SPARD

Spatial Analysis of Rural Development Measures
Contract No. 244944

Work Package No. 5
August 2012
D5.2 Estimated models in case study areas
CSA ITALY
F. Bartolini, M. Raggi, D. Viaggi

Document status:

Public use	NO
Confidential use	X
Draft No. 1	
Final	Date
Submitted for internal review	Date

COOPERATION

Contents

List of tables

Eliminato: 19

Eliminato: 20

Eliminato: 21

Eliminato: 26

List of figures

1 Introduction

This document reports the outcome of Task 5.2 of the SPARD project related to the Case Study Emilia Romagna.

2 Background information of the analysed RDP measures in the case study region

The three analysed measures are 121 (Modernisation), 214 (agri-environment) and 311 (farm diversification). The three measures where each the most prominent in the respective axis of the RDP (Rural Development Program) in terms of participation and where selected according to the common SPARD criteria.

In Emilia Romagna measure 214 is organised in several submeasures, of which the following were also modelled separately:

- Sub-measure 1: Integrated production
- Sub-measure 2: Organic agriculture
- Sub-measure 8: Meadow and grazing payments
- Sub-measure 9: Conservation of natural areas and landscape
- Sub-measure 10: Environmental set aside

3 Cross-measure issues in setting up the analysis

Information about participation to RDP were delivered by the Emilia Romagna Regional administration, Agricultural directorate and where derived from the databases used as a basis for program monitoring, reporting and evaluation.

The data available were heterogeneous across measure, a most of them are asked at the time of the application and reflect the specific needs of the measure involved. Location and measure characteristics is always present. For land-based measure (e.g. 214) also detailed information about the identification of parcels participating to the measure were available, as well as payment information.

On the contrary socio-economic and structural information on the overall participating farms were not available. Also individual information about non-participants characteristics was not available.

Data concerning participation and dependent variables are aggregated at municipality level. Such level allows to use data available with several information about determinants of participation and to have enough observations to conduct spatial analysis. For each measure, participation in the RDP 2006-2013 are aggregated at the last available years, coherently with the call opened (eg. We have used only the first call of the measure 121, up to year 2011)

4 Model estimation

4.1 Modernisation (Measure 121)[1]

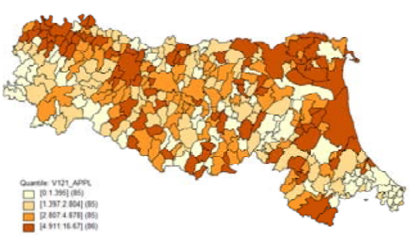
4.1.1 General features and participation

Measure 121 provides several measures for agriculture modernisation. In this section, we analyse the participation rate in the measure 121 at municipality level, highlighting as a determinant, the effect of the set of priorities established by local administration. The analysis is realised in two steps: first a Exploratory Spatial Description Analysis (ESDA) of the participation rate in Emilia Romagna (ER) municipalities is realised and then a spatial regression model (with several assumptions of the spatial weights matrix) explaining the participation rate to the measure 121 is applied.

Descriptive statistics are presented in annex 1.

4.1.2 Spatial analysis

The distribution of percentage of applicants and the percentage of beneficiaries in each municipality of the Emilia Romagna are presented in figure 1.



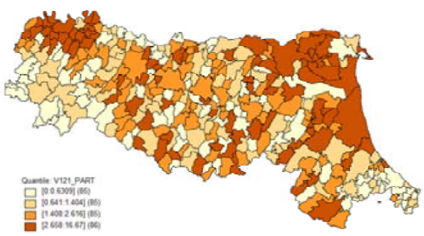


Figure 1A: Cloropeth map of applicants ; Figure 1B: Cloropeth map of beneficiaries

The maps shown the percentage of applicants to measure 121 (Figure 1A) and the participation rate for each municipality of Emilia Romagna Region (Figure 1B). In both figure applicants and participations rate are grouped in quartiles. The results shown the value of both applicants and participation are included between zero (which means that no farmers applied for the measure in the municipality) and 16%, which represent the maximum applicants rate and participation rate.

As shown by the maps, the highest applications rate are localised in the north-west part and in the Adriatic board (eastern part). The part of the region with higher percentage of applicants is located in the plain area of the Piacenza, Parma, Reggio Emilia province and in Ferrara and Ravenna provinces.

