Bio-based and Applied Economics 8(3): 297-323, 2019
DOI: 10.13128/bae-9329

Full Research Article

Spatial distribution of organic farms and territorial context: An application to an Italian rural region

Andrea Bonfiglio*, Andrea Arzeni

Research Centre for Agricultural Policies and Bioeconomy, CREA – Council for Agricultural Research and Economics, Italy

Abstract. Organic farming is increasingly promoted and supported at several levels for its capability of producing safe food and public goods. It can give an important contribution to attenuating the environmental pressure generated by conventional agriculture. This paper analyses possible determinants of the spatial distribution of organic farms in a rural region of Italy, characterised by several environmental issues. Towards this aim, a quasi-Poisson hierarchical generalised linear model with mixed effects is adopted. Results indicate that there is spatial correlation and that the distribution of organic farming is related to socio-economic, environmental and political factors. In particular, they show that public support could have favoured the spreading of organic farming where there are more problems of erosion but far from the areas where there is intensive agriculture.

Keywords. Organic farming, environmental pressures, rural development policy, hierarchical generalised linear mixed model.

JEL Codes. C25, Q18, Q56.

1. Introduction

Organic farming is defined as an overall system of farm management and food production combining best environmental practices, a high level of biodiversity, the preservation of natural resources, the application of high animal welfare standards and a production method consistent with consumer preferences for products produced using natural substances and processes (Council of European Union, 2007). Organic farming has been acquiring growing importance because of the dual role it plays: it provides quality and safe food in response to an increasing consumer demand and, at the same time, it produces public goods contributing to the protection of the environment, animal welfare and rural development.

In consideration of the important contribution of organic farming, better understanding how this phenomenon develops and which factors affect its spatial distribution can be diriment for policy makers in planning strategies pursuing objectives of sustainable development in rural areas.

*Corresponding author: andrea.bonfiglio@crea.gov.it

ISSN 2280-6180 (print)
ISSN 2280-6172 (online)

In this respect, there is a growing body of literature about the identification of factors that influence spatial patterns of organic farms. Several studies have shown that the adoption of organic farming is spatially clustered (Nyblom *et al.*, 2003; Parker and Munroe, 2007; Lewis *et al.*, 2011; Schmidtner *et al.*, 2012; Bjørkhaug and Blekesaune, 2013; Taus *et al.*, 2013; Wollni and Andersson, 2014). Moreover, a number of determinants explaining the territorial distribution of organic farming have been highlighted. They are, in particular, the neighbouring effects connected with social influence and learning (Nyblom *et al.*, 2003; Lewis *et al.*, 2011), access to information (Frederiksen and Langer, 2004; Genius *et al.*, 2006; Läpple and van Rensburg, 2011), existing agricultural systems (Häring *et al.*, 2004), characteristics of the agricultural landscape (Gabriel *et al.*, 2009; Schmidtner *et al.*, 2012) and proximity to markets and urban areas (Häring *et al.*, 2004; Koesling *et al.*, 2008). Another important factor that can justify different concentrations and development paths of organic farms between areas and regions is represented by public support (Nyblom *et al.*, 2003; Frederiksen and Langer, 2004). In Europe, the most important policy in favour of organic farming is represented by the Common Agricultural Policy (CAP), which, by means of the Rural Development Policy (RDP), provides incentives that are aimed at compensating farmers for additional costs and income foregone deriving from the conversion or the maintenance of organic farming practices (EU Regulation No. 1305/2013).

The main objective of this paper is to analyse the determinants of the spatial distribution of organic farms in a rural region located in Central Italy. Firstly, a Poisson regression model is adopted to analyse the effects of several social, economic, environmental and political factors on the territorial distribution of organic farms. The neighbourhood influence is then assessed to verify the appropriateness of a spatial model. Thus, a quasi-Poisson hierarchical generalized linear mixed (HGLM) model is applied, allowing for under- overdispersion and including a spatially autocorrelated random effect. Poisson regression models are applied by representing the regional territory as a regular grid of uniform cells and considering the number of organic farms localized in each cell as a response variable.

This paper adds in the literature concerning the identification and the assessment of contextual factors explaining the spatial distribution of organic farms. The focus is thus territorial rather than the individual farm. The context can play an important role in terms of increasing returns to adoption due to the economies of scale associated with the concentration of organic farms. These economies include, for instance, formal and informal networks of organic farmers, technical support structures and downstream structuring (Allaire *et al.*, 2015). They are the result of shared needs and collective capabilities derived from past experience by all organic farms. If this experience is transmitted to neighbouring territories, it generates spatial autocorrelation, which goes beyond individual behaviours and specific territorial determinants. The main contribution of this paper is both territorial and methodological. Firstly, it analyses a regional area, i.e. the Marche region, which exhibits interesting peculiarities. This region presents very diverse characteristics from the coast to the Apennine mountains and is therefore representative of the high heterogeneity of the Italian territory. Moreover, it is characterised by significant phenomena of erosion and negative environmental effects generated by intensive farming and related to a high specialization in arable farming (Rusco *et al.*, 2008). In this context, organic farming could therefore be very helpful in reducing the environmental impact generated by agriculture. From a methodological standpoint, rasterized data are used to determine the influence of territorial factors on a discrete variable, represented by the concentration of organic farms. Although rasters are most commonly used to represent continuous data, such as temperature and elevation values, they are often used to represent categorial or discrete data as well (Pingel, 2018). Rasterization, based on a common spatial reference, allows us to use an heterogenous set of statistical data, composed of discrete and continuous variables, and available in different formats and at different resolutions. To this aim, various techniques are adopted to rescale data to the same spatial unit. In comparison with traditional approaches based on the use of administrative borders, gridded data allow a better analysis of causes and effects of socio-economic and environmental phenomena (Eurostat, 2016). Moreover, rasterization of farms, i.e. their aggregation within cells, is one of the possible statistical-disclosure-control techniques to tackle the issue of ensuring anonymity in using micro-data (United Nations, 2007), which is instead potentially present in research works at a farm level and based on the use of point data (i.e. Läpple and Kelley, 2015)

In literature, there are a few studies that analyse the influence of territorial factors, rather than those at a farm level, on the distribution of organic farms through spatial econometric models for count data. For instance, Gabriel *et al.* (2009) use a raster approach based on a 10-km grid to investigate the spatial distribution of organic farms in England and analyse the degree to which environmental and socio-economic factors correlate with their distribution. To assess the effects of covariates, they adopt a hurdle model (Cragg, 1971). They analyse the presence/absence of organic farms using a generalized linear model (GLM) with binomial errors, and the concentration of organic farming using a GLM with Gaussian errors. Spatial information is incorporated by fitting spatial eigenvector mapping into the binomial model and by performing a spatial autoregressive (SAR) model for the Gaussian model. More recently, Allaire *et al.* (2015) analyse the influence of territorial context on the diffusion of organic farming in France, measured by the number of beneficiaries of aid for conversion to organic farming (COF) existing in micro-territories at a NUTS-4 scale. A hurdle model, composed of a spatial probit model and a non-spatial zero-truncated negative binomial regression, is applied to assess both the extent of the contracting of COF aid and its local intensity. Hurdle models, as well as zero-inflated Poisson (ZIP) models (Lambert, 1992) and their variants, are commonly used to handle the overdispersion in the response variable generated by a high number of zeros. These models are based on some theoretical assumptions. They assume that either zero observations have different origins (ZIP models) or zeros and positive counts are generated by different processes (hurdle models). These models are also restricted to overdispersion, therefore excluding the possibility of underdispersion, which is another issue, even if less frequent, that may occur. Moreover, in the cases where spatial effects are not considered, they fail in handling a further source of overdispersion generated by spatial aggregation, which increases the probability of observing zeros. A more flexible model is the quasi-Poisson HGLM model. The latter does not introduce any assumption about the distribution of zeros. Moreover, it can handle both under- and overdispersion in the response variable, generated by excessive zeros and spatial dependence. This kind of model has been recently used by Lee *et al.* (2016) in the ecology field, showing its superiority in spatial prediction in comparison with zero-inflated and hurdle models. To our knowledge, it has not yet been used in agricultural studies, in particular in the field of organic farming. Compared with existing research, a further novelty is therefore the analysis of the influence of territorial factors on the concentration of organic farms using raster data in conjunction with a quasi-Poisson HGLM model for jointly handling dispersion in both directions and spatial dependence.

The reminder of this paper is organized as follows. Section 2 is devoted to illustrating the area under study, the variables and the dataset used, and the spatial econometric model adopted. Sections 3 and 4 present and discuss the main results, also providing some policy recommendations. Final section concludes.

2. Materials and methods

2.1 The area under study

The area analysed in this study is the Marche, an Italian rural region. It is located in the Central-Eastern part of Italy and has an area of 9.7 thousand km^2, equivalent to about 3% of the national territory. Most of the territory is mountainous or hilly. The annual average temperature varies from 5 to 14 °C. In the coastal strip and middle hill, the climate is Mediterranean. It gradually turns into sub-Mediterranean moving inward, while it is similar to the oceanic one in the mountains although Mediterranean influences are still present. The average annual precipitation is 700 to 1,400 mm as we move from the coast towards mountain areas.

Agriculture is particularly important in the region from a territorial standpoint. The share of total surface managed by farms amount to 68%, against a national average of 57%. This reveals the importance of regional agriculture in managing natural resources and, thus, in affecting the overall quality of the environment.

