002)	HYV rice adoption	Spatial autoregressive probit model (SAR)	Bangladesh	Single farm	Every other farm in the same village
Nyblom et al. (2003)	Organic farming	Logistic regression	Finland	Single farm	Farms within the distance to the nearest neighbour plus 1 kilometre
Lewis et al. (2011)	Organic dairy farming	Probit Mundlak-Chamberlain maximum likelihood	USA (Wisconsin)	Single farm	Within two radii: 5 and 5-10 miles
Schmidtner et al. (2012)	Organic farming	Spatial autoregressive model (SAR)	Germany	County level	Common border
Lapple and Kelley (2015)	Organic drystock farming	Bayesian spatial Durbin probit model (SDM)	Ireland	Single farm	Within three radii: 20, 30, 50 kilometres
Storm et al. (2015)	Direct payments	Spatially lagged explanatory variable probit model (SLX) and spatial Durbin error probit model (SDEM)	Norway	Single farm	Median driving distance to the furthest field in each municipality, within a predefined radius of this distance (max 20 neighbours)
Wollni and Andersson (2014)	Organic farming	Spatial autoregressive probit model (SAR)	Honduras	Single farm	Within six radii between 1.5 and 4 kilometres

Methodology

Utility maximization

The assumptions are made that farmers maximize their utility and that the diversification decision is based on rational choices of each farmer. Utility cannot be measured directly, thus within the theory of utility maximization it is common to assume that the optimal choice is a linear function of different characteristics which explain the binary choice to diversify or not to diversify in a certain activity. We assume that neighbourhood effects influence the decision to diversify through reduced information costs and through enlarged profits due to the emergence of common pool resources. Following Schmidtner et al. (2012) and Wollni and Andersson (2014) with some adjustments to our study, farms diversify if and only if:

$$U_i^{Div}\left(\pi_i^{Div}, TC_i\left(I^{Div}(a_j)\right), \Delta\pi_j\right) > U_i^{Con}(\pi_i^{Con}) \quad (1)$$

$$\text{with } \pi_i^a = p_i^a(a_j) q_i^a\left(S_i, S_j, L_i, I^a(a_j), C_i^a(a_j)\right) - v_i^a(a_j) C_i^a \quad (2)$$

Where U_i is utility of farmer i from activity a {$Div =$ diversified, $Con =$ conventional}, π^a is profit from activity a, TC is the transaction cost of starting with diversified farming, I^a is activity specific information availability, a_j is the activity choice of the neighbouring farmer j, $\Delta\pi_j$ is the increase in profit experienced by farmer i as a result of farmer j's activity choice (common pool resource), p is the output price, q is the production function, S are structural farm characteristics including diversification in other activities, L are locational factors on the farm, C is input quantity depending on characteristics of the neighbouring farmers j, and v is input price.

Neighbourhood and spatial weights

Following literature (e.g. LeSage and Pace, 2009) neighbourhoods are modelled with a so-called spatial weight matrix W. The matrix contains either 0 when a neighbouring farm j is not in the same neighbourhood as farm i, or a value > 0 and ≤ 1 when farm j is in the same neighbourhood as farm i. The structure of this weight matrix has an influence on the estimation of the neighbourhood effect. Storm et al. (2015) refer to this problem as “W is defined rather arbitrary and does not necessarily represent the true neighbouring relationships”. Neighbourhoods have mostly been defined as all farms within a certain distance from a farm (Lewis et al., 2011; Schmidtner et al., 2012; Wollni and Andersson, 2014; Lapple and Kelley, 2015). However, the influence of neighbours may not only be determined by the Euclidian distance between farms, but also by other aspects like the accessibility between them. Moreover, the spatial extent to which the neighbourhood influences a farmer might depend on the density of farms in the region (Strom et al., 2015). If there are only a few neighbours within a radius, a neighbour on a relatively far distance may have a greater influence than when there are many neighbours within the same radius. Furthermore, there might be a maximum number of neighbours influencing the decision-making of a farmer as suggested by Storm et al. (2015). There may also be physical (e.g. rivers and roads) or institutional (e. g. municipal and province) borders, which decrease neighbourhood effects. Storm et al. (2015) thus define a neighbourhood as a maximum of twenty farms within a radius which is the median driving distance to the furthest field in each municipality. Nyblom et al. (2003) define the neighbourhood as all the farms within the distance to the nearest neighbour plus one kilometre.

The neighbourhood of a farm is here defined as its k-nearest neighbours, because most of the Dutch farms have many neighbours within a small distance from their farm (Figure 1). The k-nearest neighbour definition is the most straightforward way to circumvent the problem of dependence of the neighbourhood on the farm density. In the k-nearest neighbour model, every farm has exactly k links to other farms and there is no limitation on the distance of influence of other farms. It is reasonable to give a link in W to a nearer neighbour a higher value, as interactions with nearer neighbours might be stronger than with further ones. To include these aspects, we follow common practice and weight the links based on the inverse of their length and row-standardize the resulting spatial weight matrix. Thus, if a farm has all its k-neighbours within a small distance, differences in distance are small and similar weights will thus be assigned to its links. When a farm has some of its k-neighbours close by and some far away, the long links will get relatively low weights. When all the farm’s neighbours are far away, again similar weights will be assigned. A visualisation of the 10 nearest neighbour algorithm is shown in Figure 2. Farms are shown as dots in both the upper and the lower image. Farms active in tourism are marked as black dots and farms not active in tourism as grey dots, and all farms are linked to their 10 nearest neighbours.

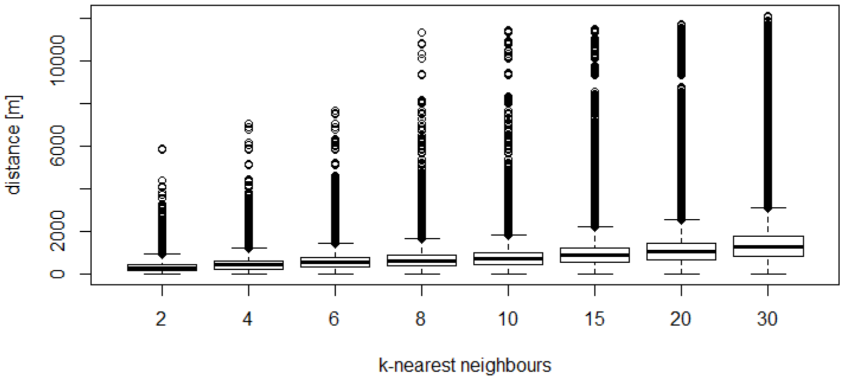

Figure 1. Distances to neighbours for different k's[3] in the Netherlands.

Moran's I Test

Before spatial autocorrelation coefficients are estimated in a spatial regression model, we explore if spatial autocorrelation of diversification is present in Dutch agriculture as well as its direction. The exploration of spatial autocorrelation of different neighbourhood definitions was used to determine an accurate number of k. Neighbourhoods with 10, 25, 50 and 100 nearest neighbours were computed. Individual, close neighbours will have a higher influence when choosing a small k. When choosing a large k, decisions of individuals will have smaller influences and the regional characteristics will have an increased importance. The exploration of autocorrelation is conducted with Moran's I tests for all activities and for 10, 25, 50, 100 neighbours.

$$\text{Moran's I test:} = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}, \quad (3)$$

where N is the number of spatial units indexed by i and j; X is the variable of interest; $\bar{X}$ is the mean of X; and w_{ij} is an element of a matrix of spatial weights.

The Moran's I test tells if neighbouring features both deviate in the same direction from the mean of an attribute. Moran's I values lie between -1 and +1, where -1 represents maximal dissimilarity, and +1 maximal similarity of neighbours, and while at 0 the spatial distribution is random. For the spatial regression, the number of neighbours referring to the highest Moran I value will be used.

Spatial regression

Following other spatial discrete choice models in agricultural literature (Holloway et al., 2002; Schmidtner et al., 2012; Wollni and Andersson, 2014; Storm et al., 2015; Lapple and Kelley, 2015) we assume that a farmer has the binary choice to diversify or not to diversify. $y*$ is assumed to be determined by the farm's characteristics as well as by spatially dependent neighbourhood effects. To control for neighbourhood effects, we follow Wollni and Anderson (2014) and use a Bayesian spatial autoregressive probit model (LeSage and Pace, 2009):

$$y* = \rho W y* + X\beta + \varepsilon \quad (4)$$

The right-sided term $y*$ is the spatial lagged version of the dependent variable $y*$. W is a matrix containing a value > 0 and ≤ 1 when there is a link between two farms and 0 when there is no link. ρ denotes the estimate correlation between neighbours. $X\beta$ captures additional predictor variables. Applied to this research, X is a matrix containing the farm's and its neighbours' structural and locational characteristics and β is the strength of correlation. ε is a random error term assumed to follow a normal distribution. The model parameters ρ, β and $y*$ are estimated using Bayesian techniques incorporating Markov Chain Monte Carlo (MCMC) sampling, as MCMC is a powerful way to avoid problems of multidimensional integration (Holloway et al., 2002).

[3] The isolated farm on the island Vlieland in the North Sea is not shown as it has no neighbours within a comparable distance.

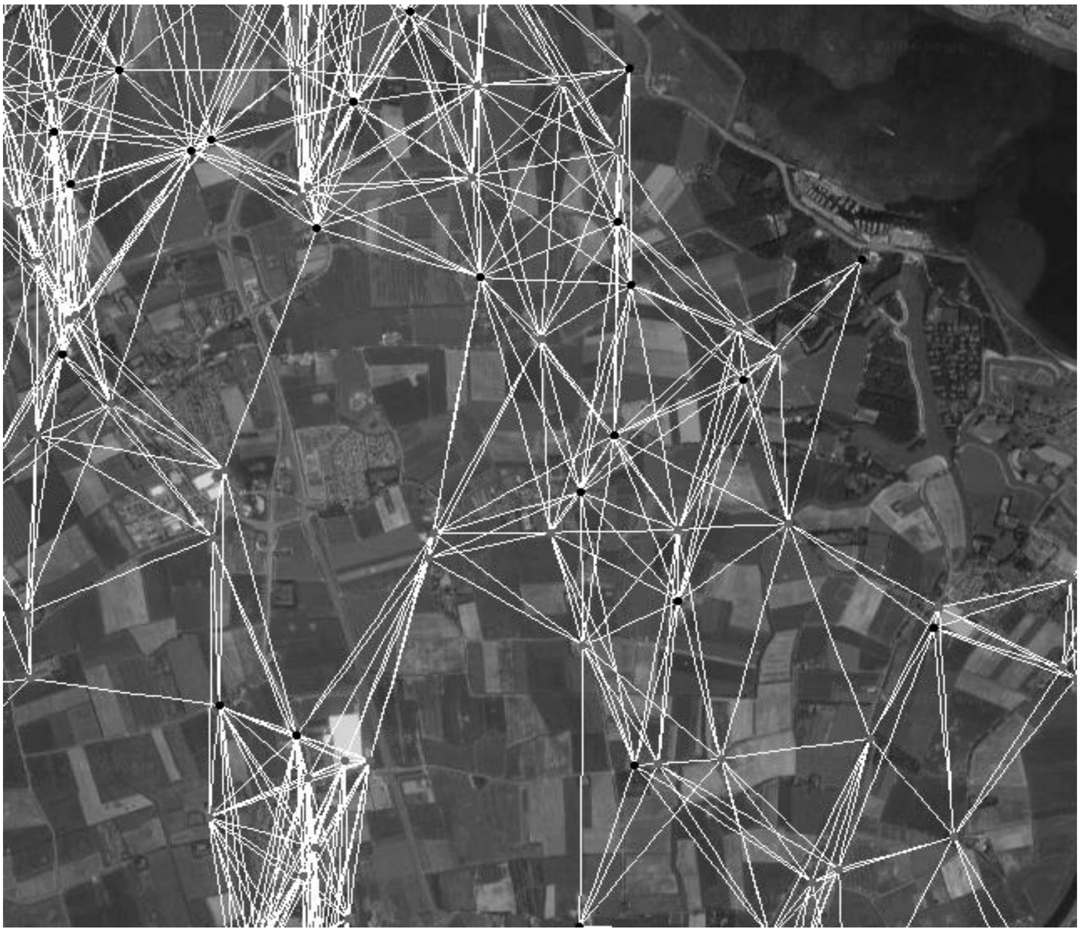

Figure 2. Farm connections generated with 10 nearest neighbours network algorithm.

Data

Individual farm data is provided by Wageningen Economic Research. Individual farm data is collected with the Dutch agricultural census on an annual basis, but this does not include the same questions every year. The most recent farm level data about diversification was collected in 2013. The data was collected from 67’481 farms, and was extended with a data set also provided by Wageningen Economic Research on farm locations and soil types, and with open access spatial data from the Dutch Central bureau for statistics (CBS) on population density and on natural areas. These additional spatial attributes were assigned to the individual farms by use of the farms’ locations with ArcGIS. A data analysis showed that there are 1’310 farm locations which are not unique. On these locations two or more farms are registered, leading to a total of 2’820 farms which share a location with at least one other farm. This is problematic when assigning weights, because no weight can be assigned to links of length zero. To solve this, multiple farms on the same location are regarded as one large farm[4], leaving a total of 65’976 farms. Locational characteristics of farms on shared locations are the same by default. The total workforce and the mean age of the farmer at the location are considered. For farm type and diversification activities, all types and activities are regarded.

As explained in the section ‘Diversification in Dutch agriculture’, structural, locational and interaction characteristics were included in our research. First, the choice of structural variables is based on previous research (e.g. Benjamin, 1994; Jongeneel et al., 2008; Meraner et al., 2015). Meraner’s (2015) research of Dutch farm diversification showed that age, farm size and farm type are significant characteristics for diversification. Second, the three most often discussed locational characteristics of the farm’s location (soil properties, landscape attractiveness and urbanization) are included to correct for spatial heterogeneity. Additionally, the farm’s activities in other diversification strategies are included as dummy variables. Table 4 shows a summary statistics of all variables.

[4] See Annex, page 29-30.

Structural characteristics

The average *age* of Dutch farmers is 55 years. Farmers on diversified farms are on average two years younger. Available farm size measures in the Netherlands include area, economic output and labour input. Measures capturing farm size are expected to highly correlate. Based on the findings of Storm et al. (2015) and the unsuitability of other measures, *workforce* measured in full-time equivalents is included as a size variable in the analysis. On average, every farm occupies 2.4 persons for a whole year, on diversified farms the workforce is slightly lower than on other farms. Storm et al. (2015) included total area, total direct payments and total workforce and found that “due to the high correlation of the three size variables, a model with just one of the three size variables has approximately the same explanatory power”. A farm’s area is not regarded suitable here, as the area of a farm is highly correlated with the farm type. The economic farm size in the Netherlands is estimated with the standard economic output (SO) of a farm. SO’s are defined as the average yield of a product in the last five years measured in Euros per ha or per animal. When multiplied with a coefficient capturing the percentage of the yields which is left after standardized costs of the output of the SO, the standard earning capacity (SVC) of a farm is found. Compared with the SO, the SVC allows a better comparability of economic farm sizes between different farm types. However, yields from diversification activities are not included in the SO’s, and thus they are not included in the SVC either (Van Everdingen, 2015). Therefore, a relation between economic size and diversified agriculture emerges from the data structure (diversified farms will have lower incomes because not all their income is counted). The correlation of workforce and SO is 0.61 in the Netherlands. For the high number of observations this correlation is clearly significant. The Dutch classification of *farm types* is based on the percentage of standardized outputs (SO) from one product category. The product categories distinguished in the Netherlands are: horticulture (‘Hort’), pasture-based livestock (‘Past’), arable farming (‘Ara), perennial farming[5] (‘Pera’) and intensive livestock (‘Inte’). If more than 2/3 of the farms income is generated by products from the same category, the farm is classified as a farm from this category. If the farm does not generate 2/3 of its income with products from the same category, the farm is classified as a mixed farm. There are comparably many pasture, perennial and mixed farms that have diversified. On the other hand, the number of intensive livestock and horticulture farms conducting diversification activities is relatively small (Table 4).