[1] This paragraph is based on Bartolini et al., (2012), presented at the I AIEAA 2012 conference.

The pictures provided by figure 1A could be representative also for the participation rate (figure1B). In fact, generally there is a correspondence between area with high participation and area with high applicants. This is particular evident in the Piacenza Province (north western part) and in the municipalities of Ferrara and Ravenna Provinces. Only in the central part of the region there is a high amount of excluded farm (see figure 2).

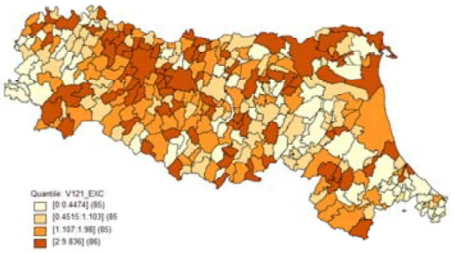


Figure 2: Cloropeth map of not eligible farmers (applicants - beneficiaries)

In this part of the chapter the results of the Moran scatter plots and the results of the LISA are presented.

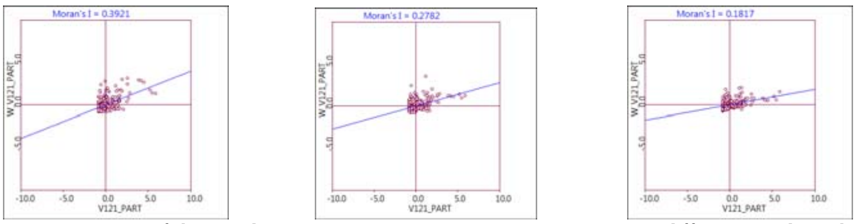


Figure 3 Moran I of the cumulative participation at 2011 to measure 121 using different spatial weight matrix (queen 1, queen 2 and queen 3) respectively.

The figure provides the value of global Moran'I and the Moran scatterplots under three hypotheses of spatial weights matrix. In each scatterplot on the *x*-axis the deviation from the mean for the observed value of the participation rate is presented, while in *y*-axis the average value of the deviation from the mean of the neighbouring observation are shown. As pointed out in the methodology the location in the four quadrants represent a regime of spatial association. Despite, due to the low value of the average of the participation rate the greater part of municipality are close to the average, and only few of municipality are clearly distant from the origins of the axis and are positioned in the high and right quadrants. Several municipalities are located in the top and right quadrant which represent a hot spot clustering.

The same queen contiguity is used, but is assumes different orders of contiguities. From the left to the right is provide the value of Moran 's I and the Moran scatterplot increasing the order of contiguity from 1 to 3. The value on the top on each scatterplot shows the global Moran's I value. In all three figures a Moran's I value greater than zero shows that there is some positive spatial association in the participation rate. In other words the participation rate of one municipality is positively influenced by the participation rate of the neighbouring region. Under the hypothesis of the first order of contiguity the value of Moran's I is quite high, about 0.40, which means that there is a relevant spatial pattern on the participation. The value is differentiate according to the spatial weight considered. Increasing the level of queen contiguity the spatial association is reduced.

Figure 3 shown that increasing the order of contiguity a reduction of the deviation from the mean is observed. This is highlighted by the more closeness of the observation to the x-axis.

In figure 4 the result of LISA are provided.

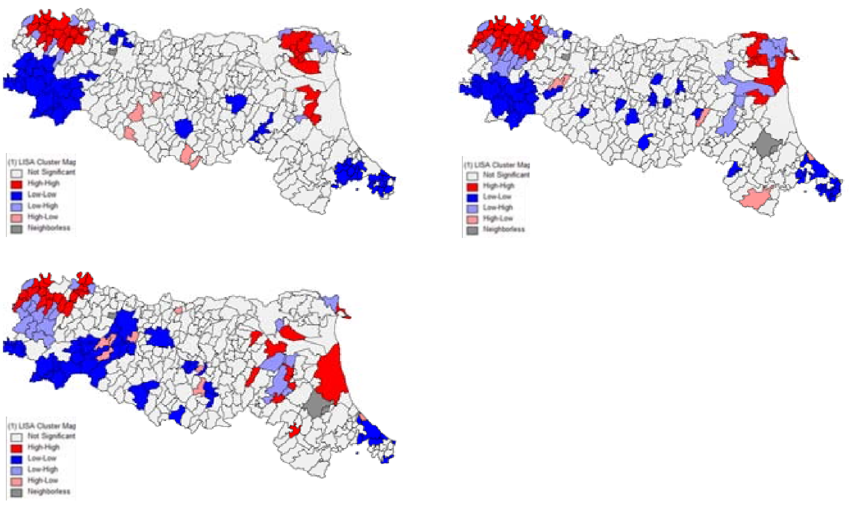


Figure 4. LISA of the Cumulative participation at 2011 to measure 121, using queen contiguity matrix of the first (A), second (B) and third level (C).