The Marche region is characterized by a marked specialization in arable crops. Based on 2013 data (ISTAT, 2015), regional farms with arable crops are more than 37 thousand units, i.e. 90% of total farms, and cultivate over 361 thousand hectares, equivalent to 81% of regional utilized agricultural area (UAA), which is by far higher than the national share (55%). However, this strong specialization, which is also associated with an intensive use of inputs and extensive application of mechanization, raises some concerns about the high pressure that agriculture can potentially exert on the environment, in consideration of the morphologic characteristics of the region (Bonfiglio *et al.*, 2017). In fact, about 90% of the agricultural area is subject to erosion with annual values of eroded soil ranging between 5 and 20 tons/ha. This phenomenon is due to the natural morphology of the territory and is quite significant in terms of geographical coverage. Intensive farming aggravates this problem and, in addition, produces pollution. The nitrate vulnerable zones cover an area corresponding to 11% of the territory, approximately 21% of total UAA. These zones fall into major regional river basins and involve both areas around river courses and the regional coastal strip (Regione Marche, 2015a).

In this context, organic farming could be one of the possible options to attenuate the environmental pressure generated by agriculture, thanks to the relevant environmental benefits. In particular, organic farming can help to preserve soil (Arnhold *et al.*, 2014; Reeve *et al.*, 2016) and reduce water, soil and air pollution by banning the use of chemical pesticides and synthetic fertilizers (Jouzi *et al.*, 2017).

Besides an increasing demand for organic products, the constant rise of organic farming has been significantly affected by policy support (Sanders *et al.*, 2011). In Italy, as in other European Member States, the most important policy instrument supporting organic farming is the RDP.[1] In the 2007-2013 programming period, measure 214 about agri-environmental payments (Reg. EC No 1698/2005) has been introduced to encourage the introduction and the maintenance of organic production. In the 2014-2020 programming period, it has been replaced by measure 11 (Reg. EC No 1305/2013). One of the main differences between the two frameworks is that, while measure 214 financed several agricultural production methods compatible with the protection and the improvement of the environment[2], measure 11 specifically supports organic farming.

Organic sector is still growing in Italy (SINAB, 2001,2008,2015). From 2007 to 2014, the number of organic operators and the agricultural area managed organically have increased by 21% and 10%, respectively. However, this expansion has slowed down in the last years since the relevant measure has exerted lower appeal. The reasons for this relate to relatively low incentives in favour of organic farming, high disparity between payment levels at a regional level and return to conventional farming, determined by sudden price increases in some commodities and more incentivizing schemes supported by measure 214, such as integrated farming (Zaccarini Bonelli, 2011).

From a regional point of view, in 2014, the Marche region was the eighth Italian region in terms of importance. There were 2,187 organic operators (4% of total organic operators) with an agricultural area of about 57 thousand ha (4% of total area). Relative to 2013 total UAA (ISTAT, 2015), the share of organic area was 13%, slightly higher than the Italian average, amounting to 11%. After experiencing a significant phase of expansion in the early 2000s (the number of operators has grown by over 60% from 2000 to 2007), the regional organic sector has been involved by a continuous process of decline: from 2007 to 2014, organic operators have decreased by 23% and the agricultural area used for producing organic products has diminished by 44%. In addition, from 2007 to 2013, the share of organic area in relation to total UAA (ISTAT, 2008,2015) has decreased by about 8% while the national one has increased, even if slightly, passing from 9% to 11%. Regarding policy application, according to 2015 data about financial implementation (Regione Marche, 2016a), the Managing Authority of the Marche region allocated €108.8 million (of which €49.6 million represented by EAFRD contribution) to measure 214 for the 2007-2013 programming period, equivalent to 22.5% of total RDP expenditure.

2.2 The variables analysed

The count (dependent) variable used is the number of organic farms operating in the Marche region. As regards potential determinants, a number of variables that could affect the distribution of organic farming are investigated. They refer to the following aspects; existing farming system; land use; environmental characteristics; demographic and social characteristics; policy.

---

[1] In addition to the RDP, there exist other policy instruments in favour of organic farming, which are not considered here, such as Article 68 of Regulation EC No 73/2009, contribution to producer organisations of the fruit and vegetable sector based on Regulation EC No 1234/2007 and further national/regional support outside the CAP. For a wider discussion about these policy tools, see for instance Sanders *et al.* (2011).

[2] In the case of the Marche region, five actions were financed by measure 214 of the 2007-2013 RDP: a) integrated production; b) organic farming; c) protection and improvement of soil; d) maintenance of local endangered varieties and breeds; e) better management of permanent pastures (Regione Marche, 2015a).

Farming system is described by the number of total farms (this serves as an exposure; see the methodological section for details), the share of the agricultural area used for arable crops, UAA per farm, labour units per hectare and the share of young farmers. Land use is described by the percentage of urban and natural areas. The environmental characteristics considered are altimetry, soil fertility and erosion. Altimetry is measured as the natural logarithm of meters above the sea level. Soil fertility is modelled using data about the percentage of soil organic matter. Erosion is approximated by the quantity of tons of soil per hectare, which are lost due to surface water flows. Demographic and social factors are represented by the shares of resident population aged 20-39 (young population) and 40-64 (adult population) and by the proportions of population with higher education (high school and university). Finally, policy is represented by the total amount of public subsidies per hectare of agricultural area.

The choice of determinants takes account of different aspects such as potential benefits of organic farming, the context considered, and the variables investigated in studies concerning organic sector. Specifically, it is based on the following considerations. Given its potential contribution to the environment, organic farming is expected to be more concentrated in territories of the Marche region where there are or there might be more critically environmental issues, i.e. where there are higher levels of erosion and lower levels of soil fertility. This would be consistent with previous research showing a higher presence of organic farms where soil is less fertile and the levels of erosion are higher (Lewis *et al.*, 2011; Gabriel *et al.*, 2009; Wollni and Andersson, 2014; Paudel and Thapa, 2004). Moreover, organic farming should be less concentrated in areas where its environmental contribution is lower because of the prevalence of natural elements, i.e. where there is a higher share of natural areas. The altimetric distribution of organic farms may also be important from an environmental point of view. In the medium-high hills of the Marche region, organic farming can prevent from phenomena of landslides and leaching, thus preserving soil integrity, while, in flatter areas, it can reduce the quantity of pollutants produced by agriculture. With reference to the existing farming system, organic farming can contribute to reducing the environmental pressure exerted by intensive agriculture. Therefore, we expect that organic farming is more concentrated where there is a higher soil exploitation, i.e. in areas characterized by higher levels of mechanization, thus a lower use of labour, and a higher specialization in mechanizable crops, such as arable crops. However, the adoption of organic farming can significantly depend on the relevant costs. In this respect, Häring *et al.* (2004) have pointed out that the conversion to organic farming is more convenient for farms using practices that are less intensive in the use of mechanization. In this case, in contrast with our expectations, we would have a higher presence where the contribution of organic farming to the environment is lower. Regarding further characteristics of farms, Läpple and van Rensburg (2011) have shown that the adoption of organic farming is more probable among smaller and younger farmers. This can be justified by the economic opportunities, in terms of relatively higher prices and public subsidies, which organic farming can give to new entrants and, in general, to farms that are too small to compete on the market. The result would be a higher concentration of organic farms in locations where there is a higher percentage of young farmers and small-sized farms. However, regarding farm size, Pietola and Lansink (2001) have shown that farmers who have large land areas are more likely to switch to organic farming since they have more possibilities of applying extensive farming technologies. In line with this result, Koesling *et al.* (2008) have found that the probability that a farmer will produce organically rather than conventionally increases if the farmer has more farmland. As regards socio-demographic characteristics, Wier *et al.* (2008) have shown that urbanization, education and population age play an important role in consumer choices. In particular, medium and long education increases the propensity to purchase organic foods. This propensity is also higher in more adult population and where the levels of urbanization are higher. Looking at the supply side, studies, such as those by Häring *et al.* (2004) and Koesling *et al.* (2008), have shown that organic farms, for which direct marketing is particularly important, tend to localize close to urban areas, because they would have an easier access to consumers. If these results were confirmed, we should find that organic farms are mostly located where the percentage of urban areas as well as the share of population aged 40-64 and with higher education are higher. Finally, policy may be another important factor to be considered since it can incentivise the spreading of organic farms by compensating higher costs or lower income resulting from organic management (Sanders *et al.*, 2011).

2.3 The dataset used

Data come from several sources and are available at a different geographical detail, i.e. points, irregular polygons and grids with different resolutions. For this reason, they are rescaled to the same territorial unit, represented by regular grids[3] composed of uniform cells.[4]

Specifically, data about organic farms refer to 2014 and come from the national register of organic operators. Organic farms are both those who are already organic and those who are converting to organic farming. For every farm, there is information about the relevant headquarter address, which is used to identify the exact geographical position and localize farms within cells. Total number of organic farms per cell is thus obtained as a sum of the organic farms localized in each cell.[5] Overall, in 2014, organic farms operating in the Marche region and enrolled in the national register amounted to 2,160 units.[6]

2010 agricultural census is used to retrieve information about total farms and the relevant data about UAA and the age of farmers. Coordinates of the firm site or, alternatively, the relevant address are used to localize farms within each cell. This allows us to derive total number of farms and of young farmers, who are less than 40 years old, existing in each cell.

---

[3] The methodology used for dividing the territory into grid cells is based on the INSPIRE Equal Area Grid system (INSPIRE, 2014). Spatial representation is defined by a specific system of geographical coordinates (ETRS89-LAEA) that can be used as a common reference for different sources and studies. According to this system, cell sides should have a length included between one metre and 100 km with multiples of 10.

[4] It should be cleared that not all cells are of regular size. In fact, the shape of the cells located on administrative borders of the region is adjusted in such a way to correctly represent the regional territory.

[5] Data on the agricultural area used by organic farms are also available. However, they are the sum of hectares that can be partly positioned in cells different from those where organic farms are localized. Of these hectares, the exact localization, which is necessary for georeferencing, is unknown. Moreover, for farms which are converting into organic farming, information about the area in conversion towards organic is not available. Accordingly, data about the agricultural area of organic farms cannot be used.