Locational characteristics

All farms were assigned a *soil type*. Most of the Dutch farms are located on sands (‘Sand’) and on clays (‘Clay’), which are also the most frequently occurring soil types throughout the country. Further, 12% of Dutch farms are located on peat soils (‘Peat’). Loam (‘Loam’) soils are mainly found in the very south of the country. Other soil types in the data set (water and urban grounds) were grouped as ‘Others’. Figure 3 (left) gives an overview of soil types throughout the country.
Urbanisation was measured as population density on the municipality level. The distribution of the data makes a logarithmic use of this data favourable. The mean Dutch farm is located in a municipality with 400 inhabitants per square kilometre. Diversified farms are on average more frequent in less population dense areas. However, standard deviations are large. Because the population density is the same within a municipality, most farms have the same score of urbanisation as their neighbours.

[5] Includes wine farms, fruit growers and other perennial farming types (Van Everdingen, 2015).

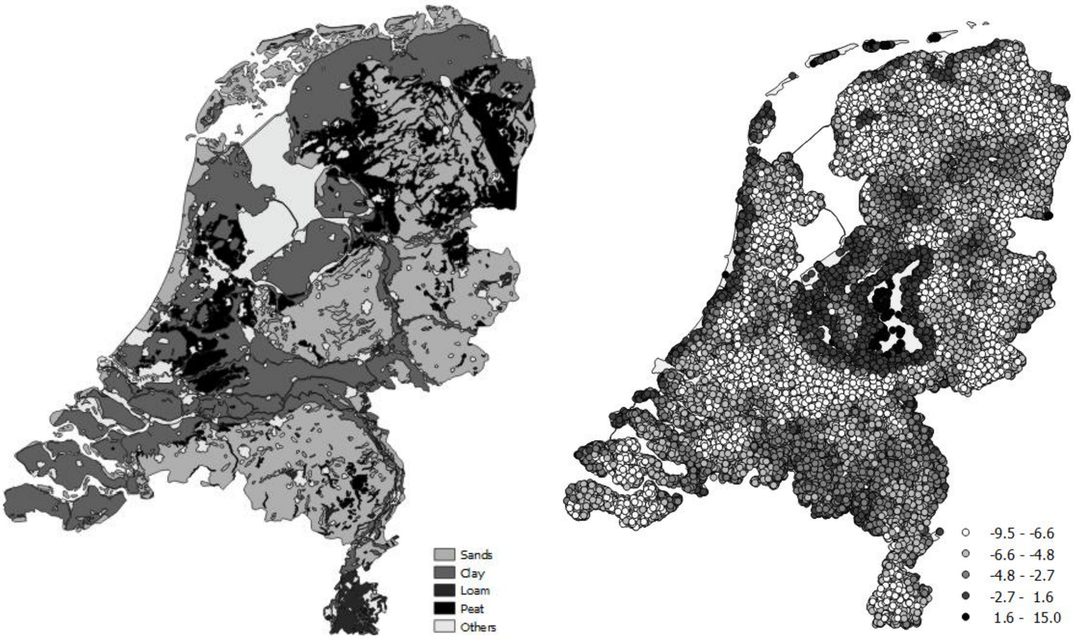

Figure 3. Soil types in the Netherlands (left) and logarithmic Reilley-index scores of Dutch farms (right)

The *attractiveness of the landscape* around a farm was estimated with the Landscape Reilly Index (short: Reilly-index) as distinguished by Geohegan et al. (1997). For this, the size of every natural area within a certain radius around the farm is normalized by the distance from the farm to the territory. The farm is then assigned a score which is the sum of all the normalized areas within this radius. We defined natural areas as areas that are classified as woods or as open nature in the land-use classification of the CBS of 2013. The radius was set to five kilometres, which is related to a twenty minutes' bike trip or a few minutes' car drive from a farm. We defined a minimum size of 10 ha for natural areas to exclude very small, single and remote natural areas which do not necessarily contribute to the regional landscape attractiveness.

Reilly-index: $R = \sum_{i=1} \quad \sum_{y=1} \frac{size\ of\ natural\ area\ y\ within\ a\ radius\ r\ of\ farm\ i}{(shortest\ distance\ of\ firm\ i\ to\ natural\ area\ y)^2}$ (5)

High scores on the Reilly-index indicate that a farm is close to large natural areas whereas small scores indicate that a farm is relatively far from rather small natural areas. 68 farms have no natural areas within less than five kilometres distance, their score on the Reilly-index was set to zero. Six farms located in the large national park 'De Hoge Veluwe' got an infinite score. They were assigned the maximum score reached by any other farm (3166523). The mean score of diversified farms is higher than the mean score of non-diversified farms (Table 3), mainly because Reilley-scores of tourism and nature conservation farms are relatively high. Sale, care and processing activities reach low scores indicating that farms carrying out these activities are located rather far from natural areas. Most farms have quite similar Reilley-scores as their neighbours' farms. However, standard deviations are high. The distribution of the Reilley-index favours a logarithmic use of the variable.

Table 3. Overview of Reilly-index statistics.

	Total number	**Median**	**Median absolute deviation**
All farms	65976	0.00488	0.00605
Non-diversified farms	53535	0.00495	0.00615
Diversified farms	12442	0.00455	0.00554
Nature conservation	7371	0.00383	0.00444
Tourism	2772	0.00788	0.01050
Sale	3113	0.00524	0.00652
Processing	1034	0.00518	0.00642
Care	868	0.00411	0.00487

Table 4. Summary statistic for explanatory variables.

	All farms (n = 65976)		No diversification (n = 53534)		Diversification (n = 12442)	
	Mean	**Std. dev**	**Mean**	**Std. dev**	**Mean**	**Std. dev**
Structural Characteristics						
Age	55.150	11.608	55.538	11.791	53.522	10.631
Workforce	2.433	7.449	2.459	8.145	2.319	2.971
Past	0.543	0.498	0.521	0.500	0.638	0.481
Ara	0.179	0.384	0.186	0.389	0.152	0.359
Hort	0.133	0.340	0.146	0.353	0.080	0.272
Inte	0.080	0.271	0.089	0.285	0.038	0.191
Pera	0.025	0.156	0.021	0.144	0.040	0.197
Mix	0.049	0.216	0.046	0.210	0.062	0.241
Locational Characteristics						
PopDense	399.900	492.331	402.030	489.945	390.745	502.381
Clay	0.353	0.478	0.341	0.474	0.402	0.490
Sands	0.490	0.500	0.515	0.500	0.383	0.486
Peat	0.120	0.325	0.110	0.313	0.162	0.369
Loam	0.016	0.125	0.012	0.111	0.031	0.173
Other soils	0.021	0.145	0.021	0.145	0.022	0.145
Diversification interactions						
Nature conservation	0.112	0.315	0.000	0.000	0.592	0.491
Tourism	0.042	0.201	0.000	0.000	0.223	0.416
Sale	0.047	0.212	0.000	0.000	0.250	0.433
Processing	0.016	0.124	0.000	0.000	0.083	0.276
Care	0.013	0.114	0.000	0.000	0.070	0.255

Statistics of Reilly-index for landscape attractiveness are presented in Table 3.

Results

Moran's I

The results of the Moran's I test show that there is a positive relation between diversification and the adoption of diversification of neighbours (Table 5). Except for nature conservation, autocorrelation is rather small but in every case significant at a five percent level. The significance of autocorrelation is robust between different numbers of neighbours included. With increasing numbers of k, autocorrelation decreases, but the significance of the phenomenon rises. This robustness of significance with varying weight matrixes is in line with the findings of Storm et al. (2015) and Lapple and Kelley (2015), whose results were mostly not affected by different definitions of W. This finding led to the decision to continue with a neighbourhood defined as the 10 nearest neighbours, since with this definition we found the largest autocorrelation. With the results of the Moran's I test, we do not yet know if autocorrelation is present due to more than spatial heterogeneity. With a visual analysis of the modelled networks[6], we found that with a larger number of k, especially intra-community links increase, but the structure of the network remains more or less the same. Therefore, there are no entirely new characteristics entering the neighbourhood with increasing numbers of k, which would be one explanation for the low sensitivity on k. Another explanation is found when regarding the construction of W. Due to the inverse distance weighting of the links, proximate neighbours are more important than further ones. With higher numbers of k, relatively further and thus less important neighbours enter the model.

Table 5. Moran's I results.

	Nature conservation	**Tourism**	**Sale**	**Processing**	**Care**
10 neighbours	0.223***	0.062***	0.020***	0.016**	0.005*
25 neighbours	0.212***	0.056***	0.019***	0.014**	0.004**
50 neighbours	0.203***	0.051***	0.018***	0.013***	0.004***
100 neighbours	0.190***	0.047***	0.017***	0.011***	0.004***

Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

[6] Figures are provided in the Annex.

Spatial regression

Table 6 shows the results of the spatial regression for the spatial autoregressive probit model. We observe that ρ is significant at a five percent level for nature conservation, tourism and on-farm sale and on a ten percent level for on-farm processing, indicating that neighbourhood effects matter in the emergence of diversified agriculture. The positive sign of ρ shows that farms are more likely to start with a certain diversified activity if their neighbouring farms are conducting this activity. For care farming the sign is found to be negative, indicating that neighbourhood effects may be disadvantageous. However, ρ is clearly not significant in this case. As the outcome of a model with binary dependent variables is non-linear, the magnitude of the impacts of the independent variables is found by deriving marginal effects. The impact of a change in the independent variable on the farm is captured in the direct effect. Indirect effects are the cumulative effect of a change in the independent variable of neighbouring farms on the adoption probability of the farm. When ρ is larger, indirect effects are more important. As ρ is not significant for care farming, we do find almost no indirect effects on the emergence of care farming. For continuous variables, marginal effects are estimated at the mean and for dummy variables at a change from zero to one (Wollni and Andersson, 2014). Marginal effects are presented in Table 7. As found also by other authors (e.g. Wollni and Andersson, 2014; Lapple and Kelley, 2015), direct effects are generally larger than indirect effects.

With respect to *structural farm characteristics*, the results presented in Table 6 indicate that younger farmers are more likely to diversify in any of the activities. The direct marginal effects in Table 7 show that with every additional life year of a farmer on farm i the chance to start with nature conservation or on-farm sale decreases with 0.1 percent. For other activities, the marginal effect is, unless significant, close to zero. Table 7 also shows that a small negative indirect effect of the age of neighbours is found for nature conservation activities: when the cumulative age of farmers of the ten nearest farms to farm i is one year higher, the chance of farm i to diversify decreases with 0.1 percent, leading to a total effect of 0.2 percent of the variable age on nature conservation.

The workforce is found to be important for the adoption of nature conservation, on-farm sale and on-farm processing, with larger farms being more often active in nature conservation and on-farm processing and smaller farms in on-farm sale (Table 6). The indirect effects in Table 7 show that there is a small positive effect of large neighbours on the emergence of nature conservation activities. Farms are 0.1 percent more likely to start with nature conservation when neighbouring farms occupy one person more.

The estimates of the dummy variables of the farm type are compared to perennial farming. They show a significant relation between pasture-based farming and diversification into all activities (Table 6). However, the direction of the effect of pasture-based farming is different for the diverse activities: pasture-based farming positively influences nature conservation, tourism and care farming. Pasture-based farms have for example a 10.6 percent higher chance to start with nature conservation. With a neighbour j changing its farm type to pasture-based farming, the chance that farm i starts with nature conservation increases by 13.8 percent. On-farm sale and on-farm processing are directly and indirectly negatively related with pasture-based farming. Table 7 shows that direct and indirect effects of arable farms are similar to those on pasture-based farms, but these effects are less strong. Horticultural farms are less likely to diversify in any diversification activity. This negative effect is significant at a five percent level for nature conservation, tourism and on-farm sale. Table 7 shows that having neighbours with horticultural farms decreases the probability of diversifying. Yet, for on-farm sale, the indirect negative effect of horticultural neighbours is almost negligible relative to the direct effect. For nature conservation, on the other hand, the indirect effect of horticultural neighbours is larger than the direct effect. Intensive livestock farms are significantly less likely to start with the diversification activities tourism, on-farm sale and on-farm processing. On-farm sale is the activity found to be the most affected by the farm type intensive livestock on the farm itself, while the emergence of tourism activities suffers most from intensive-livestock farming neighbours. Table 6 shows that mixed farming is at a five percent level significantly positively related with nature conservation and care farming, and significantly negatively related with on-farm sale.

Table 6. Results of spatial autoregressive model.

	Nature conservation	Tourism	On-farm sale	On-farm processing	Care farming
Constant	-0.811***	-0.876***	-0.768***	-1.761***	-2.583***
Structural characteristics					
Age	-0.008***	-0.005***	-0.007***	-0.009***	-0.009***
Workforce	0.003*	0.001	-0.005**	0.004***	0.002
DumPast	0.719***	0.142**	-1.184***	-0.147*	0.233**
DumAra	0.409***	0.057	-0.818***	-0.019	-0.284***
DumHort	-0.182***	-0.251***	-0.475***	-0.122*	-0.146.
DumInte	-0.022	-0.308***	-0.809***	-0.365***	0.099
DumMix	0.497***	0.088.	-0.407***	0.030	0.313***
Locational characteristics					
LogPopDense	-0.001	0.007*	0.032**	0.000	0.068***
LogReilley	0.004*	0.035***	-0.008*	0.008.	-0.013*
DumClay	0.189***	-0.098***	0.002	0.153***	-0.046.
DumPeat	0.276***	-0.156***	-0.127***	0.158***	-0.004
DumLoam	0.433***	0.112*	0.255***	-0.080	-0.151.
DumOthers	0.080*	0.053	0.111*	-0.031	0.205*
Diversification interactions					
Nature Conservation		0.270***	0.271***	0.206***	0.274***
Tourism	0.371***		0.666***	0.603***	0.696***
Sale	0.348***	0.674***		1.616***	0.557***
Processing	0.181***	0.658***	1.969***		0.164*
Care	0.389***	0.853***	0.659***	0.279***	
Rho	0.601***	0.365***	0.053**	0.127.	-0.040

Significance codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1, ‘ ’ 1.

Table 7. Marginal effects.

	Nature conservation		Tourism		On-farm sale		On-farm processing		Care farming	
	direct	**indirect**	**direct**	**indirect**	**direct**	**indirect**	**direct**	**indirect**	**direct**	**indirect**
Structural characteristics										
Age	-0.001	-0.001	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000
Workforce	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DumPast	0.106	0.138	0.011	0.006	-0.093	-0.005	-0.004	-0.001	0.007	0.000
DumAra	0.060	0.078	0.004	0.002	-0.065	-0.003	-0.001	0.000	-0.009	0.000
DumHort	-0.027	-0.035	-0.019	-0.010	-0.038	-0.002	-0.003	0.000	-0.005	0.000
DumInte	-0.003	-0.004	-0.023	-0.012	-0.064	-0.003	-0.010	-0.002	0.003	0.000
DumMix	0.073	0.095	0.007	0.004	-0.032	-0.002	0.001	0.000	0.010	0.000
Locational characteristics										
LogPopDense	0.000	0.000	0.001	0.000	0.003	0.000	0.000	0.000	0.002	0.000
LogReilley	0.001	0.001	0.003	0.001	-0.001	0.000	0.000	0.000	0.000	0.000
DumClay	0.028	0.036	-0.007	-0.004	0.000	0.000	0.004	0.001	-0.001	0.000
DumPeat	0.041	0.053	-0.012	-0.006	-0.010	-0.001	0.004	0.001	0.000	0.000
DumLoam	0.064	0.083	0.008	0.004	0.020	0.001	-0.002	0.000	-0.005	0.000
DumOthers	0.012	0.015	0.004	0.002	0.009	0.000	-0.001	0.000	0.006	0.000
Diversification interactions										
Nature Conservation			0.020	0.011	0.021	0.001	0.006	0.001	0.009	0.000
Tourism	0.055	0.071			0.053	0.003	0.017	0.002	0.022	-0.001
Sale	0.051	0.067	0.050	0.027			0.044	0.006	0.017	-0.001
Processing	0.027	0.035	0.049	0.026	0.155	0.008			0.005	0.000
Care	0.057	0.075	0.064	0.034	0.052	0.003	0.008	0.001		

Mixed farming is especially important for the emergence of nature conservation as shown in Table 7. Both direct and indirect effects of mixed farming are comparably large.