Eliminato: 3

Each map shows the spatial cluster obtained by LISA. All the painted municipalities are those municipalities which are significant at least at 0.05. The red colours represent a hot spot cluster, while the Blu the "cold spot" cluster and the pink and the sky-blu the spatial outliers.

Figure 4 confirms that the province of Piacenza, Parma, Ferrana and Ravenna have several municipalities with higher participation rate and this also applies to the neighbours. In the case of the first order of contiguity, it is possible to note that the province of Piacenza and Parma have very high heterogeneity within the territory of the province. In fact in these two provinces there are both hot spots and cold spots clusters. Otherwise, in the rest of the territory there are different regime of spatial associations. In fact it is possible to see only very distant clusters about the two regimes. Finally, there are few spatial outliers. These are mainly closer to hot spot clusters and represent a group of municipalities with lower participation closer to municipality with higher participation rate. In addition there are also few and very spread outliers with higher participation rate compared to the neighbouring municipality (pink cluster)

With higher order of the contiguity the cluster is quite differentiated, especially in the composition of the outliers. In fact under the hypothesis of second or third order the outliers increasing and are mainly those closer to the hot or cold spot clusters.

4.1.3 Modelling strategy

Participation to RDP measures has been explained using several econometrics techniques, in which the heterogeneity of the methods are generally a consequences of the data available and the analysis of the participation to the RDP measures are strongly data-driven. In fact the varieties of models applied follows the difference concerning the units observed (payments, farms, etc.), the territorial level observed (individual farm; several geographical areas; etc) and the timing of available observations (all RDP time programming; yearly etc.). The most used methods are panel data using FADN data or regression analysis focusing on payments or participation rate at one specific territorial level.

Results highlight that farmers attitude, farmers characteristics, territorial/geographical features, quality and efficiency of institutions involved and quality of consultancy services are determinants of difference in participations.

Most of these variables could have a spatial pattern, which could assume the form of spatial dependence between observations and/or spatial heterogeneity in the model (LeSage and Pace, 2009). As a results spatial location is relevant and when dependent variable are affected by the space, is violated the Gauss-Markov assumptions used in regression modelling (LeSage and Pace 2009). In fact, when in many economic processes are considered proximity or distance functions the variables observed are not independent of each other (Brady and Irwin, 2011). Concerning the participation to RDP measures, location could affect quality and efficiency of local institution, perception by farmers, agricultural systems and quality of advises and consultancy services.

Following LeSage and Pace 2009, the spatial dependency could be modelled as an extension of the standard linear regression model. As a result, the regression could be written as (Breustedt and Habermann, 2011):

$$\begin{aligned} r &= \rho W_1 r + X\beta + \varepsilon \\ \varepsilon &= \lambda W_2 \varepsilon + \mu \qquad (2) \\ E\left[\mu_i^2\right] &= \sigma^2 h(z_i) \\ E\left[\mu_i \mu_j\right] &= 0 \text{ with } i \neq j \end{aligned}$$

Where r is the observed participation rate; X is the $n \times k$ matrix of the k determinants of the participation rate, β is the regression parameter to be estimated, ε is the error term, W_1 and W_2 are the $n \times n$ matrix of spatial weights; ρ are the spatial lag parameter; and λ spatial error coefficient. Where ith element of $W_1 r$ represent the spatial weighted average of the participation rate for municipality i and $W_2\varepsilon$ are the error lag and represent a specification of the error term.

Under several assumptions about of the ρ and λ the equation 1 could yield:

with $\rho = 0$; $\lambda = 0$ the equations return a standard linear regression model (model 1);

with $\lambda = 0$; the equations return a spatial lag model (model 2);

with $\rho = 0$;the equations return a spatial error model (model 3);

Spatial lag model and spatial error model take into account differently the spatial patterns of the participation to RDP measures. In a spatial lag model it is assumed that participation of one area is affected by the participation of neighbouring areas; while in the spatial error model, some unknown variables shared with the neighbourhood influence the participation rate.