[6] In allocating farms among cells based on their headquarter address, it can occur that some organic farms fall in cells where there are not agricultural producers according to the agricultural census or there is not agricultural land according to 2012 Corine Land Cover. For instance, in the case of a grid of 3-km size, there are 7 farms of this kind. This happens when headquarters do not correspond with operational sites. Since removing these farms can produce biased estimates, they have been reallocated in the neighbouring cells where there are agricultural producers and land using the k-nearest neighbours algorithm (package spdep version 1.1.3 in R3.5.3).

UAA per farm is calculated by dividing total UAA of farms located in each cell by the number of farms. The share of young farmers is obtained by dividing the number of young farmers by total farms falling into each cell.

2012-2014 data from Italian Farm Accountancy Data Network (FADN) are used to estimate labour intensity. Specifically, regional average ratios of labour units to UAA, differentiated by land use and five altimetric zones, are firstly calculated using FADN. These coefficients are multiplied by the shares of agricultural area present in each cell, taking account of different uses according to 2012 Corine Land Cover and the altimetric zone of the cell. The sum of labour units distinguished by land use represents the total labour units employed in each cell. Dividing this sum by total agricultural area, labour units per hectare are obtained.

Information about land use comes from 2012 Corine Land Cover and is available at a 100-metre resolution, i.e., one-hectare area. The database classifies land in five main classes, of which classes "Artificial surfaces" and "Agricultural areas" are used to identify urban and agricultural areas, respectively. Sub-classes "Shrub and/or herbaceous vegetation associations" and "Open spaces with little or no vegetation" within class "Forest and seminatural areas" are instead employed to define natural areas. Urban, agricultural and natural areas are obtained by summing the relevant hectares recorded in the dataset and falling into each cell. The shares of urban and natural areas are calculated as ratios of the corresponding hectares to total area, obtained as a sum of hectares belonging to all classes. Sub-class "Arable land" within "Agricultural areas" is used to derive arable land and calculate the relevant share, obtained by dividing arable land by the total area attributed to each cell. All sub-classes within "Agricultural areas", including "Arable land", "Permanent crops", "Pastures" and "Heterogeneous agricultural areas", are used to estimate labour intensity (see above).

Data concerning soil erosion come from the 2015 European dataset "Soil Loss by Water Erosion in Europe", which offers detailed information on soil erosion by water in 2010 for the European Union at a resolution of 100 metres (Panagos *et al.*, 2015). Data on erosion are available as tonnes per hectare.

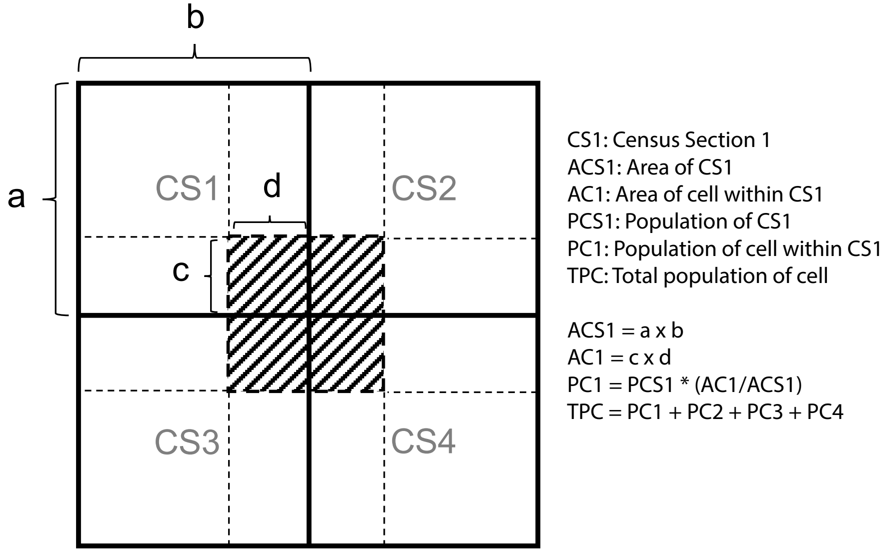
Data related to soil organic matter are from the EFSA Spatial Data Version 1.1 and are available at a resolution of 1,000 metres (Hiederer, 2012). The dataset is composed of several layers, of which that relevant to organic matter concentrations expressed as percentages is used. Shares are calculated from the map of topsoil organic carbon by applying a factor of 1.72. This factor assumes an average organic carbon content of organic matter of 58%.

Information about altimetry, expressed in meters, is retrieved from Shuttle Radar Topography Mission dataset, managed by the U.S. Geological Survey agency. In particular, we use digital elevation data published in 2014 with a resolution of 3 arc-second for global coverage, corresponding to a spatial resolution of about 90 meters.

Levels of erosion, organic matter concentrations and altimetry of each cell are estimated by summing the respective levels relevant to the areas falling into each cell and dividing these sums by the number of areas that belong to the cell, so obtaining an average level of erosion, an average percentage of organic matter concentration and average altimetry for every cell.

Data about resident population, distinguished by age and level of education, come from 2011 population census and is available by census section, which represents the minimum territorial unit of a given municipality on which the census survey is based. Spatially, it is represented by an irregular polygon. The sum of all census sections gives the entire regional territory. The polygons corresponding to census sections are firstly cut using regular grids

Figure 1. Subdivision of census sections into regular cells and estimation of the population of a given cell (square with diagonal stripes) belonging to several census sections.



Note: for the sake of simplicity, census sections are supposed to be represented by regular polygons

in order to quantify the area falling into cells (Figure 1). Then, the population of a census section, which is present in a given cell (PC1), is estimated by multiplying total population (PCS1) by the share of territorial surface falling into the cell (AC1/ACS1). The total population of the cell (TPC) is thus obtained as a sum of shares of population of all census sections, whose surfaces fall into the cell (PC1, PC2, etc.). Percentages of population aged 20-39 and 40-64 as well as that of population with higher education are calculated by dividing the relevant quantities by the total population assigned to each cell.

Finally, public subsidies are collected from the dataset of national agency disbursing agricultural funds. We use information about the payments made in the period 2008-2014 and relevant to measure 214 of the 2007-2014 RDP. In this way, we assess the influence of the 2007-2014 RDP on the concentration of organic farms existing in 2014. 2008 is the first year of effective application of the RDP (i.e. the first year when payments are made), while 2014 represents the final year of programming and corresponds with the reference year of organic farms that are present in the national register of organic operators. Data are not distinguished by sub-measure. Therefore, it is not possible to identify the amounts that are specific to organic farming. However, this should not affect results significantly, since, in the Marche region, almost the totality of the payments relevant to measure 214 are addressed to support organic farming.[7] For each payment, information about the identification code of the beneficiary is available. This code is matched with that resulting from the national register of organic operators in order to associate payments with the organic farms recorded in

[7] In the 2007-2013 programming period, 96% of total payments have been used to finance the specific action related to organic farming (Regione Marche, 2016b).

Table 1. Descriptive statistics about the data used based on a regular grid of cells with a 3-km grid size and presence of farmers (total number of observations = 1,032).

	Mean	Standard deviation	Min	Max
Organic farms (number)	2.1	3.2	0.0	46.0
Farming system				
Total farms (number)	43.5	35.5	1.0	259.0
% of arable land	57.9	32.3	0.0	100.0
UUA per farm (ha)	16.4	25.4	0.3	400.0
Labour units per ha	0.1	0.1	0.0	0.3
% of young farmers (< 40 years old)	3.0	2.8	0.0	22.0
Land use				
% of urban areas	6.2	12.6	0.0	100.0
% of natural areas	7.6	11.3	0.0	84.6
Environment				
Altimetry (meters above the sea level)	350.9	279.1	2.0	1,687.0
% of soil organic matter	3.0	1.8	0.0	10.4
Erosion (tons / ha)	12.3	5.5	0.0	26.7
Demographic and social factors				
% of 20-39 years-old population	11.3	3.2	0.0	35.3
% of 40-64 years-old population	34.6	5.9	0.0	100.0
% of population with higher education level	33.9	8.9	0.0	100.0
Policy				
2008-2014 Policy payments per ha (€)	54.3	415.6	0.0	12,513.5

Source: Authors' elaborations

the national register. In this way, the sum of all agri-environmental payments that are made to organic farms existing in each cell can be derived. Public subsidies per hectare, for every cell, are calculated by dividing the sum of agri-environmental payments by total agricultural area, obtained by using data from 2012 Corine Land Cover.

Table 1 provides some descriptive statistics about the data used. For the sake of convenience, only data referring to the cells with a 3-km size where there are farmers are shown (see methodological and results sections for an explanation).

2.4 Methods

2.4.1 The Poisson regression model

A GLM for count data, specifically a Poisson regression model, is firstly adopted. This choice depends on the objectives of the analysis and the characteristics of the dataset used. The main aim of this study is to analyse how specific characteristics of the territory affect the probability of observing organic farming in a given space. As already specified, these characteristics can be measured using data that are available at different spatial levels. Therefore, these data are converted into a common spatial reference, using regular grids composed of uniform cells. Moreover, in consideration of data availability, the distribution of organic farming can be analysed in terms of farms, represented by points that can be georeferenced within a grid, so obtaining the number of farms, i.e. a count variable, operating in each cell. In analysing count data, ordinary least squares regression cannot be adopted because count data are non-negative and discrete, tend to be highly skewed and non-normally distributed, and commonly follow a Poisson distribution (Ma *et al.*, 2012). For all these reasons, a regression model based on a Poisson distribution may be considered as appropriate.