Our results show that *locational characteristics* also matter in the adoption of diversified activities. Significances are presented in Table 6. Higher population densities increase the adoption of on-farm sale and care farming. With respect to the landscape attractiveness, we find that this is increasing the chance to start with tourism and nature-conservation, but at a five percent level significantly decreasing the chance of starting with on-farm sale and care farming. Interestingly, unless the

Reilley-index is assumed to be almost the same between neighbouring farms, there is a small positive indirect effect of the landscape attractiveness of neighbouring farms on tourism and on nature conservation. Indirect effects of the landscape attractiveness only occur when there are differences in Reilley-scores between neighbours, while both are conducting the same activities. This indicates that the positive effect of natural areas on the emergence of diversification goes beyond our artefactually set border of five kilometres.

For soil types, sands are the reference group. Nature conservation is found more often on loam soils than on sandy soils and having neighbours located on loam soils also has a positive effect on the emergence of nature conservation. Tourism farms and on-farm sale are found the least on peat soils and most on loam soils. On-farm processing is found most on peat and clay soils. However, neighbourhood effects of soil types are mainly an artefact because farms got assigned a soil type based on the location of their main farm building. This might in many cases not be the same soil type as on their fields.

Diversification interaction are also coded as dummy variables. The dummy variable ‘no diversification’ is dropped. Diversification in other activities has a strong influence on the start of other activities (Table 6). From Table 7 we see that a farm’s chance to start with any diversification activity is greater when it is already conducting a diversified activity, independent of the type of diversification. Having diversified neighbours increases the chance of a farm to diversify. Care farming is an exception to this rule: where a farm’s activity in other diversification activities enlarges the chance to start care farming, having diversified neighbours might restrict the emergence of care farming. The results show that all diversification activities combine well with nature conservation both directly and indirectly. Especially indirect effects are relatively important for nature conservation. The emergence of touristic activities is supported most by other activities on the farm and are seen more often when neighbours are conducting other diversification activities. Farms that are already active in on-farm processing have a 15.5% higher chance to start with the activity on-farm sale than farms that do not conduct any diversification activity. The chance to start with on-farm processing, on the other hand, is only 4.4% larger when a farm is already conducting sale activities.

Discussion

Neighbourhood effects

By estimating the spatial correlation coefficient ρ, we showed that neighbourhood effects matter in the adoption of diversification. Unless some of the clustering can be explained with locational attributes, the estimated ρ is found to be positive and significant at a five percent level for nature conservation activities, tourism and on-farm sale. For on-farm processing it is significant on a ten percent level. We found the largest estimate of ρ for nature conservation (0.601), indicating that the probability to start with nature conservation strongly depends on the characteristics of farms in the neighbourhood. The high estimates of neighbourhood effects in nature conservation are probably caused by three facts. First, social learning occurs more often when neighbours are successful (Munshi, 2004). As nature conservation activities are supported by the Dutch government (Van der Meulen et al., 2014), nature conserving neighbours might more often be successful than other diversified neighbours. Second, nature conservation is in many cases only possible with a local collective of farmers on favourable locations (Van der Meulen et al., 2014). Third, nature conservation needs less individual entrepreneurship than other diversification strategies (Fischer and Ypma, 2012), which might make it a preferable diversification strategy.

For touristic activities ρ is 0.365, indicating that for the adoption of tourism, the characteristics of farms in the neighbourhood are important as well. The strength of neighbourhood effects on the emergence of tourism is comparable to the strength of neighbourhood effects on the emergence of organic agriculture found by Wollni and Andersson (2014). Tourism activities occur more in neighbourhoods with sale and processing farms. A probable explanation for the cause of the benefit for touristic farm activities is the emergence of common pool resources (Fischer and Ypma,2012; Van der Meulen et al., 2014).

We found relative large positive neighbourhood interactions between sale and processing farms. This is not surprising as both activities can profit from the same common pool resources, like for example local brands. Yet, having neighbours with on-farm processing activities increases the chance of a farm to start with on-farm sale more than vice versa. A possible reason for this is found in the relatively larger size of processing farms. With the advantages of being larger, on-farm processers might be able to behave more independently and might have better access to financial resources and knowledge. Moreover, for on-farm sale, new regional networks for marketing, research, sourcing and logistics were formed with the support of the Taskforce (Fischer and Ypma, 2012) which may have contributed to the importance of neighbourhood effects. On-farm processing however, was not captured by the Taskforce.
Due to the relatively low number of care farms, the significance of the ρ of care farming is more sensitive to changes in the variables. Although not significant, it was not expected that the sign of ρ was negative. This is because regional corporations exist (Fischer and Ypma, 2012) and the Moran's I score of care farming was slightly positive. A possible explanation is that due to a limitation of people who can be helped and who can help on care farms, competition might be relevant in care farming. That the population density has a positive effect on the emergence of care farms explains why the Moran's I score can still be slightly positive: care farms do occur somewhat clustered, but this is due to the fact that they are located in population dense areas. Without this reason for clustering, care farms might occur dissimilative due to negative neighbourhood effects probably caused by competition. The relatively large positive effect of neighbouring care farms on the emergence of other activities on farm 's i is less expected. A care farm apparently contributes to the common pool resource of a region. However, care farms do not seem to profit from diversification in the neighbourhood.

Other neighbourhood characteristics influencing diversification

Our results additionally indicate that the indirect neighbouring effects are to a high extent due to neighbourhood effects of diversification activities. However, other neighbourhood characteristics are found to influence the adoption of diversification as well. Regarding the *structural characteristics*, a higher age of neighbours seems to have a small negative effect on the adoption of nature conservation. Having older neighbours might reduce the local social acceptance of nature conservation in the neighbourhood, leading to decreased adoption chances (Home et al., 2013). With respect to workforce, again only nature conservation activities are found to be slightly influenced by the workforce on neighbouring farms. This might be a result of competition, since smaller farms might sooner give up a field and transform it into a natural area when they are surrounded by larger farms (Storm et al., 2015). The farm types in the neighbourhood are of more importance for the emergence of diversification. Nature conservation is often found in the same neighbourhood as pasture-based, arable and mixed farms. A possible explanation is that these farm types are more often diversified (Weltin et al., 2017), which might lead to an increased social acceptance in the neighbourhood (Home et al., 2013). Moreover, as nature conservation is often conducted by collectives (Van der Meulen et al., 2014), having neighbours with farm types which are suitable for nature conservation might increase the chance of farm i to take up nature conservation as well. Tourism activities do also profit from pasture-based farms in the area, while especially horticultural and intensive-livestock neighbours reduce the chance of farm i to start with tourism. It is possible that the attractiveness of the landscape is not only dependent on the attractiveness of natural areas as measured in our research, but also on the attractiveness of the main farm type in the neighbourhood. The remaining diversification activities are less influenced by the neighbours' farm types. On-farm sale emerges most when neighbours are perennial farmers and least when they have pasture-based farms. It is possible that farmers with a farm shop do not only sell their own products, but also local products produced by their neighbours. The fact that perennial farms produce relatively valuable crops (Van Everdingen, 2015), might explain why they increase the chance of their neighbours to start with on-farm sale. When there are many pasture-based farms in the neighbourhood, sale activities emerge less frequently. A possible explanation for this is that the accessibility of farms is probably lower when they are surrounded by fields. On-farm processing is slightly disadvantaged by neighbours with intensive-livestock farms. Since for on-farm processing the existence of common pool resources are probably relevant, our results indicate that the opposite effect also exists. Possibly, these farms harm the image of the region and in this way reduce the market position of the farm-made products from the region.

Neighbourhood effects of *locational characteristics* are mainly an artefact of their measurement. As long as the locational attributes of neighbours do not differ, no indirect effects result. The Reilley-score is continuous, so Reilley-scores of neighbours are similar and in most cases no indirect effects occur. There is however a small indirect effect for tourism and nature conservation where landscape attractiveness is important. This indicates that the attractiveness of the landscape works beyond five kilometres, because apparently farms outside the border do profit from the attractive landscape characteristic of farms within the border.

Additional insights in direct effects

We found that younger farmers engage more in diversification activities, which is in line with earlier findings (Jongeneel et al., 2008; Barbieri and Mahoney, 2009; Meraner et al., 2015, Weltin et al., 2017). Barbieri and Mahoney (2009) explain this with the assumption that younger farmers are more in need of reducing risks due to their longer-term activities. Unlike Meraner et al. (2015), we did also find this effect for nature conservation activities. Unless significant at a high level, the direct effect of the workforce is relatively small compared to earlier studies (Jongeneel et al., 2008; Pfeifer et al., 2009; Meraner et al., 2015). This indicates that the inclusion of neighbourhood and interaction effects led to a reduced importance of the workforce compared to earlier studies.

Our results are also in line with earlier findings that the proximity to cities is important for on-farm sale and care farming (Ilbery, 1991; Pfeifer et al., 2009; Meraner et al., 2015). Both activities profit from closeness to larger markets. As also reported by Pfeifer et al. (2009), we found that attractive landscapes contribute to the emergence of tourism. Providing tourism activities is relatively more valuable when people are attracted by the landscape (Pfeifer et al., 2009). Moreover, we found a positive relation between nature conservation and attractive landscapes. A probable explanation is that nature conservation activities are especially supported when proximate to natural areas (Van der Meulen et al., 2014). On-farm sale and care farming appear more when they are further away from attractive landscapes, probably because natural areas are often in remote areas where markets are smaller.

We do find that nature conservation occurs often on farms located on peat soils, which according to Pfeifer et al. (2009) is because conserving nature occurs more on less productive soils. The indirect effect of peat soils gives rise to the assumption that the effect of peat soils may have been underestimated. This is because a farm's soil type was here defined as the soil type at the main farm building. This is however, not necessarily the same as on at least a part of a farm's fields.

Additionally, we find that direct diversification interactions are of large importance which is also found by Weltin et al. (2017). We find the largest interdependences on the farm i itself between on-farm sale and on-farm processing. When already producing on the farm, the step to selling farm-made products is small. We do also find that farms that are already active in nature conservation have a higher chance to start with a new activity, but the chance to start with nature conservation when already doing another activity is much higher. A possible explanation might be that when the first activity is successful, investing work hours in the activity might be more profitable than working on the field, which again increases the benefit of transforming the field towards nature conservation. Care farming is most advanced on tourism and on-farm sale farms. Yet again, the interaction is stronger in the reversed direction: other activities with the exception of processing are increasingly taken up on care farms. This might result from the potential of cared-for people to contribute in other diversification activities.

Conclusion

Our results show that neighbourhood effects are important in the emergence of agricultural farm diversification. Farms with diversified neighbours have more incentives to diversify as well. Neighbourhood effects are found to be especially important for nature conservation activities. For tourism and on farm-sale we did also find that these effects are significant at a five percent level.

We believe neighbourhood effects are mainly caused by three phenomena. First, diversifying farms create common pool resources which increases the benefit of diversification for other farms in the neighbourhood. Second, diversified farms reduce the information costs for other farms in the region, making it easier for other farms to diversify as well. Third, local social acceptance of diversification may increase when more farms in the neighbourhood diversify, which may lead to a greater chance to adopt diversification. We can additionally suggest that not only the neighbours' diversification activities matter for the decision to diversify, but also the main farm type of surrounding farms influences the emergence of diversification. The direction of effects of neighbouring farm types differ for varying diversification activities.

We think that further research should investigate how neighbourhood effects can be used for the support of new farm strategies in the Netherlands. Neighbourhood effects might affect the effectiveness of support programmes, since the decision of a farmer seems to influence the decision-making of his neighbours. We also think that deeper investigation into the structure of networks which cause neighbourhood effects is necessary. Because we did not find neighbourhood effects for care farming, but regional cooperation alliances for care farming exist, neighbourhood effects are likely to not being captured by a spatial modulation of a network. We suggest that a definition of local networks based on activities of farmers in local collaborations, alliances etc. which does also respect spatial borders and/or farm accessibility would better capture neighbourhood effects. Moreover, we think future research could come up with an index for landscape attractiveness which does also respect the main farm type in the region.

Our findings highlight the importance of neighbourhoods for farm diversification. Moreover, our research underlines that diversification promotes not only income diversification of the farm itself, but also creates new economic chances for farms in the neighbourhood. Therefore, we could quantify the local impacts of diversifying farms. We could show that by enlarging chances both on the farm itself and on farms in the neighbourhood, farm diversification contributes directly and indirectly to rural development.

Acknowledgements

We would like to thank Tom Kuhlman for his valuable inputs about spatial characteristics and advice on measurement of spatial data.

Bibliography

Augère-Granier, M.-L. 2016. Briefing Farm diversification in the EU. European Union: European Parliamentary Research Service.

Anselin, L. 1988. Spatial econometrics: methods and models. Dordrecht, Netherlands: Kluwer Academic Publishers.

Barbieri, C., and E. Mahoney. 2009. Why is diversification an attractive farm adjustment strategy? Insights from Texas farmers and ranchers. *Journal of Rural Studies* 25 (1):58-66.

Benjamin, C. 1994. The growing importance of diversification activities for French farm households. *Journal of Rural Studies* 10 (4):331-342.

Case, A. 1992. Neighborhood Influence and Technological Change. *Regional Science and Urban Economics* 22 (3):491-508.

Everdingen, van, W. 2015. NSO-typering 2015 : typering van agrarische bedrijven in Nederland. Den Haag, Netherlands: LEI Wageningen UR.

Fischer, M., and T. Ypma. 2012. Vier jaar impuls voor de multifunctionele landbouw: inzet Taskforce Multifunctionele Landbouw 2008-2012. Taskforce Multifunctionele Landbouw.

Geoghegan, J., L.A. Wainger, and N. E. Bockstael. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics* 23 (3): 251-264.

Hassink, J., H. Agricola, and J. Thissen. 2016. Participation rate of farmers in different multifunctional activities in the Netherlands. *Outlook on Agriculture* 45 (3):192–198.

Heringa, P.W., C.M. van der Heide, and W.J.M. Heijman. 2013. The economic impact of multifunctional agriculture in Dutch regions: An input-output model. *NJAS-Wageningen Journal of Life Sciences* 64-65:59-66.

Holloway, G., B. Shankar, and S. Rahmanb. 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. *Agricultural Economics* 27 (3):383–402.

Home, R., O. Balmer, I. Jahrl, M. Stolze, and L. Pfiffner. 2014. Motivations for implementation of ecological compensation areas on Swiss lowland farms. *Journal of Rural Studies* 34:26–36.

Ilbery, B.W. 1991. Farm Diversification as an Adjustment Strategy on the Urban Fringe of the West Midlands. *Journal of Rural Studies* 7 (3):207-218.

Jongeneel, R.A., N.B. Polman, L.H. Slangen. 2008. Why are Dutch farmers going multifunctional? *Land use policy* 25 (1):81-94.

Lange, A., A. Piorr, R. Siebert, and I. Zasada. 2013. Spatial differentiation of farm diversification: How rural attractiveness and vicinity to cities determine farm households' response to the CAP. *Land Use Policy* 31:136-144.

Lapple, D., and H. Kelley. 2015. Spatial dependence in the adoption of organic drystock farming in Ireland. *European Review of Agricultural Economics* 42 (2):315-337.