4.1.4 Modelling results

In the table 3 the results of the three regression models are shown.

Table 3. Model results.

	model 1		model 2						model 3					
			queen 1		queen 2		queen 3		queen 1		queen 2		queen 3	
Variable	Coeff	sign	coeff	sign	coeff	sign	coeff	sign	coeff	sign	coeff	sign	coeff	sign
d_lfa	-0.5007	**	-0.4339	*	-0.4245	*	-0.4709	**	-0.4071	*	-0.4257	*	-0.5261	**
dens_ab	0.0000		0.0002		0.0002		0.0002		0.0002		0.0001		0.0002	
only_hhlab	-0.0252	*	-0.0199		-0.0197		-0.0257	*	-0.0267	*	-0.0247	*	-0.0276	**
cond_dir	0.0477	**	0.0455	**	0.0494	***	0.0490	***	0.0485	***	0.0506	***	0.0480	***
part_colle	0.6665	**	0.6916	**	0.6698	**	0.6763	**	0.8007	***	0.7352	***	0.6457	**
corpi_av	-0.0015		0.0042		0.0025		-0.0023		0.0020		0.0029		0.0009	
Potsuccess	-0.0067		-0.0057		-0.0045		-0.0077		0.0069		-0.0022		-0.0088	
Arable	-0.0093		-0.0072		-0.0073		-0.0078		-0.0038		-0.0059		-0.0088	
Fruit	-0.0075		-0.0065		-0.0052		-0.0067		-0.0072		-0.0073		-0.0090	
Grazing	-0.0001		-0.0001		0.0024		0.0019		-0.0017		0.0008		-0.0012	
Forest	0.0072		0.0047		0.0043		0.0054		0.0000		0.0012		0.0068	
Livestock	-0.0108		-0.0094		-0.0074		-0.0071		-0.0022		-0.0034		-0.0093	
Pigs	-0.0647	**	-0.0504	**	-0.0479	*	-0.0563	**	-0.0533	**	-0.0445	*	-0.0536	**
prop_mach	0.0103		0.0088		0.0089		0.0080		0.0080		0.0110		0.0096	
tractor_le	-0.0170	**	-0.0171	***	-0.0182	***	-0.0183	***	-0.0175	***	-0.0170	**	-0.0176	***
tractor_mo	0.0009		0.0018		0.0034		0.0020		-0.0013		0.0000		-0.0005	
labcon_sup	-0.0300		-0.0227		-0.0269		-0.0271		-0.0228		-0.0271		-0.0267	
young	0.0714	*	0.0581		0.0521		0.0654	*	0.0793	**	0.0551		0.0714	*
age_more65	-0.0002		0.0054		-0.0003		0.0017		-0.0003		-0.0049		0.0013	
partime	-0.0165		-0.0141		-0.0151		-0.0144		-0.0059		-0.0141		-0.0165	
Agredu	-0.0156		-0.0134		-0.0099		-0.0122		-0.0077		-0.0083		-0.0132	
uaa_more50	0.2539	***	0.2280	***	0.2459	***	0.2343	***	0.2667	***	0.2804	***	0.2494	***
ave_farmsize	0.0171		0.0073		-0.0003		0.0139		0.0008		-0.0138		0.0104	
sau_sup	2.7057	**	2.0641	*	2.3008	**	2.4869	**	1.5448		2.0352	*	2.6825	**
prob_crops	0.0836	***	0.0768	***	0.0790	***	0.0811	***	0.0758	***	0.0794	***	0.0845	***
_cons	-1.9035		-2.4285		-2.8863		-2.2222		-2.8602		-2.0341		-1.4382	
RHO			0.2515	***	0.2598	***	0.2249	**						
LAMBDA									0.3130	***	0.3206	***	0.2244	
R2	0.45		0.53		0.52		0.51		0.53		0.51		0.49	

The first model refers to the standard linear regression, while model 2 and 3 refer to the spatial lag model and spatial error models respectively. Those last two models are performed using three different orders of queen contiguity.