The Poisson regression model has two fundamental components: the response distribution is not necessarily Gaussian distribution and a monotonic link function is used to transform the mean of response variables into a linear form. The probability density function of a Poisson random variable Z is given as:

$$P(Z = z) = \frac{e^{-\lambda}\lambda^{z}}{z!} \tag{1}$$

where parameter λ is the mean and the variance of random variable Z, i.e., $E(Z) = \lambda$ and $\mathrm{Var}(Z) = \lambda$. Thus, in the absence of other information, one should expect to see λ events, represented, in our case, by a given number of organic farms, in any spatial unit. Assuming that event rate λ is not constant but depends on a number of variables, which are supposed to affect the probability of observing a given number of organic farms in a given spatial unit, the Poisson regression model takes the following form:

$$\log(E(Z)) = \log(\lambda) = \mathbf{x'\beta} \tag{2}$$

where $\mathbf{x'}$ is a row vector of explanatory variables and $\boldsymbol{\beta}$ is a column vector of unknown regression coefficients. The log function is the link between the mean of the Poisson random variable and linear predictors. It ensures that the mean remains positive for all linear predictors and parameters. The Poisson regression assumes that observations are independent. However, this assumption could be invalid since the number of organic farmers, which operate in a given area, can be spatially dependent. In other words, it might also depend on the farmers located in the neighbouring space. If there is no evidence of significant spatial autocorrelation in model residuals, non-spatial methods may be appropriate. However, if statistical tests indicate significant spatial autocorrelation, methods that also consider the spatial component should be adopted (Ma *et al.*, 2012). In order to verify the existence of spatial autocorrelation, the Moran's *I* test (Moran, 1950) is carried out on the residuals of the Poisson regression model. Values of Moran's *I* range from -1 (indicating perfect dispersion) to +1 (perfect autocorrelation). The regression model is applied using different sizes of cells in order to confirm spatial autocorrelation in correspondence with different grids.[8] The choice

[8] This kind of autocorrelation could also exist between farms operating within the same cells and this would contrast with the assumption of the Poisson regression model, according to which observations should be independent. To ensure that this assumption is not violated, it would be necessary to carry out an analysis based on the use of points rather than areas. However, this would not be consistent with the objectives of this analysis and the dataset available. In fact, the aim is to assess the influence of territorial characteristics on the concentration

of distances is often an empirical issue since exact information on the size of the neighbourhood does not exist (Roe *et al.*, 2002). Following Lapple and Kelley (2015), we assume that beyond a certain distance, a spatial effect, if any, does no longer affect the adoption of organic farming. In order to allow for several neighbours per farm, 1 km is chosen as the minimum distance cut-off and Poisson models with 1, 2, 3, 4 and 5 km distance cut-off are applied. This requires the creation of grids composed of a decreasing number of cells, i.e. 9,721 (1 km), 2,525 (2 km), 1,160 (3 km), 675 (4 km), 441 (5 km).[9] The regional territory is also represented by non-agricultural areas, where zero observations are due to the absence of agriculture. This implies that the spatial autocorrelation, identified by the Moran's *I* test, could be the result of a clustering of agricultural and non-agricultural areas rather than a clustering of organic and non-organic farming. To avoid this, the Poisson regression model is applied only to those cells where there are agricultural producers, amounting to 5,785 (1 km), 2,087 (2 km), 1,032 (3 km), 622 (4 km), and 409 (5 km).

2.4.2 The quasi-Poisson hierarchical generalized linear mixed model

After confirming spatial autocorrelation, a HGLM model is then applied to analyse the influence of spatial dependence in addition to other possible factors on the distribution of organic farms. Different from linear models, HGLM models allow for inclusion, besides the usual fixed linear covariates, of an independent random location effect accounting for heterogeneity or, as in this study, a spatially autocorrelated random effect.

Specifically, a quasi-Poisson HGLM model with CAR-type specification[10] of spatial covariance (hereinafter, quasi-Poisson CAR-HGLM model) is adopted.[11] Following Lee *et al.* (2016), this model takes the following form:

of organic farms. Data about these characteristics are not all at level of single farms (thus point data) but are, mostly, at level of space. Moreover, the reference unit of spatial analysis is given by the group of farms existing in a given cell rather than the single farm, i.e., the objective is to analyse spatial dependence between groups of farms, which can be composed of one or more units. As in Allaire *et al.* (2015), we assume that the farms located in a given space are a homogenous group, sharing, because of their location, the same territorial characteristics that influence farmers' strategies.

[9] The threshold of 2 km corresponds with the squared root of the average territorial density of regional organic farms, which is specifically one farm every 4.5 km^2.

[10] There are two auto-Poisson models that are commonly used: the SAR model (Whittle, 1964) and the conditional autoregressive model (CAR) (Besag, 1975). CAR models are very popular in spatial analysis of count data (Lee *et al.*, 2016). They are unsuitable when the spatial weight matrix is asymmetric (Dormann *et al.*, 2007) but they are appropriate in the opposite case, as happens when proximity or distance between areas is modelled. The SAR specification is a special type of CAR models, at least in a continuous-response context, and generates a less natural spatial structure (Cressie, 1995). The SAR approach is harder to apply for more complex and limited-response situations, especially when large datasets are used (Wang *et al.*, 2012,2014) and yields parameter estimates that are similar to those estimated by the CAR model (Kim and Lim, 2010). A formal and more exhaustive presentation of SAR and CAR models can be found in Wall (2004) and Glaser (2017).

[11] It should be remarked that the main interest, here, is not methodological. In other words, the aim is not to find and adopt the best model among a battery of possible alternatives but to evaluate the influence of some factors on the observed distribution of organic farms. This is conducted using one of the possible models available in literature, which can fit to the context and to the data analysed. Comparison of results associated with different models and, thus, issues related to model selection can be an interesting and future research direction.

$$
\begin{gathered}
f(z_i|v_i) = \phi^{-\frac{1}{2}} \exp\left(-\frac{\lambda_i}{\phi}\right) \frac{\exp\left[\left(\frac{1}{\phi} - 1\right) z_i\right] z_i^{z_i}}{z_i!} \left(\frac{\lambda_i}{z_i}\right)^{\frac{z_i}{\phi}} \approx \phi^{-1} \exp\left(-\frac{\lambda_i}{\phi}\right) \frac{\left(\frac{\lambda_i}{\phi}\right)^{\frac{z_i}{\phi}}}{\left(\frac{z_i}{\phi}\right)!} \\
E(z_i|v_i) = \lambda_i \\
\mathrm{Var}(z_i|v_i) = \phi\lambda_i \\
\log(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta} + v_i \\
\boldsymbol{v} \sim N(\mathbf{0}, \Sigma = \tau(\mathbf{I} - \rho\mathbf{W})^{-1})
\end{gathered} \tag{3}
$$

where $z_i|v_i$ is the conditional distribution of the response (count) variable given the location specific random effect v_i; ϕ is a dispersion parameter; λ_i is a random intensity for location i, which equals the conditional mean, i.e. the expected value $E(z_i|v_i)$; $\phi\lambda_i$ is the variance Var $(z_i|v_i)$, where ϕ =1 gives the Poisson distribution, while $\phi < 1$ and $\phi > 1$ allow for underdispersion and overdispersion, respectively; **v** is a vector of random effects that are supposed to be normally distributed; **Σ** is the covariance matrix of the n-dimensional normal density with a CAR-type specification; **I** is an identity matrix; τ and ρ are coefficients to be estimated. In particular, ρ is known as spatial dependence parameter (Hodges, 2014), with $\rho = 0$ corresponding to independence and $\rho = 1$ corresponding to strong spatial autocorrelation. **W** is a spatial weight matrix, which defines the relationship among different locations. In other words, it defines the spatial neighbourhood for every location. There are several choices of spatial matrices, depending on the neighbouring criterion (Anselin, 2002). In this study, we use the queen contiguity, according to which two cells are neighbours if they share a common side or a vertex. Moreover, we opt for a binary approach, i.e. diagonal elements are all 0 while off-diagonal elements (i,j) are 1 if locations i and j are neighbours.[12] The grid size used is that which exhibits the highest spatial autocorrelation, since it corresponds with that for which the spatial model can be more appropriate.

The choice of this model is also conditioned on the characteristics of the sample used. In our data, there is a significant share of zero counts.[13] This situation is not infrequent in that spatial counts are often characterised by a high number of zeros (Agarwal *et al.*, 2002; Dénes *et al.*, 2015; Zuur *et al.*, 2012). The presence of more zeros than expected is a source of overdispersion, meaning that the variance is higher than the mean. Under such circumstances, a standard Poisson regression model would be inappropriate. In the literature, for modelling counts with excessive zeros, ZIP models (Lambert, 1992), hurdle models (Cragg, 1971) and their modifications have been proposed (Zuur *et al.*, 2012). These models are based on some theoretical assumptions. A ZIP model assumes that zero observations have two different origins: "sampling" and "structural". More specifically, the population is considered to consist of two types of individuals. The first type involves counts of event in a Poisson or Poisson-like process, which might also contain zeros ("sampling zeros"). The second type always gives

[12] Clayton and Berardinelli (1996) point out that a binary specification of the spatial matrix is not internally consistent in the case where the number of neighbours varies, which occurs with most irregular lattices. In this case, matrix standardization is necessary. Since the grid cells we use are mostly regular lattices, a simple specification is kept.

[13] For instance, using a grid of 3-km size, zero counts amount to about 30%.

a zero count ("structural zeros"). In contrast, a hurdle model assumes that all zero data are from one "structural" source, while the positive (i.e., non-zero) data have "sampling" origin, following either truncated Poisson or truncated negative- binomial distribution.[14] Moreover, these models apply only when there is overdispersion in the response variable. However, there are studies finding that many zeros may be associated with underdispersion (i.e. the variance is lower than the mean) for which ZIP and hurdle models would not be appropriate (i.e., Oh *et al.*, 2006; Tin, 2008). A further source of overdispersion is spatial aggregation, which leads to a higher probability of zero counts (Gabriel *et al.*, 2009). In other terms, due to spatial correlation, co-occurrence of zero counts may generate higher shares of zeros in some samples than expected under the assumption of independent observations (Lee *et al.*, 2016). In these cases, the inclusion of spatially autocorrelated random effects in regression models represents an effective way to handle the problem of overdispersion associated with zero inflation. Neglecting the issues of over- and underdispersion in analysing count data can cause several estimation problems, such as poor fit of the model, different estimates of regression parameters, and wrong inferences concerning the model parameters (Ridout and Besbeas, 2004; Tin, 2008; Ver Hoef and Boveng, 2007).