LeSage, J.P., and R.K. Pace. 2009. Introduction to Spatial Econometrics. Boca Raton, Florida: CRC Press.

Lewis, D.J., B.L. Barham, and B. Robinson. 2011. Are There Spatial Spillovers in the Adoption of Clean Technology? The Case of Organic Dairy Farming. *Land Economics* 87 (2):250–267.

Meert, H., G. Van Huylenbroeck, T. Vernimmen, M. Bourgeois, and E. van Hecke. 2005. Farm household survival strategies and diversification on marginal farms. *Journal of Rural Studies* 21 (1):81–97.

Meulen, van der, H. A. B., M. Dijkshoorn-Dekker, J. Jager, H. Schoorlemmer, M. Schoutsen, E. Veen, G. Venema, M. Vijn, M. Voskuilen, and J. op de Weegh. 2014. Kijk op multifunctionele landbouw. Omzet en impact 2007-2013. Den Haag, Netherlands: LEI Wageningen UR.

Meraner, M., W. Heijman, T. Kuhlman, and R. Finger. 2015. Determinants of farm diversification in the Netherlands. *Land Use Policy* 42:767-780.

Mishra, A.K., H.S. El-Osta, and C.L. Sandretto. 2004. Factors affecting farm enterprise diversification. *Agricultural Finance Review* 64 (2):151 - 166.

Munshi, K. 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics* 73 (1):185-213.

Nyblom, J., S. Borgatti, J. Roslakka, and M.A. Salo. 2003. Statistical analysis of network data - An application to diffusion of innovation. *Social Networks* 25 (2):175-195.

Ostrom, E. 2002. Common-pool resources and institutions: Toward a revised theory. In *Handbook of Agricultural Economics*, eds. Gardner, B.L., and G.C. Rausser, 2A:1315-1339. North-Holland, Elsevier.

Pfeifer, C., R.A. Jongeneel, M.P. Sonneveld, and J.J. Stoorvogel. 2009. Landscape properties as drivers for farm diversification: A Dutch case study. *Land Use Policy* 26 (4):1106-1115.

Ploeg, van der, J.D., and D. Roep. 2003. Multifunctionality and rural development: the actual situation in Europe. In *Multifunctional agriculture: a new paradigm for European agriculture and rural development,* eds. Van Huylenbroeck G., and C. Durand, 37-54,. Aldershot, Hampshire, England: Ashgate.

Polman, N., K.J. Poppe, J-W. van der Schans, and J.D. van der Ploeg. 2010. Nested markets with common pool resources in multifunctional agriculture. *Rivista di economia agraria* 65 (2):295-318.

Rutten R. and F. Boekema. 2007. Regional social capital: Embeddedness, innovation networks and regional economic development. *Technological Forecasting and Social Change* 74 (9): 1834-1846.

Schmidtner, E., C. Lippert, B. Engler, A.M. Häring, J. Aurbacher, and S. Dabbert. 2012. Spatial distribution of organic farming in Germany: does neighbourhood matter? *European Review of Agricultural Economics* 39 (4):661-683.

Small, M. L. and K. Newman. 2001. Urban Poverty after The Truly Disadvantaged: The Rediscovery of the Family, the Neighborhood, and Culture. *Annual Review of Sociology* 27:23-45.

Storm, H., K. Mittenzwei, and T. Heckelei. 2015. Direct payments, spatial competition, and farm survival in Norway. *American Journal of Agricultural Economics* 97 (4): 1192-1205.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46:234-240.

Walford, N. 2001. Patterns of development in tourist accommodation enterprises on farms in England and Wales. *Applied Geography* 21 (4):331–345.

Weltin, M., I. Zasada, C. Franke, A. Piorr, M. Raggi and D. Viaggi. 2017. Analysing behavioural differences of farm households: An example of income diversification strategies based on European farm survey data. *Land Use Policy* 62:172-184

Wollni, M., and C. Andersson. 2014. Spatial patterns of organic agriculture adoption: Evidence from Honduras. *Ecological Economics* 97:120–128.

Annex

Forming of the hypotheses

Visual interpretation of clusters

Figure 4 shows the location of diversified farms in the Netherlands. Nature conservation activities show large clusters in the north (Friesland), in the middle of the country ('the Green Heart') and in the south-east (Zuid-Limburg). The largest tourism clusters emerged in the south-west (Zeeland) and on the islands in the North Sea (Wadden islands). Clusters of other activities are smaller (on-farm sale, on-farm processing) or non-existent (care farming). The largest processing clusters are found in the 'Green Heart', sale activities are more frequent in the south of the country. Care activities do not show clusters. Figures like Figure 4 have been made for the Netherlands by Meraner et al. (2015) and by Van der Meulen et al. (2014). The main difference from their figures to Figure 4 is that in this figure, the scale is the same for all activities. When normalizing for the frequency of diversified observations, the clusters in activities low in number are better visible. Here, however, colouring is the same for all activities and not normalized for the frequency of observed activities. This makes it possible to visually compare the strength of clustering of different activities without distortions. Figure 4 clearly shows the differences in cluster sizes of different activities. Care activities do not show clusters on this scale. However, as care activities are low in number, two care farms next to each other might be random but it looks like a cluster when normalizing for the number of farms. As care farming seems to be clustered in the figures of Van der Meulen et al. (2014), I came up with the hypothesis that clusters of care farming exist unless they are not visible here. The results of the research showed that this hypothesis was wrong.
Note: In Table 1 which gives an overview of the directions of influences in the hypothesis, the expected direction of the clustering coefficient can be found on the diagonal of diversification interactions.

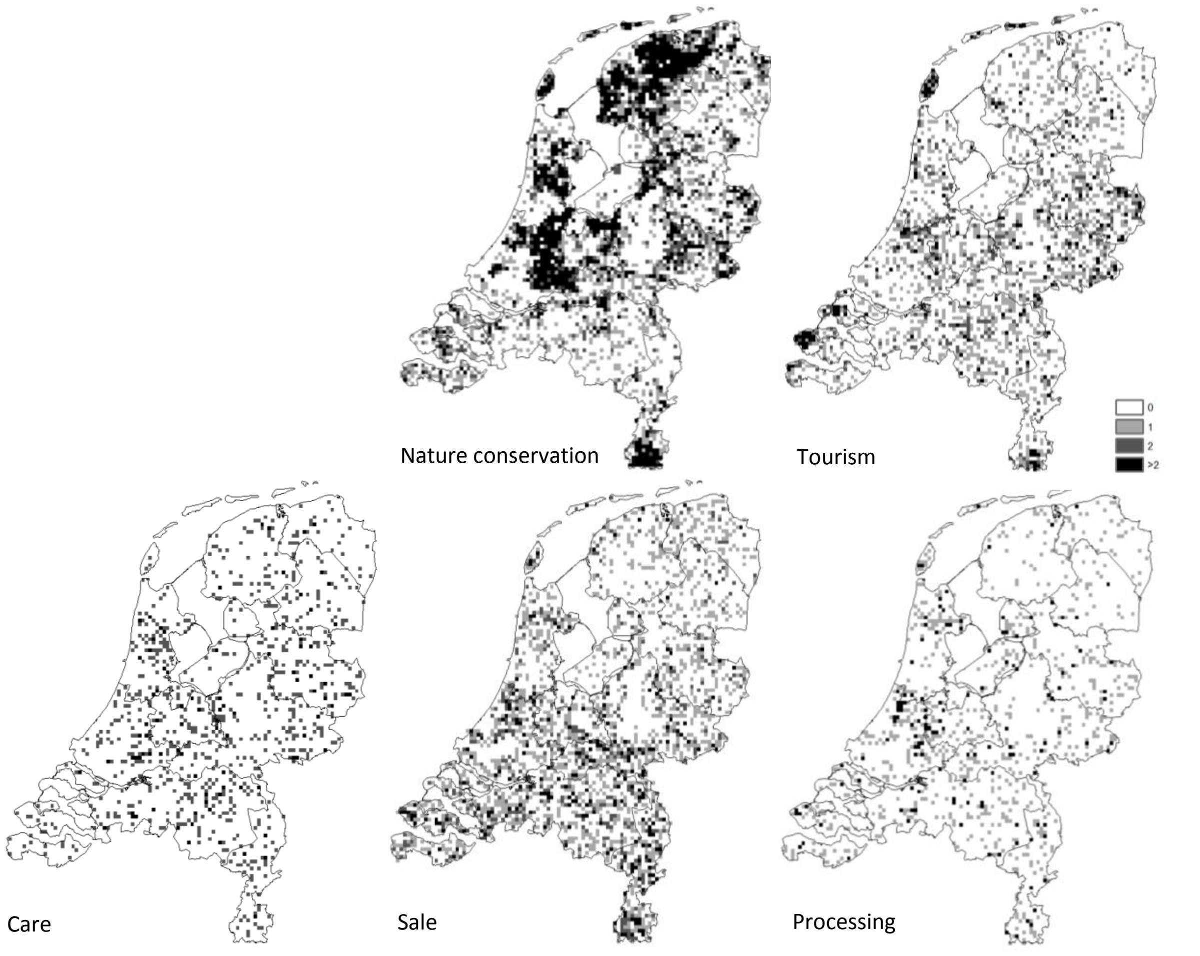

Figure 4. Clusters of diversification

Data

Dependent variables

The Dutch agricultural census includes more forms of diversified agriculture than captured by the Taskforce for diversified agriculture. The additional categories are 'processing' and 'aquaculture'. Processing is included in the work of Meraner et al. (2015) and is also relatively large in number in 2013 (1'041 farms). As processing activities are expected to be clustered especially in the Green Heart (Figure 5), processing is included in this analysis. In 2013, only 25 farmers said to be active in aquaculture. If these aquaculture farms are all located in the same region, aquaculture would be a good and easily understandable example to explain neighbourhood effects. To see if this is the case, the dataset was exported from R as a shapefile and mapped in ArcGIS. Unfortunately, as shown in Figure 5, no aquaculture farming clusters seem to exist in the Netherlands.

Figure 5. Locations of aquaculture activities in the Netherlands

Yields from diversified agriculture

The data set contains 8785 farms which have positive percentage yields from diversification but are not active in the selected diversification activities. These farms decided for other strategies (amongst others aquaculture, work for third parties and child education) which are seen as diversified by the CBS but not by the definition of this research. They are treated as non-diversified farms in the analysis.

Location

The data analysis showed that in some cases the farm's locations are duplicated. The dataset of 87'481 farms only showed 56'065 unique X-coordinates and only 57'558 unique Y-coordinates. A further data analysis show that there are 1'310 farm locations which are not unique. On these locations 2 or more farms are registered, leading to a total of 2'820 farms which share a location with at least one other farm (Table 8). When weights are assigned to links of a certain length, this is problematic because no weights can be assigned to links of length zero. It is not exactly known why farms share locations in the Netherlands, but it is assumed that farms sharing a location can occur for example when a farm is taken over by a son or daughter, but parents still keep a small part of the firm. This very close "neighbour" probably has a very high influence, but the communication with the rest of the neighbourhood is probably the same. In this case, it can be assumed that on many of these locations one of the farms is smaller. If this is true, it would make sense to only keep one farm per location and exclude the others. To see if this is true, the farm workforce is regarded. 1% of all Dutch farms have a workforce of less than 0.1 and 7% of the farms have a workforce of less than 0.2.

From farms on shared locations 1.2% of the farms have a workforce of less than 0.1, and 9.5% of the farms on shared locations have a workforce smaller than 0.2. As the difference in emergence of very small farms is not very large between all farms and farms on shared locations, excluding farms might not be an optimal decision. A more detailed analysis of the data showed that on 16 locations 5 or more farms are based. These 16 locations are all at the country's border with Germany or Belgium. This is also the case for most locations (30 from 41) on which four farms are based. This finding gives rise to another assumption why farms might share locations. It could be that in some cases farms are located (partly) abroad but keep an address in the Netherlands on the same location. As there might be two different reasons for shared locations, it is seen as the best solution to regard 'farm locations' instead of farms. Or differently said, regarding farms on shared locations as one large farm. Locational characteristics of farms on shared locations are the same by default. Workforce is summed up. For age the mean is taken and for farm type, all farm types are regarded. This means that very few farms in the dataset will have two farm types. The location is seen diversified if one or more of its farms are diversified.

Table 8. Shared farm locations

Farms sharing a location	Number of shared locations	Number of farms on shared location
2 or more	1316	2820
3 or more	117	542
4 or more	41	194
5 or more	16	136

One is located outside the Netherlands (Figure 6, encircled). For this farm, data from the agricultural census are known, but it's soil type, population density and possible neighbours in Belgium are unknown. This farm is thus excluded. After this data cleaning process 65'976 (67'481 -2'820 + 1'316 – 1) farm locations are left.

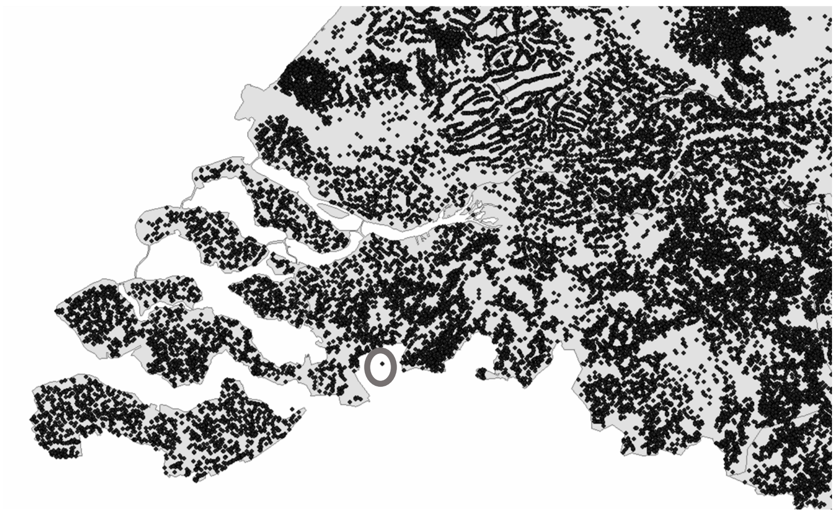

Figure 6. A farm located outside the Dutch territory.

Age

A data transformation of the variable age might be desirable because there is a chance that the distribution of the farmers age is not normal, but that there are more older farmers. This would result in a right-skewed distribution of the variable. Table 9 shows that taking the square root of the variable would lead to a minimal skewness of the distribution. However, for the 'no transformation' case, the skewness is rather low as well. Regarding the histogram of the variable (Figure 7, left), a slight right skewness can be detected. The qqplot (Figure 7, right), on the other hand, shows that the distribution is still close to normality. To not make interpretations more complicated than necessary, the variable age is used without any transformation.

Table 9. Statistics of different transformation of the variable age.

	Mean	SD	Median	Min	Max	Skewness	Kurtosis
No transformation	55.15	11.61	54.79	17	100	0.28	-0.17
Square	3176.28	1329.99	2916	289	10000	0.82	0.58
Logarithmic	3.99	0.22	3.99	2.83	4.61	-0.32	0.18
Square root	7.38	0.78	7.35	4.12	10	-0.01	-0.16

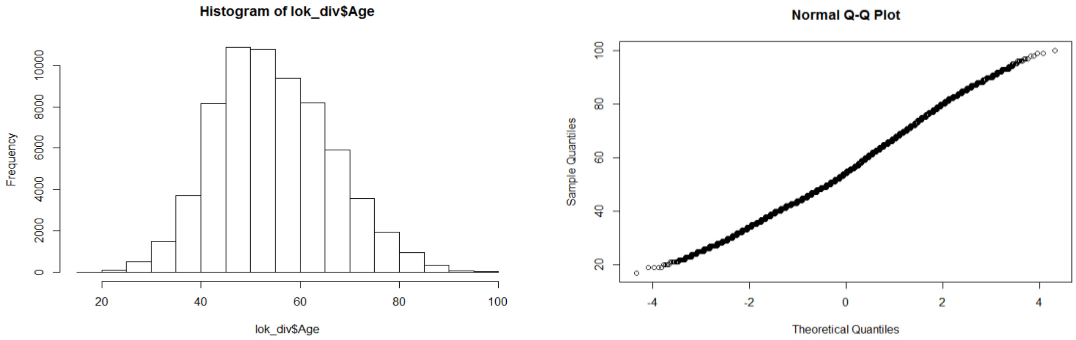

Figure 7. Histogram and qqplot of the variable age.