The first model highlights that the variables which contribute positively to the participation rate are those connected with the farm structure such as farm size and the percentage of farmers with large farm and labour such as the direct conduction of the farm. In addition municipality with high percentage of young farmers and high percentage of farms organised in associative legal status have higher participation rate. In addition the priority score of the municipality, concerning the ranking of farms for the priorities of the measure 121 increase strongly the participation rate. The first model shows that municipalities with high percentage of farmers invest less capital and with high percentage of farm which use only household labour have lower participation rate. In addition municipalities with high percentage of farms specialised in pork production and the location of the municipality in less favourable areas influence negatively the participation rate. Finally, the model return a quite satisfactory value of R2.

The second model refers to the spatial lag model. Such a model allows to include neighbouring effect in the dependent variable. With exception of the percentage of the young farm, the significant variables are the same and have the same sign, but with a lower magnitude than the model 1. The model provides an estimation of the spatial lag coefficient, which is about 0.25 for the first two order of the contiguity and is reduced for the third order of contiguity. Moving to the spatial lag model the R2 has been strongly increased, which means that the ability of the model to reproduce the observed with the inclusion of the spatial pattern is increased.

The third model refers to the spatial error model. Such model allows to include the spatial patterns into the error term, which means that the error term have a component which is spatially distributed. With exception of the ratio between uaa_taa in each municipality, the significant variable are the same and have the same sign, but with a different magnitude with respect the model 1 and model 2. Different to the model 2 increasing the order of contiguity there a reduction of significance in the spatial error parameters, and in the case of the third order is not significant. Again, adding the spatial error in the error term there is a strong increasing of the R2 compared to the model 1.

4.1.5 Discussion

In this section, the spatial patterns of the participation rate in the municipalities of Emilia Romagna of measure 121 are investigated. Results highlight the relevance of the spatial analysis in improving the predictability of the participation on rural development measures. In particular they show a positive effect of the neighbouring and of the spatial location in the explanation of the participation rate; the results also show the effect of the priority mechanism implemented by each province in determine the eligibility of the application.

An important results is about the spatial weight considered. In fact results are affected by the different hypothesis about the neighbourhood and the relation between locations. Altogether for a correct interpretation of the spatial issues is very important to identity a priory which form of spatial relation should improve the explanation of the observed phenomena. In our case due to the dimension of the analysis the first order of contiguity could be enough to explain the spatial pattern which affect the participation. However, additional insight about the effect of the spatial pattern in the explanation of the participation rate is needed to enrich the predictability of the participation.

Additional weakness of this analysis concerns the data availability at municipality level which could better describe territorial issues and allow for a better distillation of the spatial components by the territorial components. However, the analysis could be improved using data at farm level which allows to include further spatial effects such as for example imitation which could provide more insight in the explanation of the participations.

4.2 Agri-environmental measures (Measure 214)

4.2.1 General features and participation

Agri-environmental measures cover a substantial part of the RDP budget in Emilia Romagna. The measure 214 is organised in several sub-measures (=”sub-measure”), strongly differentiated in terms of environmental objectives and target area for application.

No surprise that the distribution on the regional area, in terms of percent of participating farms per municipality, is rather differentiated and is different between the aggregate and specific sub-measures (Figure 5 to Figure 10).

Eliminato: Figure 4

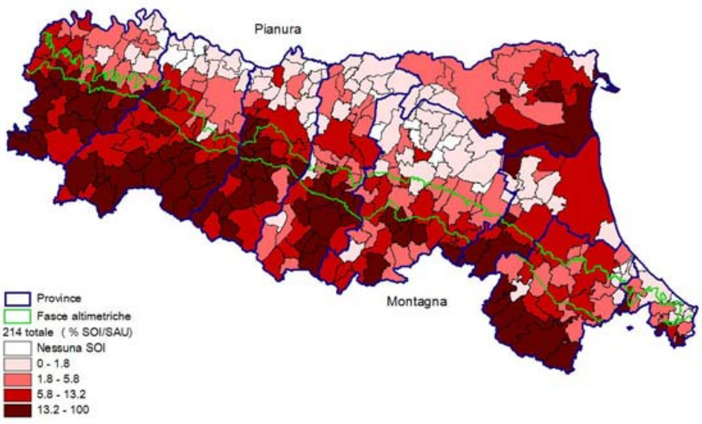
Eliminato: Figure 9

In particular, while sub-measures 1, 9 and 10 are mainly located in the plain areas, sub-measures 2 and 8 are mainly located in the hill-mountain area.