Compared with ZIP and hurdle models, the quasi-Poisson HGLM model is a more flexible solution. It does not make any specific assumption about the process that generates zeros. Moreover, it can handle both under- and overdispersion, by allowing situations where the variance differs from the mean and including random effects to capture spatial autocorrelation.[15] A quasi-Poisson regression model gives a correction term for testing the parameter estimates under the Poisson model and produces an appropriate inference if overdispersion is modest (Cox, 1983). However, in conjunction with a HGLM model, not only it improves inference but it may also produce better fits in comparison with ZIP and hurdle models, in presence of zero observations (Lee *et al.*, 2016).

Models (2) and (3) allow us to estimate the expected number of organic farms within a given space, depending on some factors. However, the probability of observing organic farms is also conditioned on the presence of farms. To take account of this aspect, a simple remedy consists in adding an exposure (or offset), in our case the logarithm of total farms, to linear predictors (Green, 2012, pp. 847-848). This is equivalent to estimate a model where the response variable is a rate rather than a count variable. This approach assumes that there is exact proportionality between the number of organic farms and total farms, since the parameter associated with the exposure is constrained to one. It is like stating that if total farms increase by 1%, also organic farms increase by 1%. However, this could be untrue since an

[14] In our study, sampling zeros correspond with farms that are organic but do not appear in the national register of organic operators. This can happen for errors, possible delays in updating the register or because farms, which are organic *de facto*, still do not undertake the procedure of certification or, also, renounce the organic certification for avoiding the relevant costs and bureaucratic obstacles. Conversely, structural zeros derive from the presence of only conventional farms. It is evident that both assumptions, i.e. there exist both sampling and structural zeros or there are only structural zeros, might be true. Therefore, the choice between ZIP and hurdle models is not neutral because, not only the initial assumption could be untrue, but the relevant estimates and interpretations could be very different (Hu *et al.*, 2011).

[15] An alternative to quasi-Poisson models, allowing for both overdispersion and underdispersion, is the so-called Conway–Maxwell–Poisson (COM-Poisson or CMP) distribution, which is a generalization of the Poisson distribution (Conway and Maxwell, 1962). One of the main differences between quasi- and COM-Poisson is that the latter requires estimation of two rather than one parameter, making the procedure of estimation more complex.

increase in total farms can be due to an increase in conventional farms. To relax this assumption, which could produce biased estimates, the parameter of the exposure is let to vary, by treating the logarithm of total farms as a further explanatory variable to be included in **x'**.

A model based on a Poisson distribution does not allow the derivation of a natural counterpart to the *R*-squared of a linear regression model, because the conditional mean function is nonlinear, and the regression is heteroscedastic. Therefore, to measure goodness of fit, several alternatives have been suggested (Green, 2012, p. 844). Here, in order to compare non-spatial and spatial Poisson-based models, we use a *R*-squared measure based on deviance residuals, which satisfies all the criteria requested (Cameron and Windmeijer, 1996).[16] Let n be the number of events, this measure takes the following form:

$$R^2_{DEV} = \frac{\sum_{i=1}^{n}\left[\lambda_i \log\left(\frac{\hat{\lambda}_i}{\bar{\lambda}}\right) - \left(\hat{\lambda}_i - \bar{\lambda}\right)\right]}{\sum_{i=1}^{n}\left[\lambda_i \log\left(\frac{\lambda_i}{\bar{\lambda}}\right)\right]} \tag{4}$$

3. Results

Table 2 shows the Moran's *I* statistic calculated on residuals of predictions in Poisson regression models in correspondence with different grid sizes. As can be noted, the value of the statistic is always positive, indicating that the number of organic farms in a given cell increases (decreases) as the number of organic farms in the neighbouring cells increases (decreases). The relevant trend appears to be concave: it increases until 3 km, where it reaches the highest value, and then decreases. As expected, for lower values of grid size, spatial autocorrelation tends to vanish. This is because the distance between farms increases and the number of farms in each cell decreases. On the contrary, for higher values, the number of cells decreases and the concentration of farms within single cells increases with the consequence that spatial autocorrelation becomes weaker.

Results therefore confirm the presence of spatial autocorrelation, which is stronger in correspondence with a grid of 3-km size. The spatial analysis is therefore conducted using this grid size.

Figure 2 shows the distribution of organic farms over the regional territory represented by uniform cells of 3-km size. From the figure, it turns out that organic farming is a phenomenon that spreads over the entire territory. Nevertheless, it mainly localizes in the hinterland, i.e. in the medium-high hills. It is less widespread in the Eastern part of the region,

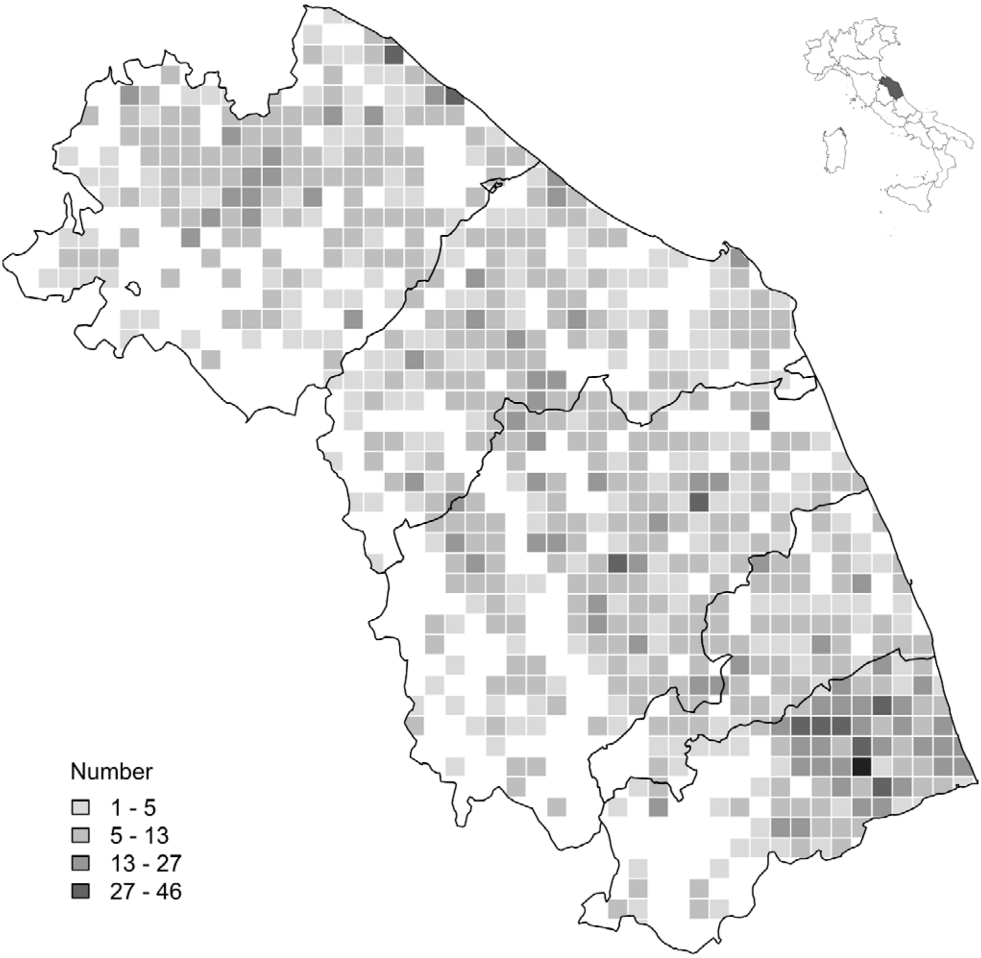
Table 2. Moran's *I* statistics for residuals of predictions in Poisson regression models for different grid sizes, Marche (Italy), 2014.

Size	Moran's *I* statistic	Variance	P-value
1 km	0.0659	0.0001	< 0.001
2 km	0.0896	0.0001	< 0.001
3 Km	0.1561	0.0003	< 0.001
4 Km	0.0954	0.0004	< 0.001
5 Km	0.0563	0.0004	< 0.001

Source: Authors' elaborations.

[16] According to Cameron and Windmeijer (1996), a *R*-squared measure should satisfy the following criteria: 1) it is included between zero and one; 2) it does not decrease as regressors are added; 3) it coincides with *R*-squared based on explained variation; 4) there is correspondence between *R*-squared and significance test on all slope parameters and between changes in *R*-squared as regressors are added and significance tests; 5) it has an interpretation in terms of information content of the data.

Figure 2. Spatial distribution of organic farms within a 3-km-size grid, Marche (Italy), 2014.



Note: administrative borders at NUTS-3 level are shown.
Source: Authors' elaborations.

characterised by flat areas. Moreover, there is a zone in the Southern part of the region and located in the Ascoli Piceno district where organic farms appear to be more concentrated. Spatial distribution of all the predictors analysed are reported in Figure A1 in the Appendix.

Table 3 shows the results relevant to application of both the non-spatial Poisson model and the quasi-Poisson CAR-HGLM model.[17] The *R*-squared based on deviance residuals are

[17] The Poisson regression model and the quasi-Poisson CAR-HGLM model are estimated by using the packages stats version 3.5.3 and hglm version 2.2.1 in R3.5.3, respectively. To solve the quasi-Poisson CAR-HGLM model, the EQL1 method available in the package is used. This method has been conceived to improve estimation for Poisson models with a large number of levels in the random effects (Lee and Lee, 2012).