Farm size

Different measures can be used to estimate a farm's size. Storm et al. (2015) include three variables related to farm size, namely farm income, farm area and labour input. The authors conclude that "due to the high correlation of the three size variables, a model with just one of the three size variables has approximately the same explanatory power as the full model". The correlation of workforce and SO is 0.61 in the Netherlands. For the high number of observations this correlation is clearly significant. Additionally, Storm et al. (2015) have found that "despite the fact that all three variables are highly correlated, the large sample size if sufficient to uniquely identify coefficients for the three variables". Available farm size measures in the Netherlands include area, economic outputs and labour inputs. A farm's area is not regarded suitable, since the area of a farm is highly correlated with the farm type. The standard economic output (SO) is used in the Netherlands to estimate the economic farm size. SO's are defined as the average yield of a product in the last five years measured in Euros per ha or per animal. SO's can change from year to year. For example, in the year 2015 the SO of a ha summer wheat was 1'190 Euro's and the SO of a milk cow was 2'880 Euros. Yields from diversification activities are not included in the SO's (Van Everdingen, 2015). When multiplied with a coefficient capturing the percentage of the yields which is left after standardized costs of the output of the SO, the standard earning capacity (SVC) of a farm is determined. Compared with the SO, the SVC allows a better comparability of economic farm sizes between different farm types. Yields from diversification activities are not included in the SO's and thus not in the SVC either (Van Everdingen, 2015). Therefore, a relation between economic size and diversified agriculture emerges from the data structure (diversified farms will have lower incomes because not all their income is counted). This measure is therefore not very suitable for this research. For each farm, the percentage of income from diversification activities is known. This variable contains 4 classes: <10%, 10-30%, 30-50% and >50%. An aggregation of SO's or SVC with the income from diversification activities might be interesting to get a more suitable measure for the economic farm size. However, aggregating the percentage of income from diversification activities to correct for the fact that diversification is not included in the SO, is not appropriate for two reasons. Firstly, the categories are very imprecise. Secondly, SO is not the same as farm income.

Regarding workforce as a size variable in research on diversification can be problematic since diversification might also have influence on the workforce. We tested if this effect is present in the Netherlands with a paired sample t-test of farm data from 2011 and farm data from 2013. In three separate tests, we included all farms that diversified between 2011 and 2013, that stopped their diversified activities, and looked at both types of changed strategies together. The results presented in Table 10 show that we found no indication that diversification influences the workforce on Dutch farms.

Figure 8 shows the distribution of SO's, which was regarded before the decision to exclude SO's was made. The figure shows that a logarithmic distribution of SO's would have been preferred. Based on the findings of Storm et al. (2015) and the unsuitability of other measures, workforce is the only size variable included in the analysis.

Table 10. Paired sample t-test of workforce.

Group	Mean of difference	p-level
Started diversification	-0.051	0.570
Stopped diversification	-0.119	0.351
Taken together	-0.080	0.230

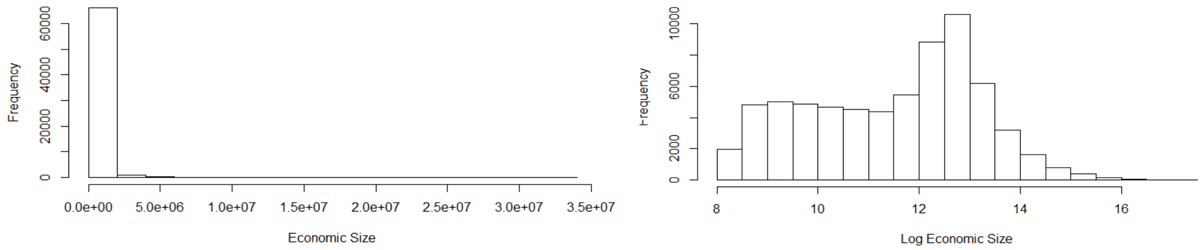

Figure 8. Distribution of SO's.

Urbanisation

As closeness to urban centres is found to have influence on diversification (Meraner et al., 2015; Hassink et al., 2016), this spatial attribute should be included. Otherwise, neighbourhood effects might be overestimated. Urbanisation can be captured in different ways. Possibilities include the spatial distance to an urban centre with a certain population minimum (Ilbery 1991; van Leeuwen et al., 2013), population density (Meraner et al., 2015) or accessibility (Pfeifer et al., 2009). In this research on the Dutch situation population density was used to capture urbanity. The main reason for including urbanity in this way, is because it is the most straightforward option. Both distance to an urban centre and accessibility need new definitions, making the conceptual model more difficult to understand. Data on population density in each municipality is made available by the Central bureau for statistics (CBS) and is open-source. Figures 9- 10 show that a logarithmic use of the population density is preferable, as the distribution is more like a normal distribution when using the logarithm.

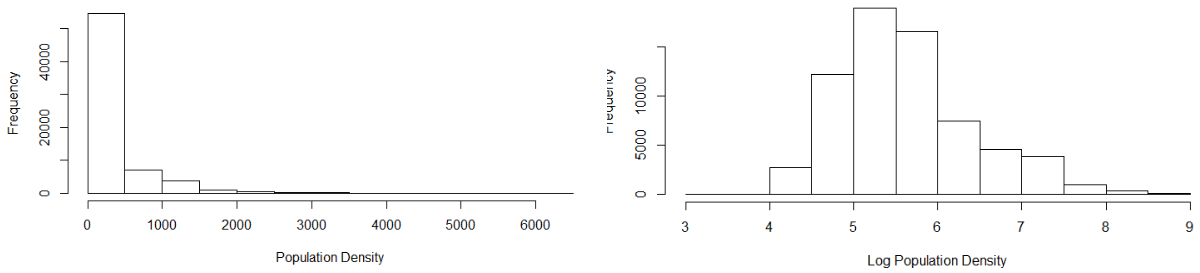

Figure 9. Distribution of population density.

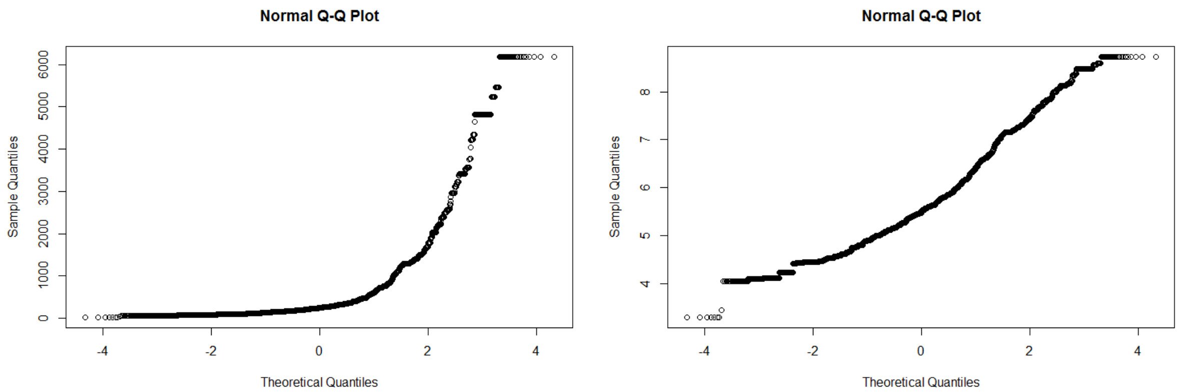

Figure 10. Q-Q Plots of population density (left) and logarithmic population density (right).

Soil type

Due to border artefacts, 69 farms were not assigned a soil type after combining the farm's location with the soil map. This problem is caused by the fact that the border line of the soil map has relatively few vertexes which causes smooth borders. A visual extrapolation was done to assign soil classes to these farms. Figure 11 shows a part of the soil map in a border region in the south of the Netherlands. grey areas indicate two different soil types, the black line represents the borders of the Dutch municipality map which is less smooth. The figure shows that some farms (red dots) are located in the Netherlands but "fell of" the soil map.

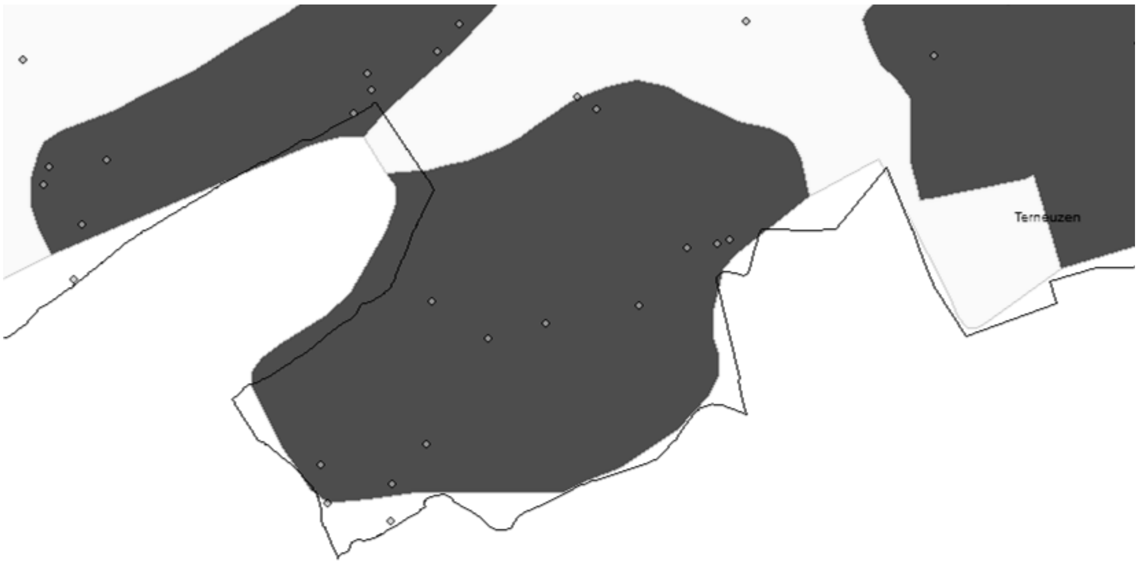

Figure 11. The soil map (tones of grey), the municipality map (black line) and farms (dots) in a border region in the south of the Netherlands.

The soil type "urban grounds" is of special interest. When the hypothesis that farms diversify more when they are close to cities is true, farms on the soil type 'urban grounds' should diversify more. Or, when the hypothesis is true that sale activities occur more close to cities, there should be more on-farm sale activities on farms located on urban grounds. The variable population density could be extended with data from the soil map. Farms located within urban areas could for example get a "bonus" score on population density.
The result of an independent t-test does not show evidence for any of these assumptions (Table 11). There is no significant difference in the amount of farms on urban grounds between diversified and non-diversified farms, also not for on-farm sale activities. However, when only regarding nature-conservation farms, there is a significant difference: farms located on the soil type "urban grounds" are less active in nature-conservation activities than farms on other soil types. There are also less

care farms on urban grounds, but this is not a significant difference. Farms on urban grounds probably have less possibilities to conserve nature as nature is simply further away. Being outside a city might be an important characteristic for the attractivity of care farms.

Table 11. T-test results.

	mean		p-value
	urban grounds	other	
Diversification	0.188	0.192	0.743
Care	0.008	0.013	0.060
On-farm sale	0.049	0.047	0.763
On-farm processing	0.017	0.016	0.742
Tourism	0.047	0.042	0.352
Nature conservation	0.095	0.112	0.036

Intensification

Agricultural intensification in the farm's region was assumed to be another possible factor influencing diversification. This factor would influence diversification negatively, since a diversified farm might not be competitive in regions with intensive agriculture. In order to include this factor as another independent variable, a region's intensification score is needed. This raises multiple questions: on what scale (municipality, region, province) does intensification influence diversification? How many intensified farms are needed in a region to have influence on diversification? Is there a difference in influence depending on the farm type of intensive agriculture? To this comes that farm intensification is not independent from the factor landscape attractiveness. In a region that is intensively used for agriculture, the landscape might be less attractive. In many cases this might mean that there is less open green area, or that there are fewer forests and water bodies. Therefore, the farm will have a low Reilly-index score in this research. Additionally, intensive farming is to some extent already captured with the farm type 'intensive-livestock farming'. Therefore, it was decided not to include agricultural intensification as an additional variable in this research.

Creation of the Regional Reilly-index for landscape attractiveness

Many spatial datasets of natural areas in the Netherlands exist, with often different definitions of natural areas. The most relevant datasets for the creation of a variable capturing the landscape attractiveness include the map of national parks, the map of woods of "Staatsbosbeheer" (the Dutch government organisation for forestry and management of nature reserves) and the nature map ("Natuur op kaart") from the portal for nature and landscapes ('Portaal natuur en landschap"). Here, the latest (2010) land use map of the CBS is used to create a variable of landscape attractiveness. The CBS has classified the whole country into different land use areas. It includes three different categories of nature and 13 different categories of water bodies. We decided to include the landscape attractiveness by use of the Reilley-index (see paper) based on areas of and distances to areas defined as natural by the land use map. To create a Reilly-index for landscape attractiveness, both nature and water bodies are considered. Nature includes forests, dry open natural areas and wet, open natural areas. The 13 water categories include categories for the large salt water bodies (Noordzee, Waddenzee, Oosterschelde, Westerschelde) as well as for the large fresh water bodies (IJsselmeer/Markermeer, Grevelingen and Haringvliet, Rijn and Maas, Randmeer). Other water categories are reservoirs, recreational internal waters, internal water for resource mining, industrial used water bodies and others. The category "others" includes most of the lakes, rivers and ditches. Here, the large salt and fresh water bodies as well as recreational and other water bodies were included to create an "attractive landscape" map.

The land-use dataset from the CBS includes many single polygons (Figure 12) within one larger natural area. For example, when a small road or an agricultural parcel divides the natural area into multiple pieces, this does not necessarily break an attractive area into individual pieces. The calculation time, however, increases strongly with the number of polygons. The total number of polygons was thus decreased to aggregating polygons of divided natural areas. As the shapes from water bodies and natural areas differ, two different methods were used to reduce the amount of polygons. Natural areas polygons which are situated on less than 200 m distance from each other were aggregated. This was done by making a 100 m buffer around all polygons (Figure 13). All the intersecting areas were then connected to one larger polygon. Afterwards the buffer was deleted. This resulted in a decrease in the total number of polygons and a reclassification of certain land covers. As shown in Figure 14, roads and agricultural areas surrounded by nature are now classified as nature. This increases the total natural area. However, roads and agricultural areas within a forest might not make the forest less attractive but maybe even increase landscape attractiveness since nature is made more accessible. Therefore, the addition of these areas is not seen as problematic. A visualisation of this process is provided in the Figures 12-14, where the darkest areas represent forests and the lightest represent dry open natural areas.