Also, the concentration of participation is very different across municipalities, which mainly follows the zoning applied.

Figure 5: Spatial distribution for measure 214 (all sub-measures)

Eliminato: *4*



In particular, sub-measure 1 (integrated production) is very focused in areas characterised by large share of fruit production (eastern part of the region). This is largely connect to a deliberate strategy of valorisation and targeting to the sector.

On the contrary, organic production (sub-measure 2) is much more contracted in hill-mount areas, characterised by more extensive systems and easies plant protection. This is true with the exception of Ferrara Province, which is a completely flat area, and the main farming systems are connected with cereal and alfalfa productions, in the municipalities with higher participation rate

Finally figure 7,8,9 shows a strong spatial agglomeration of the participation in specific areas. Such result by the application of the zoning applied to each measure. Measure 8 related to the meadow and grazing conservation are mainly distributed in the hill and mountain area, and in the planning area of the Parma and Reggio Emilia Provinces. On the contrary, participation to the measure 9 and 10 are mainly located in the plain area.

Figure 6: Spatial distribution for measure 214 (sub-measure 1: Integrated production)

Eliminato: *5*

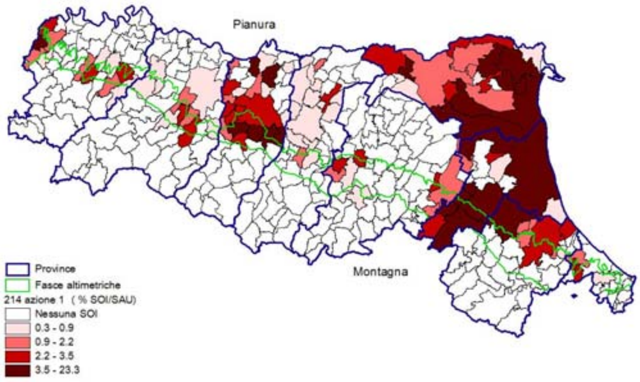


Figure 7: Spatial distribution for measure 214 (sub-measure 2: Organic production)

Eliminato: *6*

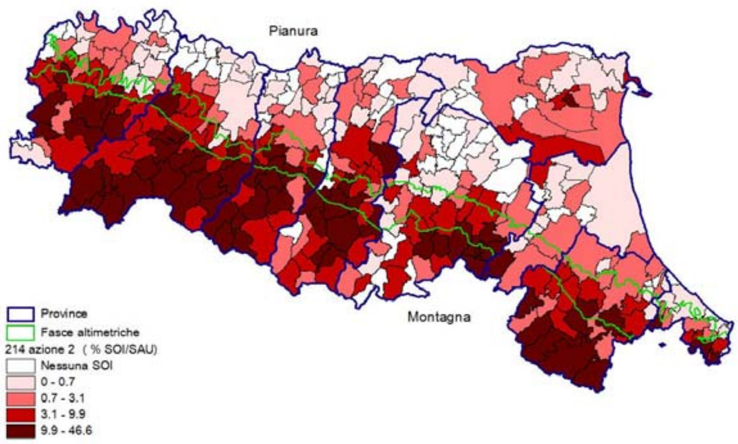


Figure 8: Spatial distribution for measure 214 (sub-measure 8: meadow and grazing payments)

Eliminato: *7*

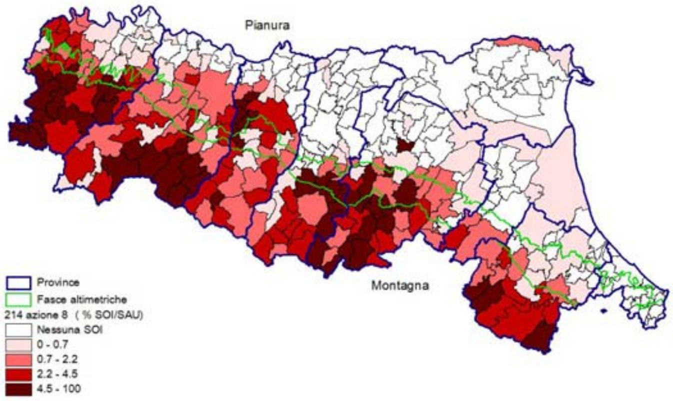


Figure 9: Spatial distribution for measure 214 (sub-measure 9: Conservation of natural and landscape elements)