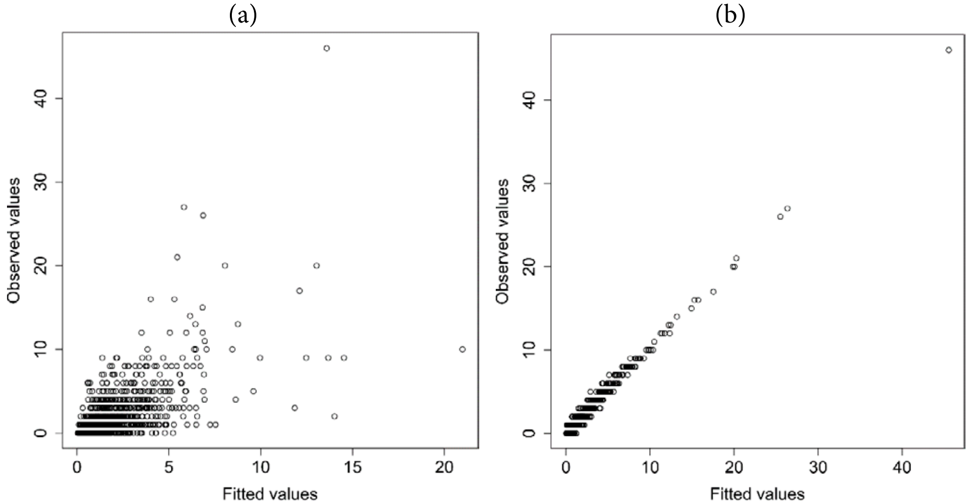
Table 3. Estimated parameters for the Poisson regression model and the quasi-Poisson CAR-HGLM model based on a 3-km-size grid, Marche (Italy).

	Poisson regression model		quasi-Poisson CAR-HGLM model	
	Coefficient	Standard Error	Coefficient	Standard Error
Intercept	-4.053**	0.547	-11.670**	1.199
Number of total farms (log)	0.584**	0.037	0.383**	0.067
% of arable land	-0.678**	0.086	-0.629**	0.214
UUA per farm	0.007**	0.001	0.006**	0.002
Labour units per ha	-2.493**	0.821	6.938**	1.978
% of young farmers (< 40 years old)	0.920 *	0.374	0.666	0.563
% of urban areas	2.734**	0.256	5.434**	0.597
% of natural areas	0.700 *	0.317	-0.491	0.611
Altimetry – meters above the sea level (log)	0.399**	0.073	0.767**	0.169
% of organic matter	-28.210**	3.684	19.360**	7.745
Erosion (tons / ha)	0.039**	0.006	0.039 *	0.013
% of 20-39 years-old population	0.051	1.185	1.941	1.905
% of 40-64 years-old population	1.488 *	0.743	0.749	1.227
% of population with higher education level	1.437**	0.360	3.193**	0.798
Policy payments (€) per ha***	0.190**	0.000	0.295**	0.000
ϕ (mean model)	1.0		0.705	
ϕ (random effects model)	-		3.262	
τ	-		0.894	
ρ	-		0.127	
Moran's I statistic	0.156**		-0.050	
AIC	3842.9		-	
R^2_{DEV}	0.415		0.904	

* p-value < 0.05; ** p-value < 0.01; *** coefficients are multiplied by 1,000 for improving reading.
Source: Authors' elaborations.

0.42 and 0.90, respectively. These values indicate a significantly higher capability of prediction of the spatial model. The different performances can also be observed in Figure 3, which shows a plot of the observed responses against the fitted values. It can be noted that the spatial model fits the observed data much better than the Poisson regression model. Looking at the parameters estimated, several differences in terms of extent, direction and significance between the two models emerge. One evident difference lies in the lack of significance related to the percentage of young farmers, the share of natural areas and the percentage of population aged 40-64 in the spatial model. These coefficients are instead significant and markedly higher (in absolute terms) in the non-spatial model. Further deviations are the extent and the direction of significant coefficients in the spatial model. The relevant coefficients are clearly

Figure 3. Plots of observed and fitted values for (a) Poisson regression model and (b) quasi-Poisson CAR-HGLM model based on a 3-km-size grid, Marche (Italy).



Source: Authors' elaborations.

higher (in absolute terms) in the cases of labour units per hectare, proportion of urban areas, altimetry, share of population with higher education and public subsidies, while they are lower especially with reference to the number of total farms and the percentage of organic matter. Labour units per hectare and the percentage of organic matter have also opposed sign. These results stress the importance of using spatial models if spatial autocorrelation is detected.

Focusing on the spatial model, coefficient ρ indicates the presence of some spatial dependence consistently with the results related to residuals of predictions in the non-spatial regression model. The Moran's I statistics is around zero and the hypothesis that its value is significantly different from zero can be rejected, indicating that there is no more spatial autocorrelation in residuals.

Dispersion parameter ϕ regarding the mean model is lower than one, indicating the presence of underdispersion, while the one related to random effects, which captures spatial autocorrelation, is higher than one, suggesting, in this case, overdispersion. The joint presence of under- and overdispersion legitimates the use of models that can take into account both phenomena.

Analysing the estimated parameters, the results related to farming system firstly show that organic farms are present where there are also other farms. This is an obvious result. However, it is interesting to note that there is not a direct relationship. The relevant coefficient indicates that if total farms increase by 1%, organic farms increase by 0.4%. This means that the distribution of organic farms and that of conventional farms do not follow the same pattern. The coefficient related to the share of arable land shows a negative and significant relationship between productive specialization in arable crops and numerosity of organic farms. The parameters associated with the average farm size and labour intensity are also significant but with opposite signs.

Looking at land use, the number of organic farms is positively and significantly related to the percentage of urban area. In relation to environmental characteristics, positive and significant coefficients are also found in the cases of altimetry, share of organic matter and levels of erosion.

With regard to socio-demographic factors, there is a positive and significant relationship between the share of population with higher education and the presence of organic farms. Finally, the coefficient associated with the incidence of public subsidies is positive and significant. Its value measures the logarithmic difference of expected organic farms due to an increase of one unit in public subsidies per hectare, keeping the other predictor variables constant. It suggests that, to increase organic farms of one unit, it would be necessary to increase total public payments of about 3,400 €/ha, corresponding to an annual payment of 485 €/ha. This value can be interpreted as an average measure of the willingness of regional farmers to convert towards organic farming and could be used as a reference parameter to define premiums in favour of organic farming.

4. Discussion

Results show that the distribution of organic farming is spatially clustered, confirming previous studies. From mapping, it turns out that organic farming tends to concentrate particularly in the hinterland. The concentration of organic farms in the Southern part of the region coincides with a specific cluster of organic operators, the so-called "Bio-distretto Piceno"[18], born in 2014 and promoted by AIAB, an Italian association of organic farms. The econometric model confirms that there are spatial spillovers across territorial units. In other terms, the presence of organic farms in one territorial unit both affects and is affected by the presence of organic farms in the neighbouring units.

Results also indicate that the distribution of organic farming, in addition to neighbouring effects, relates to a number of territorial and political factors. In particular, in line with previous studies (Häring *et al.*, 2004; Koesling *et al.*, 2008), the concentration of organic farming is stronger where there are higher levels of urbanization. This can depend on the adoption of marketing strategies requiring a closer and stricter relationship with consumers, mainly localized in urban areas. Education is a further socio-demographic factor that relates to the distribution of organic farming. Organic farms are more numerous where there is a higher share of more educated people. The reason for this could be a greater propensity of people with higher levels of education to purchase organic products (Zepeda and Li, 2007; Wier *et al.*, 2008).

With reference to farming system, results also show that organic farms are generally located in territories that are not specialized in arable crops and where there is a more intensive use of labour. Moreover, in the areas where there is a higher concentration of organic farms, existing farms have generally a higher average size, confirming that a larger size offers more chances to apply extensive technologies that are typical of organic farming (Pietola and Lan-

[18] A "Bio-distretto" is defined as a geographical, functional and non-administrative area where a partnership between farmers, citizens, tourist operators, associations and public administration is established for ensuring sustainable management of resources. This synergy occurs on the basis of principles and activities of organic production and consumption (short supply chains, organized groups of consumers and producers, high-quality restaurants, public canteens) (Basile, 2014).

sink, 2001). As regards environmental characteristics, there is a higher diffusion of organic farming in less flat areas. Organic farms are also more present where there are higher levels of erosion but also a higher content of organic matter in soils, in contrast with previous studies (Lewis *et al.*, 2011; Gabriel *et al.*, 2009; Wollni and Andersson, 2014). These findings can be read together. In fact, in medium-high hills, characterized by stepper slopes, problems of soil degradation due to surface water flows are graver. However, these areas are also less involved by intensive agriculture and are therefore richer in terms of organic matter. Therefore, we can conclude that organic farms localize far from more competitive agriculture, characterised by a high specialization in arable crops and a more intensive use of mechanisation and chemicals. A reason for this could be lower costs of conversions to organic farming for farmers adopting agricultural methods that are already based on low levels of chemicals and mechanization (Häring *et al.*, 2004). Another explanation can be the level of public payments per hectare, which can be too low for more competitive farms to render organic farming profitable, if compared with their relatively higher economic returns (Pietola and Lansink, 2001). Therefore, organic farming seems to represent an economic alternative for that part of farms located in areas that are morphologically less suited to intensification.