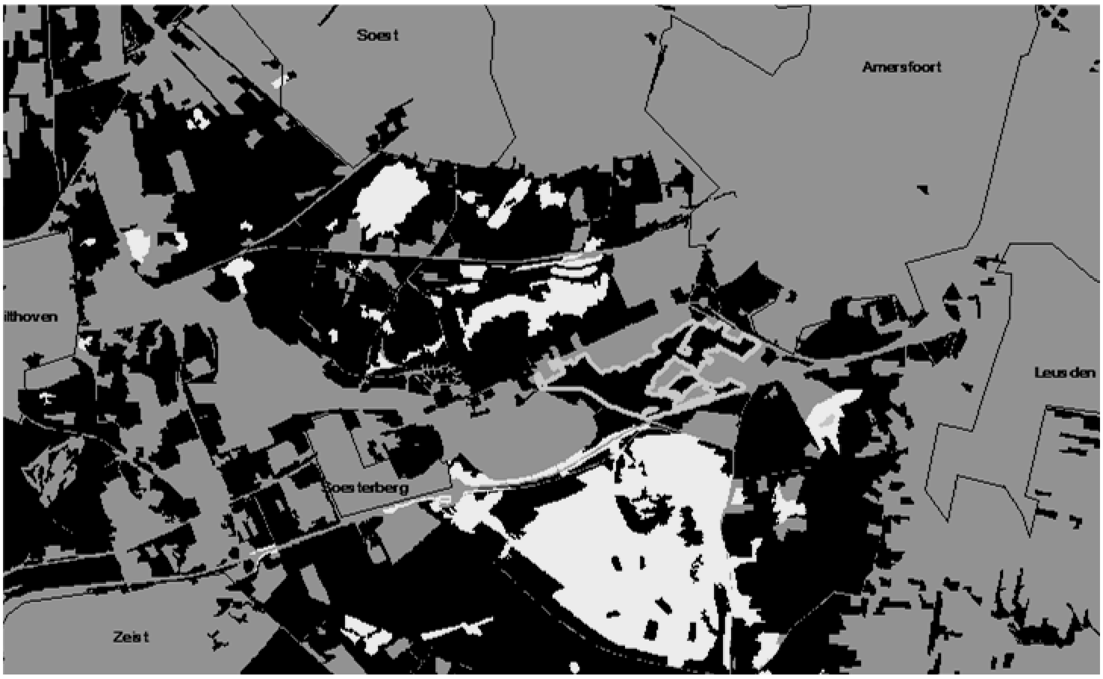

Figure 12. Original data: the natural area around the Soesterduinen is divided into many small natural areas.

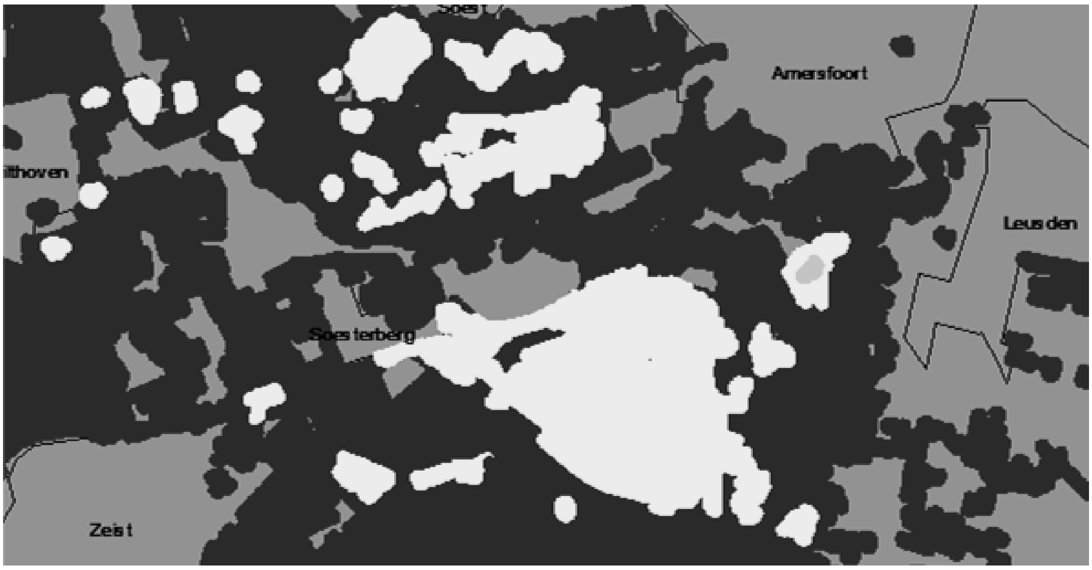

Figure 13. A 100 m buffer enlarges natural areas.

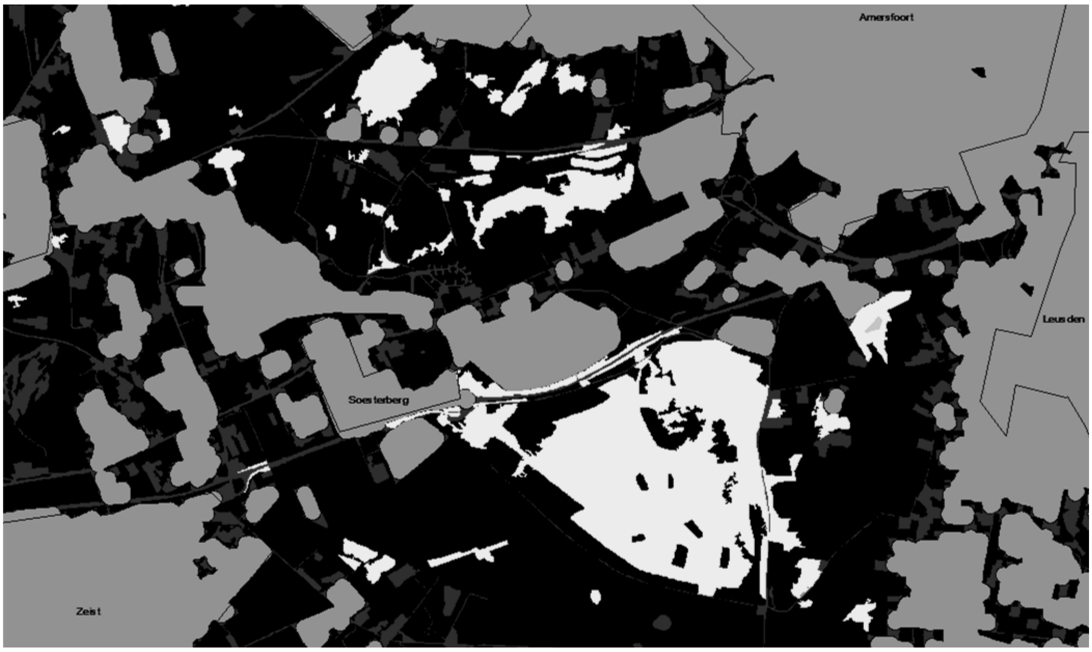

Figure 14. Edited data: the natural area around the Soesterduinen is a connected natural area.

Even after these computations, many very small and remote natural areas remain. These are often single, very small areas of forest which do not necessarily contribute much to the regional landscape attractiveness. For this reason and in order to further reduce calculation time, a minimum area was set. Areas smaller than 10 ha were deleted from the data. Dark areas in Figure 15 show areas which are excluded in the middle of the country, the area of the Soesterduinen is highlighted with the rectangle. The lighter grey areas are the remaining natural areas.

Figure 15. Darkest areas are excluded national areas, the rectangle shows the area Soesterduinen.

As water data is classified in several categories, it is possible to exclude water bodies which do not increase landscape attractiveness, namely reservoirs, internal water for resource mining and industrially used water bodies. The large salt and fresh water bodies are each saved as single large polygons. The remaining categories are "recreational internal water" and "others". These classes include again many polygons. A polygon reduction of water is however, more difficult than with nature areas since characteristics of water bodies differ from characteristics of natural areas. Especially in the case of small rivers and ditches, buffer aggregation does not make sense (the whole

area between the parallel ditches would be classified as a water body). Therefore, only areal minima were used. To exclude small ditch systems, a rather high minimum (15 ha) was used for “others”. As recreational water is by definition making the landscape attractive, a lower minimum (5 ha) was used for this class. With the minimum of 5 ha, town channels were excluded.
By merging the natural areas map and the water bodies map, a binary attractive landscape map was created (Figure 16). White areas indicate non-attractive landscapes, grey areas indicate attractive landscapes.

Figure 16. Attractive landscapes in the Netherlands.

To get from this map to a Reilly-index for all the farms, a few more steps are needed. First, the sum of distances from every farm to all the natural areas within a certain distance is calculated (with a near table in ArcGIS). Second, the sum of areas of nature which have a part of their area within a certain distance is used. When both factors are known for every farm, an individual Reilly-index score is assigned. As relative values are used, it is not crucial how large the radius is. To get an idea of how large the buffer around the farm could be, a radius of 5 km as proposed by Polman et al. (2010) was used. This buffer was compared with a smaller (3 km) buffer. Figure 17 shows again the natural area Soesterduinen in the south of Amersfoort. The circles show that 5 km buffers are relatively large compared to a middle-sized natural area in the Netherlands. However, 5 km is a distance to travel easily, as it is only a short bike trip. Therefore, a 5 km buffer was used. Figure 18 shows the resulting Reilly-index score of every farm in the Netherlands. Our adoptions in the measurement of natural areas have led to the fact that some farms are within a natural area. Their distance to natural areas is thus zero. The deviation by their distance of zero was no longer possible. Therefore, these farms were assigned the maximum Reilley-score reached by other farms.

Figure 17. 5 km buffers (left) and 3 km buffers (right) around farms.

Figure 18. Reilly-index scores with water bodies.

Figure 18 shows farms which are in an attractive landscape (dark) and farms in a less attractive landscape (light). With this classification, farms around the large natural area the Veluwe got a high score, also farms at attractive coastal sites score high (Waddenzee, Zeeland). However, farms in regions with many small natural areas (Zuid-Limburg, Achterhoek) score low and farms at unattractive coastal sites (for example Flevoland) have high scores. This results from the fact that some of the water bodies are disproportionally large and farms on less than 5 km distance from these waters get a very high score. Therefore, a second Reilly-index map as presented in Figure 3 in the paper was created. This time, all water bodies were excluded. The main difference is that farms in the coastal zones of Zeeland, Noord-Holland and Friesland score lower. Sites at attractive coasts still get high scores, due to the fact that they are close to the dunes. From this picture, it looks as if some farms got a higher value. However, this is not true but follows from a smaller distribution and therefore class borders were set a little different.

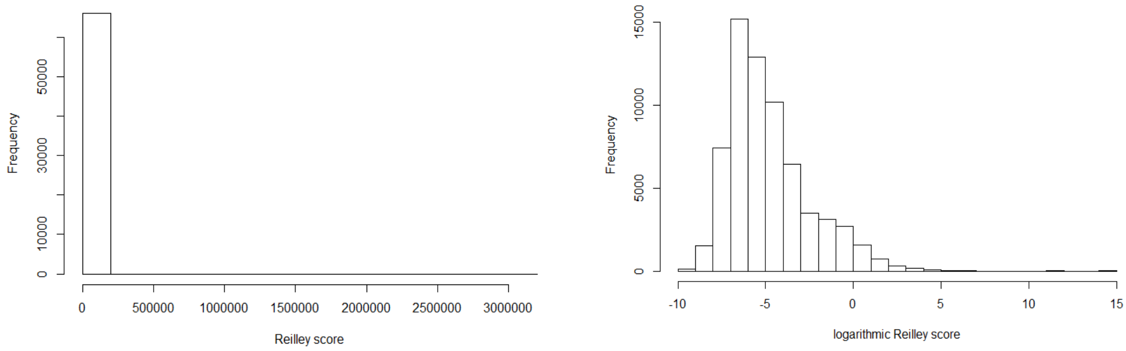

Figure 20. Histograms of Reilly-index score distribution**.**

Based on the distribution of the scores on the Reilly-index (Figure 20), a logarithmic use of the Reilly-index is preferred. The logarithmic Reilley-index has infinite small values which the spatial regression cannot handle. These values were set to -10, which is very small in a logarithmic scale. Figure 21 shows the boxplot of the remaining values. In the first analysis, the index was used like this. To check for the impact of the outliers, all values larger than two were set to two in a second attempt and the regression was conducted for nature conservations. Resulting estimates were slightly different, however, directions and significances did not change. Therefore, the results of the first attempt were not replaced.

The Reilley-index had quite some outliers. To check how this influences the result, an attempt for nature conservation was done, in which all logarithmic Reilley-scores higher than two were set to two. As can be seen from Table 12, only minor changes occurred.

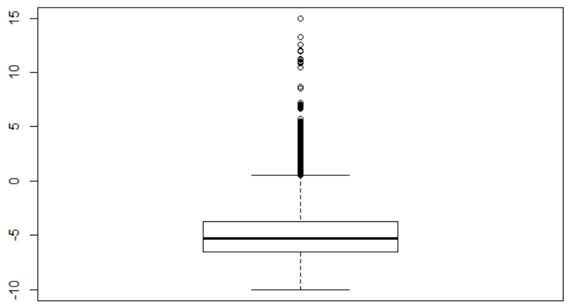

Figure 21. Boxplot of logarithmic Reilley.

Table 12. Results with max 2 for log Reilley.

	Estimate	**Std. Dev**	**p-level**
(Intercept)	-0.58159	0.073106	0
Age	-0.00768	0.000606	0
Workfrc	0.002937	0.00113	0.01
logpop	0.002234	0.007336	0.382
logReilley	-0.00203	0.002346	0.195
Dumhort	-0.25253	0.049989	0
Dumpast	0.697434	0.04219	0
Dumarab	0.386275	0.044242	0
Dumhok	-0.03504	0.056787	0.255
Dummix	0.484058	0.055182	0
Dumzkle	-0.05089	0.024934	0.017
Dumrkle	-0.0904	0.025602	0
Dumveen	0.020128	0.025358	0.225
Dumplei	-0.26668	0.023042	0
tourism	0.366153	0.030741	0
sale	0.352767	0.035062	0
care	0.388017	0.052275	0
prcssng	0.183903	0.051767	0
rho	0.599645	0.00742	0

Other aspects concerning dependent variables

Table 13. Summary statistics for specific activities.

	Nature conservation (n=7371)		**Tourism (n=2772)**		**Sale (n=3113)**		**Processing (n=1034)**		**Care (n=868)**	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev
Structural characteristics										
Age	53.74	10.92	53.52	10.26	52.62	10.09	51.66	9.77	51.74	9.07
Workforce	1.96	1.49	2.36	2.63	3.12	4.53	3.63	4.59	2.78	2.66
Farm type										
Past	0.79	0.41	0.64	0.48	0.30	0.46	0.44	0.50	0.70	0.46
Arab	0.13	0.34	0.17	0.37	0.16	0.37	0.17	0.38	0.05	0.22
Hort	0.01	0.11	0.07	0.25	0.23	0.42	0.15	0.36	0.07	0.26
Inte	0.02	0.15	0.03	0.18	0.07	0.25	0.03	0.18	0.06	0.24
Pera	0.01	0.10	0.04	0.20	0.13	0.34	0.12	0.32	0.03	0.17
Mix	0.04	0.20	0.07	0.25	0.12	0.32	0.11	0.31	0.09	0.29
Locational characteristics										
LogPop	5.54	0.76	5.59	0.83	5.73	0.82	5.68	0.83	5.73	0.86
LogReilley	-4.89	2.42	-4.18	2.81	-4.74	2.30	-4.63	2.39	-4.94	2.17
Soil type										
Sands	0.32	0.46	0.50	0.50	0.43	0.50	0.38	0.49	0.49	0.50
Clays	0.41	0.49	0.34	0.47	0.42	0.49	0.45	0.50	0.32	0.47
Peat	0.22	0.42	0.10	0.30	0.08	0.28	0.12	0.32	0.14	0.35
Loam	0.03	0.18	0.03	0.17	0.04	0.19	0.03	0.16	0.02	0.13
Others	0.02	0.13	0.03	0.17	0.03	0.17	0.03	0.16	0.04	0.20
Diversification interactions										
Nature	1.00	0.00	0.23	0.42	0.16	0.37	0.23	0.42	0.26	0.44
Tourism	0.09	0.28	1.00	0.00	0.17	0.38	0.27	0.44	0.24	0.43
Sale	0.07	0.25	0.19	0.39	1.00	0.00	0.65	0.48	0.21	0.41
Processing	0.03	0.17	0.10	0.30	0.21	0.41	1.00	0.00	0.09	0.29
Care	0.03	0.17	0.08	0.27	0.06	0.24	0.08	0.27	1.00	0.00

Neighbourhoods

Computations

To get a feeling for a good neighbourhood decision, neighbourhoods where created within municipalities. On the municipality level the number of farms and thus links between farms is limited. Three municipalities where chosen: Amersfoort as an urban municipality (72 farms, Figures 21-24), Barneveld since it is famous for its clustered chicken farms with small farm areas (645 farms, Figures 25-28) and Winsum as a remote area (157 farms, Figures 29-32). This view is limited, since farms on the border of municipalities will not only have links within the municipality, but also with farms in other municipalities. Therefore, link densities are too high in the following figures. However, the figures are useful to indicate differences between various neighbourhood definitions. Figure 21 , 22, 25, 26, 29 and 30 show that with a k-nearest neighbours definition, isolated farms and isolated communities are very rare. With an increasing number of k, the density in links increases. Especially intra community links increase. The links between communities do not increase relatively less. This explains why the results are robust when changing k. The Figures 23, 24, 27, 28, 31 and 32 show that with a small radial definition of neighbours, isolated communities exist. In a flat, modern country as the Netherlands, isolation is not very likely. When increasing the radius, some farm networks become very dense. Everyone knowing everyone might not be a realistic situation either and needs a great computation time.