From a policy point of view, results show that there are more organic farms where payments per hectare of total agricultural area are higher. This contributes to confirming the importance of policy support in promoting organic farming. In the light of the other results, it appears that policy support mainly provides incentives in more remote areas, subject to erosion and more fertile. This can be positive for maintaining agriculture where the risk of abandonment is higher, also ensuring environmental protection from phenomena of erosion. However, there might be some inconsistency between the objectives of favouring organic farming and those of increasing environmental sustainability. If it is true that organic farming can help to reduce the environmental pressure generated by farming, it could be strategic to favour diffusion of organic farming also in areas where this pressure is higher, i.e. areas specialized in arable farming where there is a higher level of soil exploitation, and a higher use of fertilizers, chemicals and mechanization. Support to organic farming, especially if it represents most agri-environmental payments, as is in the region under study, should also be targeted in relation to environmental characteristics (European Court of Auditors, 2011). This is because differentiating support spatially on the basis of given environmental issues can increase cost-effectiveness of agri-environmental measures (Uthes *et al.*, 2010; van der Horst, 2007). Analysing the system of incentives in favour of organic farms, it turns out that, during the 2007-2014 programming period, the maximum annual payment per hectare was on average 372 €/ha for introducing organic farming and 320 €/ha for maintenance (Regione Marche, 2015b). These values are lower than the average threshold estimated by the model to favour the appearance of an additional organic farms and are about 40-50% lower than the average threshold established by the Regulation (EC) No 1698/2005 (650 €/ha). The system of compensation differentiated between crops and between areas, ensuring higher premiums for farms located in areas other than the mountains, but only with reference to maintenance. In other words, a single premium for introducing organic farms, distinguished by type of crop, was assigned to farms independently of their localization. As for maintenance, the difference between the premiums granted to the farms located in areas other the mountains and those received by mountain farms was 40€/ha, i.e. just 13% higher. In addition, the regional system awarded with similar scores applications from farmers located in nitrate vulnerable zones and in Natura 2000 sites, as well as livestock farmers raising cattle with organic methods. Therefore, several critical elements emerge, i.e. a relatively low level of premiums, the lack of territorial differentiation with regard to the premium for introducing organic farming, the small difference in terms of premiums for maintenance between areas, and the scoring system, which equates farmers operating in environmentally critical areas, those located in protected areas and livestock farmers that are already organic. Granting a higher support, in consideration of available margins, and awarding, to a higher extent, conventional farmers located in environmentally critical areas in both phases of application (introduction and maintenance) could have produced better results in terms of achievement of environmental objectives. Looking at the 2014-2020 RDP of the Marche Region, it turns out that, consistently with our suggestions, premium levels were increased by about 20% and the scoring system was revised by giving higher priority to conventional farmers with intensive production, who decide to convert to organic farming (Regione Marche, 2017a,b). However, territorial differentiation of premiums was completely abolished and only a single premium, though distinguished by type of crop, was introduced. Moreover, with reference to the maintenance of organic farming, in selecting applications, a lower priority was attributed to farmers with intensive production, giving more importance to those located in Natura 2000 sites, where constraints are surely more stringent, but the environmental benefits of organic farming are also lower. Therefore, although a few improvements have been made in comparison with the previous programming period, some policy choices are still questionable and should be re-discussed to better target policy support in relation to environmental issues.

In interpreting results, some caution should be taken owing to a few possible limitations. Besides potentially different results that can derive from applying alternative models (such as ZIP, hurdle, negative binomial or COM-Poisson models), a possible drawback is that the regression analysis applied might suffer from endogeneity. Technically, this problem occurs when a predictor variable in the model is correlated with the error term. This can happen for a variety of reasons (Fingleton and Le Gallo, 2008). However, the most relevant one in this study is the possibility that the outcome variable is also a predictor and not only a response (the so-called "simultaneity bias"). Firstly, simultaneity bias may concern factors related to the prevailing farming system, since the characteristics of existing farms may reflect those of organic farms, especially in the cases where organic farms represent the majority. However, considering that organic farms only represent 5% of total farms and that the percentage of organic farms is higher than half only in less than 2% of cells where there are agricultural producers, the influence of organic farming on the existing farming system, if present, appears to be rather limited. Endogeneity issues can also concern environmental factors such as the levels of erosion and the content of organic matter. This is because the presence of organic farms could help to reduce soil erosion and to increase or, at least, to maintain the percentage of organic matter contained in soils. However, the results related to soil erosion indicate the contrary, i.e. organic farms tend to localize where the levels of erosion are higher. Moreover, observing the territorial distribution of soil organic matter (Figure A1-i), the content of organic matter increases as we move towards the hinterland. This highlights a natural phenomenon due to the morphological characteristics of the region, which is possibly affected by the distribution of all agricultural activities, rather than a result of organic farming. Endogeneity can also be concerned with the variable related to policy. In fact, a higher concentration of organic farms could be the reason for a higher incidence of policy support per hectare. In this case, it is like stating that organic farms are not incentivized by policy, but they recur to policy support only because there exists this financing opportunity. In other words, a farm becomes organic also without policy. This would be an even stronger conclusion which does not take into account the selective criteria and the obstacles that exist when applying for receiving policy support. Nonetheless, although the possibility of some endogeneity still remains, the main objective of this study is to investigate contextual factors that are more favourable to the concentration of organic farming rather than analysing causal relationships at level of single farm, for which the issues related to endogeneity can be more relevant.

5. Concluding remarks

This study has analysed the spatial distribution of organic farms in the Marche, an Italian rural region. This region is an interesting case because, in consideration of its peculiar environmental characteristics, organic farming could give an important contribution to mitigating or avoiding current and potential environmental impacts that agriculture generates. In particular, this analysis has assessed whether and the extent to which the distribution of organic farms is related to some economic, social, environmental and political factors. To this aim, a quasi-Poisson CAR-HGLM model has been adopted. This model provides high flexibility since it does not make any assumption about the distribution of zeros, allows for both under- and overdispersion, and takes account of spatial dependence in measuring the influence of some possible explanatory factors.

Results indicate that there is a tendency to spatial concentration, i.e. organic farming develops where there already exist other organic farms. Besides neighbouring effects, there are other factors affecting the spatial distribution of organic farming. Indeed, results show that organic farms concentrate where there are more favourable market conditions, i.e. an easier access to consumers and a higher propensity to purchase organic products. In addition, organic farms tend to localize far from the areas where there is a more intensive use of mechanization and chemicals, and where they could give an important contribution to attenuating environmental pressure. Factors such as different conversion costs and low incentives could have acted on this spatial distribution. From a policy perspective, organic farms reveal to be sensitive to public intervention. Policy appears to be effective in stimulating organic farming where there are more problems of erosion, but it seems to be ineffective where the environmental pressure exerted by intensive farming is higher. This is attributed to an inadequate territorial differentiation of support and to the scoring system for selecting applications, in addition to low levels of premiums. For ensuring more consistency with environmental objectives, it is suggested that payments in favour of organic farming should be increased and that the management system must be revised, by territorially differentiating incentives to a larger extent and by giving more priority to farmers with intensive production.

6. Acknowledgments

Although this article is common to both authors, the authorship can be attributed as follows: Sections 1, 2.4, 3, 4, 5 to Andrea Bonfiglio; Sections 2.1, 2.2 and 2.3 to Andrea Arzeni.

The authors are particularly grateful to three anonymous reviewers for their valuable comments and suggestions, which contributed to significantly improving the quality of this manuscript.

7. References

Agarwal, D.K., Gelfand, A.E. and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* 9(4): 341–355.

Allaire, G., Poméon, T., Maigné, E., Cahuzac, E., Simioni, M. and Desjeux, Y. (2015). Territorial analysis of the diffusion of organic farming in France: Between heterogeneity and spatial dependence. *Ecological Indicators* 59: 70–81.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27(3): 247–267.

Arnhold, S., Lindner, S., Lee, B., Martin, E., Kettering, J., Nguyen, T.T., Koellner, T., Ok, Y.S. and Huwe, B. (2014). Conventional and organic farming: Soil erosion and conservation potential for row crop cultivation. *Geoderma* 219–220: 89–105.

Basile, S. (2014). Bio-distretti: istruzioni per l'uso. *Bioagricultura* 145–146: 4–8.

Besag, J. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician* 24(3): 179–195.

Bjørkhaug, H. and Blekesaune, A. (2013). Development of organic farming in Norway: A statistical analysis of neighbourhood effects. *Geoforum* 45: 201–210.

Bonfiglio, A., Arzeni, A. and Bodini, A. (2017). Assessing eco-efficiency of arable farms in rural areas. *Agricultural Systems* 151: 114–125.

Cameron, A.C. and Windmeijer, F.A.G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business and Economic Statistics* 14(2): 209–220.

Clayton, D. and Berardinelli, L. (1996). Bayesian methods for mapping disease risk. In Elliott, P., Cuzick, J., English, D. and Stern, R. (eds.), *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*. Oxford University Press, 205–220.

Conway, R.W. and Maxwell, W.L. (1962). A Queuing Model with State Dependent Service Rates. *Journal of Industrial Engineering* (12): 132–136.

Council of European Union (2007). Council Regulation (EC) No 834/2007 of 28 June 2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91. *Official Journal of the European Communities L* 189: 1–23.

Cox, D.R. (1983). Some Remarks on Overdispersion. *Biometrika* 70(1): 269.

Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica* 39(5): 829–844.

Cressie, N.A. (1995). Statistics for Spatial Data. Revised Edition. New York: John Wiley & Sons, Inc.