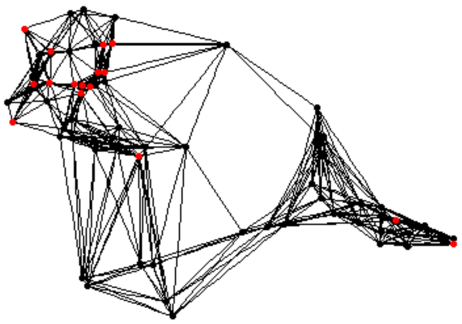

Figure 23. 10 nearest neighbours Amersfoort.

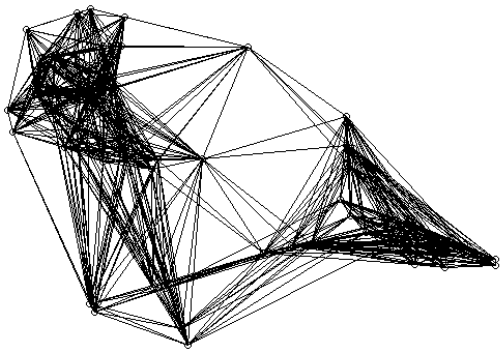

Figure 22. 20 nearest neighbours Amersfoort

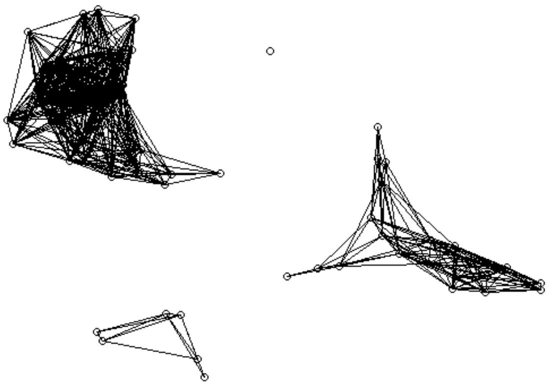

Figure 25. 2 km radial neighbours Amersfoort.

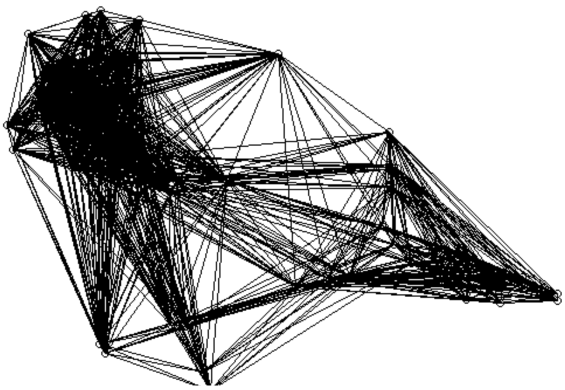

Figure 24. 5 km radial neighbours Amersfoort.

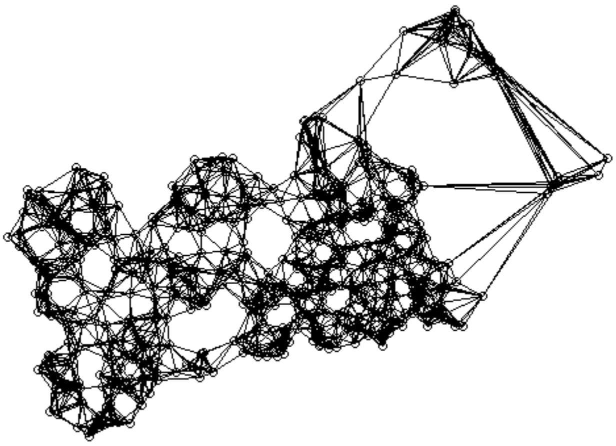

Figure 27. 10 nearest neighbours Barneveld.

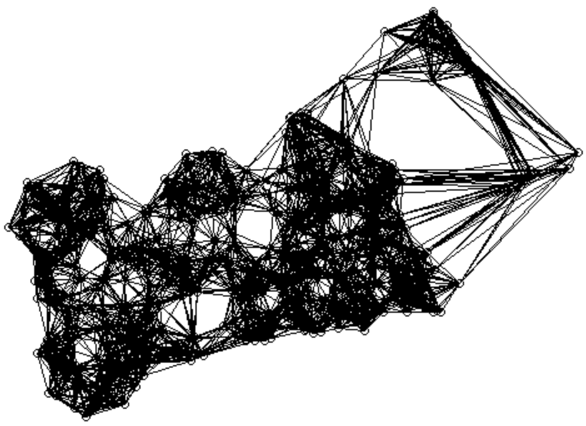

Figure 26. 20 nearest neighbours Amersfoort.

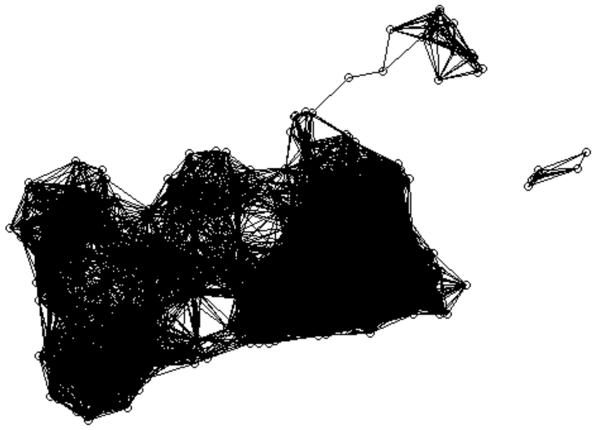

Figure 29. 2 km radial neighbourhood Barneveld.

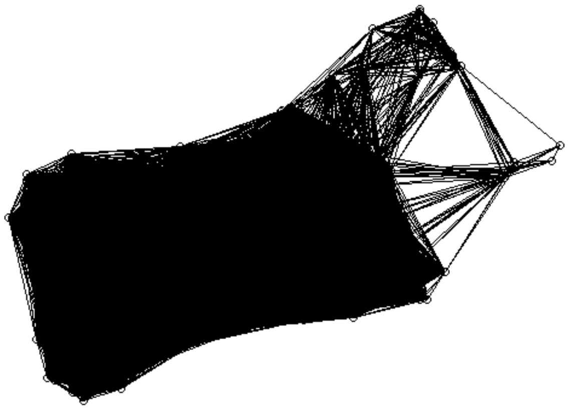

Figure 28. 5 km radial neighbourhood Barneveld.

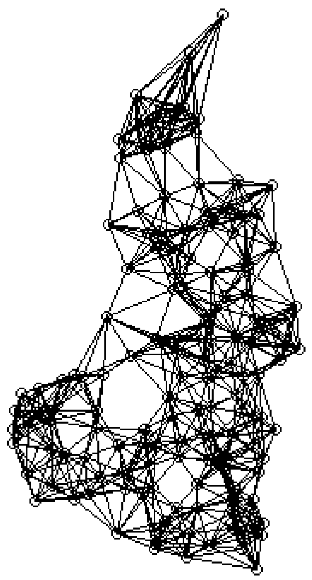

Figure 31. 10 nearest neighbours Winsum.

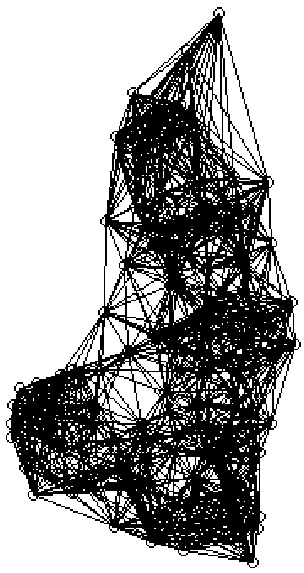

Figure 30. 20 nearest neighbours Winsum.

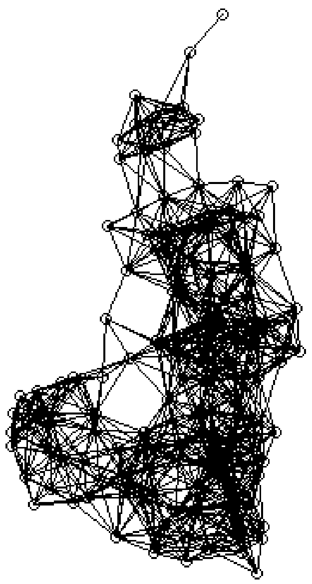

Figure 33. 2 km radial neighbours Winsum.

Figure 32. 5 km radial neighbours Winsum

From Figures 23-34 it is visible that in many agricultural regions, a radial neighbourhood of only five kilometres (15-20 min cycling) farms can have very high numbers of neighbours (in Barneveld all 645 farms are connected in the 5 km case). When reducing the radius, this might lead to the fact that in some areas with high farm densities farms have still many connections (for example Barneveld with two kilometres radii) and others might be in isolated neighbourhoods (for example Amersfoort with two kilometres radii). With a k-nearest neighbourhood, there is an exact number of neighbours a farm is influenced by and influences are not dependent on link densities. Based on the findings in the Figures 21 to 32, a k-nearest neighbourhood is found to better represent the situation in the Netherlands than a radial neighbourhood definition.

Figure 33 shows how the neighbour distance distribution changes with the number of k-nearest neighbours included. First, the mean distance to neighbours is smallest when only two neighbours are included and the variance increases as k is increased. The furthest nearest neighbours are similar for the different k's. This is because of a single farm on Vlieland which has all its neighbours at about twelve kilometres distance. When excluding the farm on Vlieland and other outliers, the difference in distance to the furthest neighbour increases with the number of neighbours included, as expected (Figure 33, right). The size of the storage space needed is 7.3 mb for two neighbours and 15.2 mb for 20 neighbours.

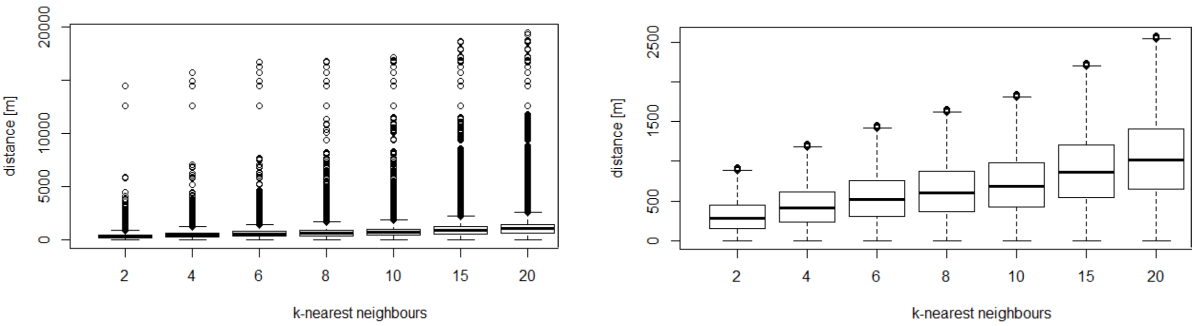

Figure 34. Distance to neighbours, with (left) and without outliers (right).

With the package ggmaps in R it is possible to plot networks on Google Maps. Different network models can be plotted quickly. Some examples are shown in the Figures 34-39. In these examples, diversified farms are marked grey. There are clusters of diversification visible. If there was perfect spatial autocorrelation, all farms within one cluster should be grey and all farms in another cluster should stay black. Figures 34-35 show the province of Utrecht and are zoomed in at Amersfoort and show that autocorrelation is clearer when only looking at one type of diversification (Figure 35) than at diversification in general (Figure 34). Figures 36-39 show the province of Zuid-Holland, zoomed in at Alphen aan de Rijn and show how networks change when more neighbours are added. The major change is that there are more links between clusters when the number of neighbours is higher, but the network structure changes not that much. Based on this finding, a low dependence of the result on changes in k is expected. It is important to note that links are directed: if central farm j is one of the nearest neighbours of farm i, farm i might not be one of the nearest neighbours of farm j.

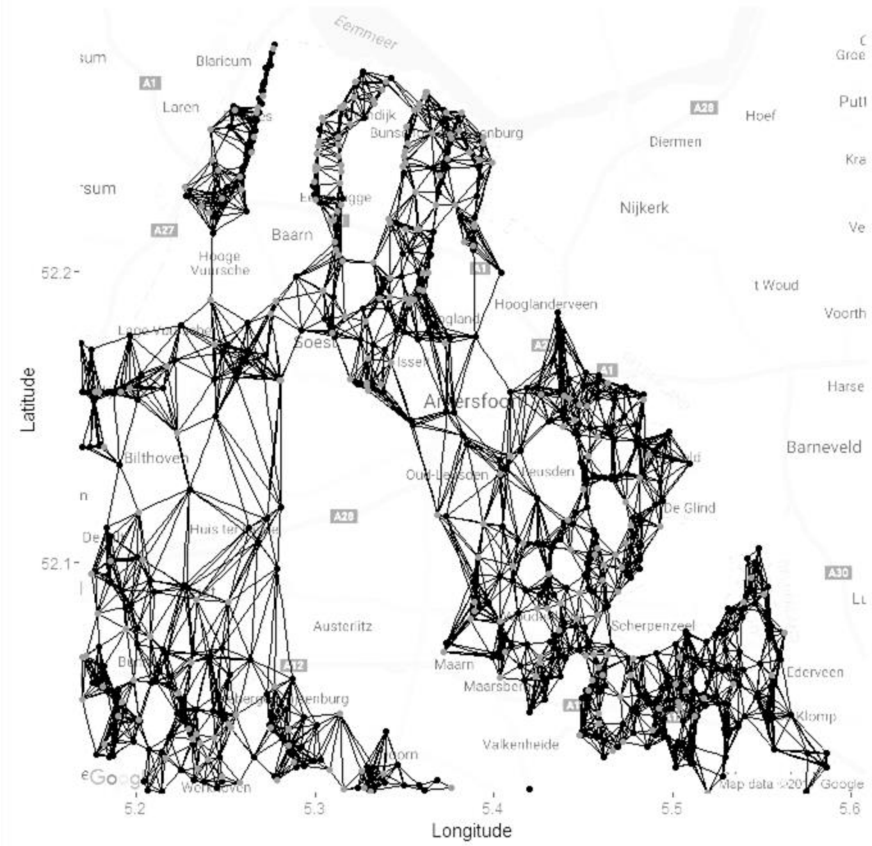

Figure 35. 10 nearest neighbours network Utrecht, diversified farms marked grey.