Dénes, F. V., Silveira, L.F. and Beissinger, S.R. (2015). Estimating abundance of unmarked animal populations: Accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution* 6(5): 543–556.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D.W., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. and Wilson, R. (2007). Methods to account

for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* 30(5): 609–628.
European Court of Auditors (2011). Is agri-environment support well designed and managed? Special Report No 7, Luxembourg.
Eurostat (2016). Population grids. Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_grids.
Fingleton, B. and Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science* 87(3): 319–339.
Frederiksen, P.I.A. and Langer, V. (2004). Localisation and concentration of organic farming in the 1990s - The Danish case. *Tijdschrift voor economische en sociale geografie* 95(5): 539–549.
Gabriel, D., Carver, S.J., Durham, H., Kunin, W.E., Palmer, R.C., Sait, S.M., Stagl, S. and Benton, T.G. (2009). The spatial aggregation of organic farming in England and its underlying environmental correlates. *Journal of Applied Ecology* 46(2): 323–333.
Genius, M., Pantzios, C.J. and Tzouvelekas, V. (2006). Information Acquisition and Adoption of Organic Farming Practices. *Journal of Agricultural and Resource Economics* 31(1): 93–113.
Glaser, S. (2017). A review of spatial econometric models for count data. Hohenheim Discussion Papers in Business, Economics and Social Sciences No. 19–2017, Available at: http://nbn-resolving.de/urn:nbn:de:bsz:100-opus-13975.
Green, W.H. (2012). Econometric Analysis - Seventh Edition. Edinburgh Gate, Harlow, England: Pearson Education Limited.
Häring, A.M., Dabbert, S., Aurbacher, J., Bichler, B., Eichert, C., Gambelli, D., Lampkin, N.H., Offermann, F., Olmos, S., Tuson, J. and Zanoli, R. (2004). Organic farming and measures for European agricultural policy. Organic Farming in Europe: Economics and Policy, Vol. 11. Stuttgart, Germany: University of Hohenheim/Department of Farm Economics.
Hiederer, R. (2012). EFSA Spatial Data Version 1.1 Data Properties and Processing. Publications Office of the European Union.
Hodges, J.S. (2014). Richly Parametarized Linear Models: Additive Linear and Time Series and Spatial Models Using Random Effects. Boca Raton, FL: Taylor & Francis Group, LCC.
Ver Hoef, J.M. and Boveng, P.L. (2007). Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data? *Ecology* 88(11): 2766–2772.
van der Horst, D. (2007). Assessing the efficiency gains of improved spatial targeting of policy interventions; the example of an agri-environmental scheme. *Journal of Environmental Management* 85(4): 1076–1087.
Hu, M.-C., Pavlicova, M. and Nunes, E. V. (2011). Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial. *The American Journal of Drug and Alcohol Abuse* 37(5): 367–375.
INSPIRE (2014). D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines. Available at: https://inspire.ec.europa.eu/file/1725/download?token=2-WQia90.

ISTAT (2015). La struttura delle aziende agricole. Anno 2013. Rome.

ISTAT (2008). Struttura e produzioni delle aziende agricole. Anno 2007. Rome.

Jouzi, Z., Azadi, H., Taheri, F., Zarafshani, K., Gebrehiwot, K., Van Passel, S. and Lebailly, P. (2017). Organic Farming and Small-Scale Farmers: Main Opportunities and Challenges. *Ecological Economics* 132: 144–154.

Kim, H. and Lim, H. (2010). Comparison of Bayesian Spatio-Temporal Models. *Journal of Data Science* 8: 189–211.

Koesling, M., Flaten, O. and Lien, G. (2008). Factors influencing the conversion to organic farming in Norway. *International Journal of Agricultural Resources, Governance and Ecology* 7(1–2): 78–95.

Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* 34(1): 1–14.

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

Läpple, D. and van Rensburg, T. (2011). Adoption of organic farming: Are there differences between early and late adoption? *Ecological Economics* 70(7): 1406–1414.

Lee, W. and Lee, Y. (2012). Modifications of REML algorithm for HGLMs. *Statistics and Computing* 22(4): 959–966.

Lee, Y., Alam, M.M., Noh, M., Rönnegård, L. and Skarin, A. (2016). Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts. *Ecology and Evolution* 6(19): 7047–7056.

Lewis, D.J., Barham, B.L. and Robinson, B. (2011). Are there spatial spillovers in the adoption of clean technology? The case of organic dairy farming. *Land Economics* 87(2): 250–267.

Ma, Z., Zuckerberg, B., Porter, W.F. and Zhang, L. (2012). Spatial Poisson Models for Examining the Influence of Climate and Land Cover Pattern on Bird Species Richness. *Forest Science* 58(1): 61–74.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37(1–2): 17–23.

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

Oh, J., Washington, S.P. and Nam, D. (2006). Accident prediction model for railway-highway interfaces. *Accident Analysis and Prevention* 38(2): 346–356.

Panagos, P., Borrelli, P., Poesen, J., Ballabio, C., Lugato, E., Meusburger, K., Montanarella, L. and Alewell, C. (2015). The new assessment of soil loss by water erosion in Europe. *Environmental Science & Policy* 54: 438–447.

Parker, D.C. and Munroe, D.K. (2007). The geography of market failure: Edge-effect externalities and the location and production patterns of organic farming. *Ecological Economics* 60(4): 821–833.

Paudel, G.S. and Thapa, G.B. (2004). Impact of social, institutional and ecological factors on land management practices in mountain watersheds of Nepal. *Applied Geography* 24(1): 35–55.

Pietola, K.S. and Lansink, A. (2001). Farmer response to policies promoting organic farming technologies in Finland. *European Review of Agriculture Economics* 28(1): 1–15.

Pingel, T. (2018). The Raster Data Model. In Wilson, J. P. (ed.), *Geographic Information Science & Technology Body of Knowledge*.

Reeve, J.R., Hoagland, L.A., Villalba, J.J., Carr, P.M., Atucha, A., Cambardella, C., Davis, D.R. and Delate, K. (2016). Organic farming, soil health, and food quality: Considering possible links. *Advances in Agronomy* 137: 319–367.

Regione Marche (2017a). Reg. UE n. 1307/2013 – PSR Marche 2014-2020 – Bando - Sotto Misura 11.1 "Pagamento al fine di adottare pratiche e metodi di produzione biologica" – Annualità 2017. Decree of the Director of Environment and Agriculture No 157 of 24 April 2017, Ancona.

Regione Marche (2017b). Reg. UE n. 1307/2013 – PSR Marche 2014-2020 – Bando - Sotto Misura 11.2 "Pagamenti per il mantenimento dei metodi di produzione biologica" – Annualità 2017. Decree of the Director of Environment and Agriculture No 157 of 24 April 2017, Ancona.

Regione Marche (2016a). European Agricultural Funds for Rural Development EAFRD. Financial Implementation report 2015: Marche - Programma di Sviluppo rurale 2007–2013.

Regione Marche (2016b). PSR 2007-2013. Relazione annuale di esecuzione. Anno 2015.

Regione Marche (2015a). Italy - Rural Development Programme (Regional) - Marche.

Regione Marche (2015b). Reg. CE n. 1698/05 – PSR Marche 2007-2013 – Bando - Misure 2.1.4. sottomisura b) sostegno all'agricoltura biologica – Campagna 2015. Decree of the Director of Environment and Agriculture No 303 of 5 May 2015, Ancona.

Ridout, M.S. and Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling* 4: 77–89.

Roe, B., Irwin, E.G. and Sharp, J.S. (2002). Pigs in space: Modeling the spatial structure of hog production in traditional and nontraditional production regions. *American Journal of Agricultural Economics* 84(2): 259–278.

Rusco, E., Maréchal, B., Tiberi, M., Bernacconi, C., Ciabocco, G., Ricci, P. and Spurio, E. (2008). Case study report (WP2 findings) - Italy. Case studies on soil/land management and policy measures - SoCo Project: Sustainable agriculture and soil conservation. Available at: https://esdac.jrc.ec.europa.eu/projects/SOCO/Case Studies/casestudyIT_002.pdf.

Sanders, J., Stolze, M. and Padel, S. (2011). Use and efficiency of public support measures addressing organic farming. Study Report, Johann Heinrich von Thünen Institut (vTI) – Institute of Farm Economics.

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

SINAB (2015). BIO in cifre 2015. Available at: http://www.sinab.it/content/bio-statistiche.

SINAB (2008). BIO in cifre 2008. Available at: http://www.sinab.it/content/bio-statistiche.

SINAB (2001). BIO in cifre 2001. Available at: http://www.sinab.it/content/bio-statistiche.

Taus, A., Ogneva-Himmelberger, Y. and Rogan, J. (2013). Conversion to Organic Farming in the Continental United States: A Geographically Weighted Regression Analysis. *The Professional Geographer* 65(1): 87–102.

Tin, A. (2008). Modeling zero-inflated count data with underdispersion and overdispersion. In SAS Institute Inc. (ed.), *Proceedings of the SAS Global Forum 2008 Conference*. Cary, NC, USA: SAS Institute Inc.

United Nations (2007). Managing Statistical Confidentiality & Microdata access. Principles and guidelines of good practice. New York and Geneva: United Nations publication.

Uthes, S., Matzdorf, B., Müller, K. and Kaechele, H. (2010). Spatial targeting of agri-environmental measures: Cost-effectiveness and distributional consequences. *Environmental Management* 46(3): 494–509.

Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* 121(2): 311–324.

Wang, X.C., Kockelman, K.M. and Lemp, J.D. (2012). The dynamic spatial multinomial probit model: Analysis of land use change using parcel-level data. *Journal of Transport Geography* 24: 77–88.

Wang, Y., Kockelman, K.M. and Damien, P. (2014). A spatial autoregressive multinomial probit model for anticipating land-use change in Austin, Texas. *Annals of Regional Science* 52(1): 251–278.

Whittle, P. (1964). On stationary process in the plane. *Biometrika* (41): 434–449.

Wier, M., O'Doherty Jensen, K., Andersen, L.M. and Millock, K. (2008). The character of demand in mature organic food markets: Great Britain and Denmark compared. *Food Policy* 33(5): 406–421.

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

Zaccarini Bonelli, C. (2011). Bruxelles premia l'agricoltura biologica. *PianetaPSR* 4. Available at: http://www.pianetapsr.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/331.

Zepeda, L. and Li, J. (2007). Characteristics of organic food shoppers. *Journal of Agricultural and Applied Economics* 39(1): 17–28.

Zuur, A.F., Saveliev, A.A. and Ieno, E.N. (2012). Zero inflated models and generalized linear mixed models with R. Highland Statistics Ltd.