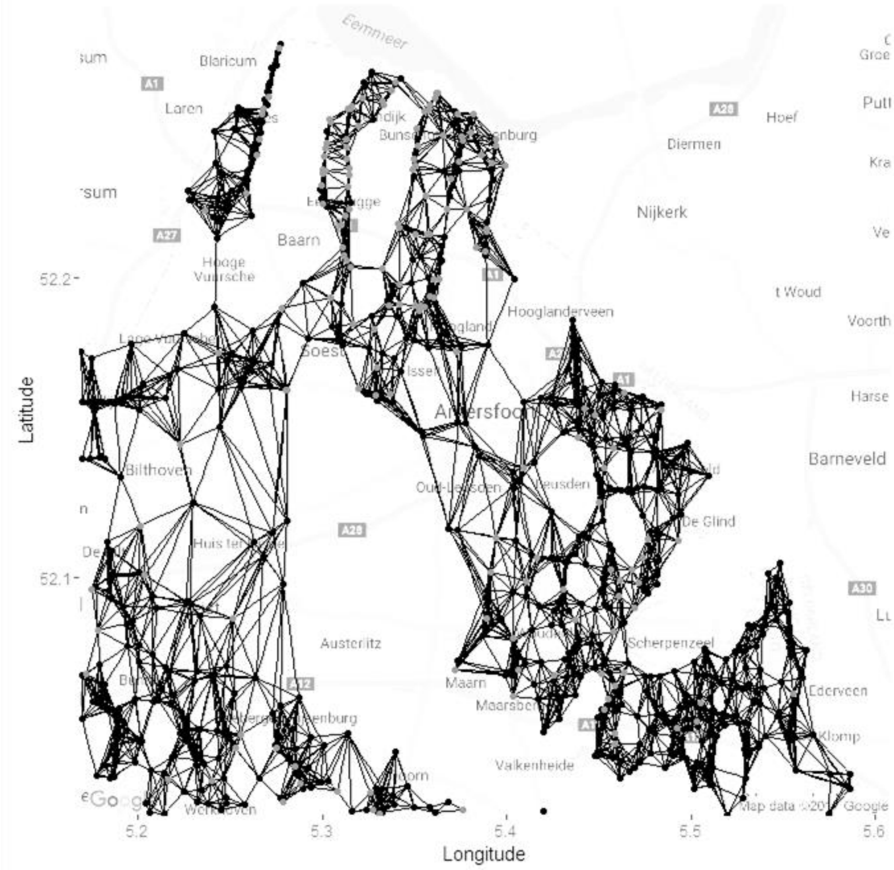

Figure 36. 10 nearest neighbours network Utrecht, nature conservation farms marked grey.

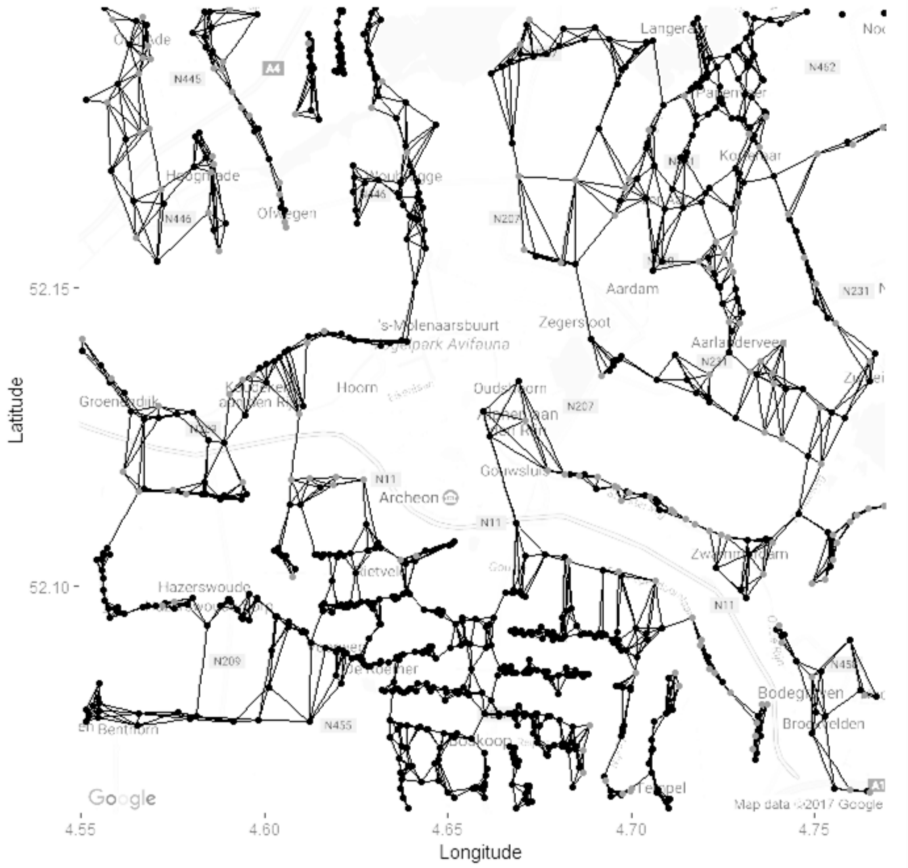

Figure 37. 5 nearest neighbours network Zuid-Holland, nature conservation farms marked grey.

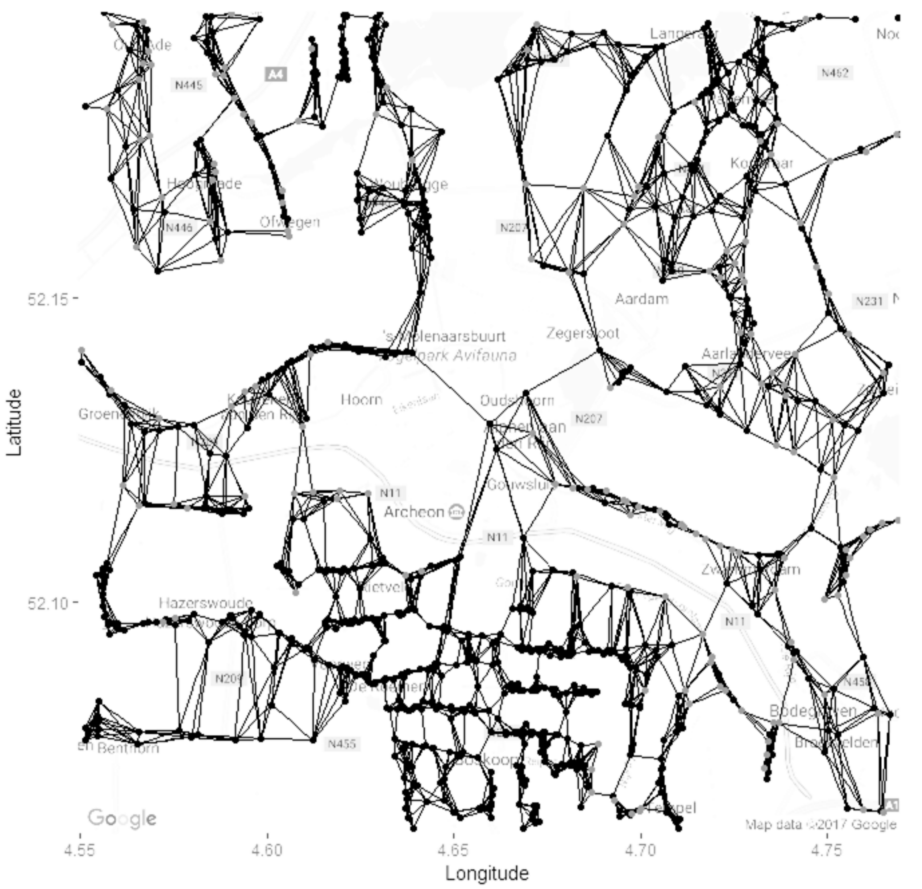

Figure 38. 7 nearest neighbours network Zuid-Holland, nature conservation farms marked grey.

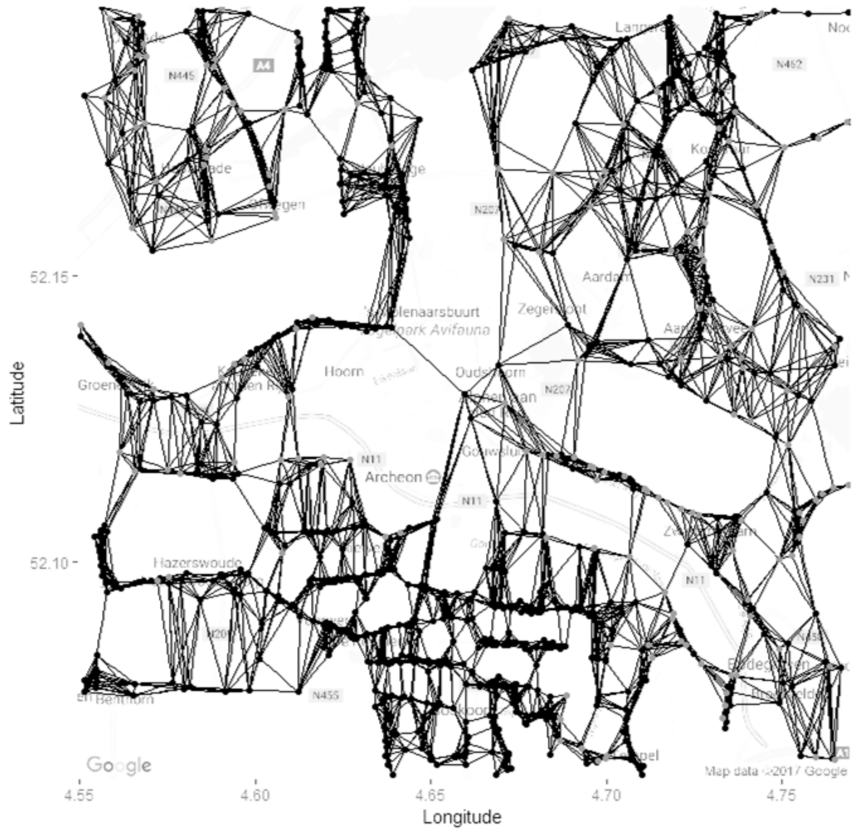

Figure 39. 10 nearest neighbours network Zuid-Holland, nature conservation farms marked grey.

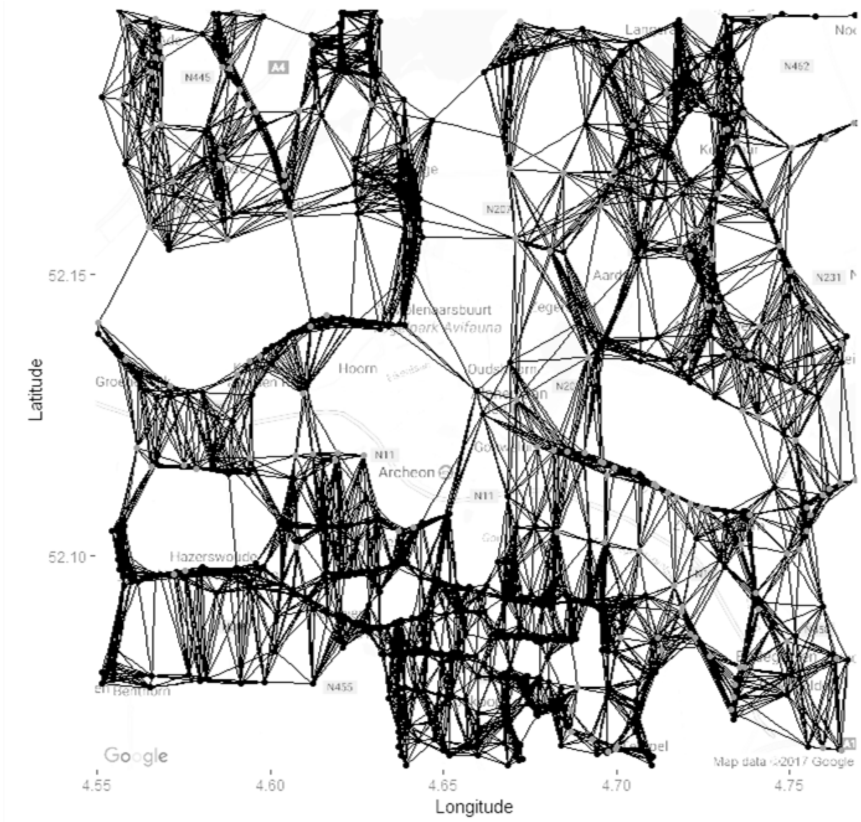

Figure 40. 15 nearest neighbours network Zuid-Holland, nature conservation farms marked grey.

The number of farms active in each diversification activity differs. It depends on the activity how many of the ten neighbours need to be active in the same activity before clustering emerges. For example, on average every farm has one neighbour which is active in nature conservation and every second farm has a neighbour which conducts sale activities. This means that sale activities can be clustered when a great number of the farms have clearly more and less than 0.5 active neighbours. With nature conservation, the deviation from having one nature conserving neighbour is important.

Table 14. Overview nearest neighbour statistics.

	Total number of farms	Expected percentage of neighbours with this activity	Standard deviation
Nature conservation	7371	0.11172	0.315
Tourism	2772	0.04202	0.2006
Sale	3113	0.04718	0.212
Processing	1034	0.01567	0.1242
Care	868	0.01316	0.1139
Diversification	12650	0.19174	0.3937

Sensitivity test

To test if results are really as robust with changing k's as expected, the analysis was done for nature conservation for the 157 farms in Amersfoort and Leusden with 10 and with 100 neighbours. Results for marginal effects are presented in Tables 14-15 and show no major deviations to each other.

Table 15. Marginal effects in Amersfoort and Leusden with 10 neighbours.

	Direct	Indirect	Total
Age	-0.00801	-0.00086	-0.00887
Workfrc	0.021602	0.002311	0.023913
logpop	0.04943	0.005289	0.054718
logReilly	0.028515	0.003051	0.031567
Dumhort	0.177718	0.019015	0.196733
Dumpast	0.188564	0.020175	0.20874
Dumarab	0.098345	0.010522	0.108868
Dumhok	-0.00657	-0.0007	-0.00727
Dumpera	0.416338	0.044546	0.460884
Dummix	0.307982	0.032953	0.340935
Dumzkle	0.472215	0.050525	0.522739
Dumveen	0.045246	0.004841	0.050088
Dumplei	0.243234	0.026025	0.269259
tourism	-0.03507	-0.00375	-0.03882
sale	0.221657	0.023716	0.245373
care	0.702538	0.075168	0.777707
prcssng	-0.45447	-0.04863	-0.5031

Table 16. Marginal effects in Amersfoort and Leusden with 100 neighbours.

	Direct	Indirect	Total
Age	-0.00813	-0.00142	-0.00956
Workfrc	0.021583	0.00378	0.025363
logpop	0.047349	0.008293	0.055642
logReilly	0.028572	0.005004	0.033577
Dumhort	0.181955	0.031869	0.213824
Dumpast	0.189905	0.033261	0.223166
Dumarab	0.098783	0.017301	0.116084
Dumhok	-0.00371	-0.00065	-0.00436
Dumpera	0.414521	0.072602	0.487123
Dummix	0.308258	0.05399	0.362249
Dumzkle	0.475733	0.083323	0.559057
Dumveen	0.049536	0.008676	0.058212
Dumplei	0.253835	0.044458	0.298293
tourism	-0.03802	-0.00666	-0.04468
sale	0.219972	0.038527	0.258499
care	0.694868	0.121704	0.816571
prcssng	-0.45391	-0.0795	-0.53341

Annex Bibliography

Leeuwen, van, E., and J. Dekkers. 2013. Determinants of off-farm income and its local patterns: a spatial microsimulation of Dutch farmers. *Journal of Rural Studies* 31:55-66.