Eliminato: *8*

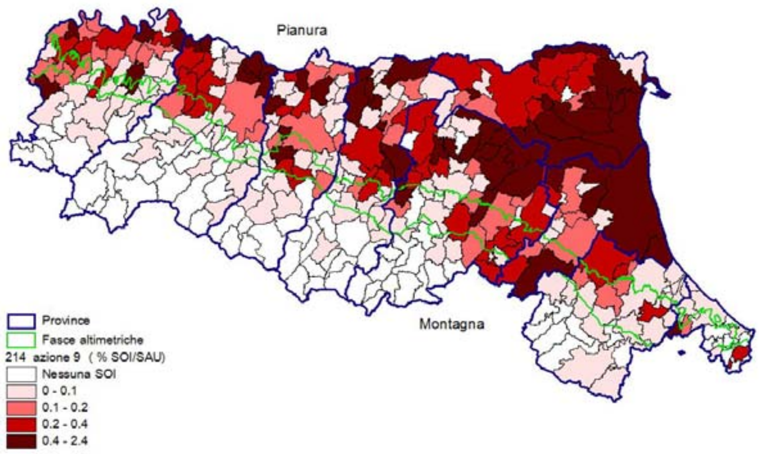
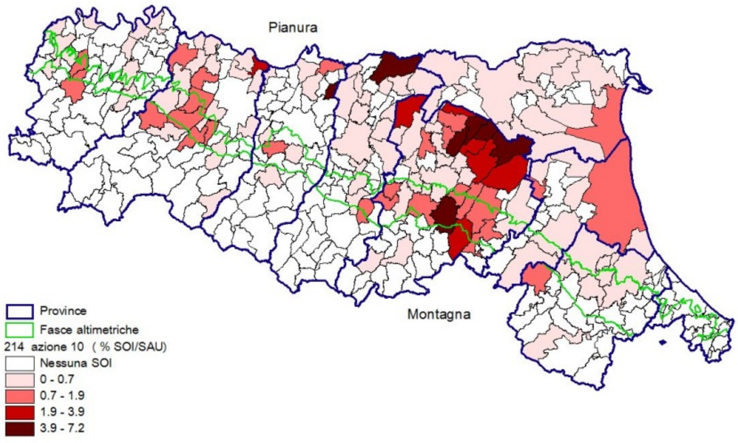


Figure 10: Spatial distribution for measure 214 (submeasure 10: twenty years environmentak set-aside)

Eliminato: *9*



4.2.2 Spatial analysis

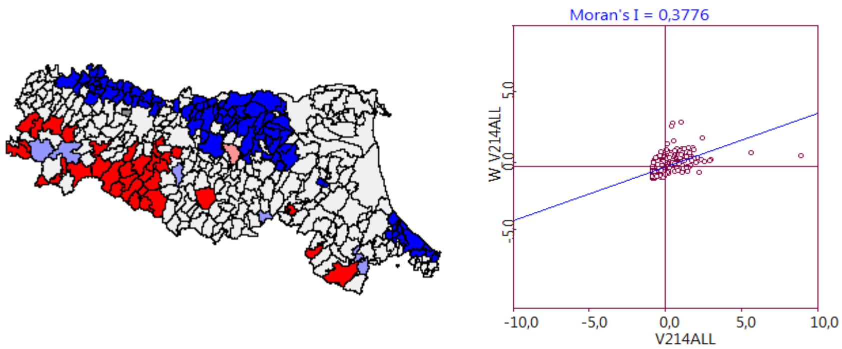
A LISA cluster map and Moran scatter plot are presented below. In all figure the dependent variable is the participation (uptake/total number of farms) in each municipality.

Observing the figure from 7 to 10 it is evident the different spatial distribution of participation in the sub-measure of 214.

The Moran Index varies from 0,28 to 0,45 representing a not so strong evidence of spatial correlation.

Figure 11: LISA Cluster Map and Moran I for measure 214 (all submeasures)

Eliminato: *10*



Two sub-measure (1 and 2) have higher Moran’s I compared to the overall. This largely reflect the concentration and occurrence of hotspots already noted in the participation map.

Figure 12: LISA Cluster Map and Moran I for measure 214 (sub-measure 1)

Eliminato: *11*

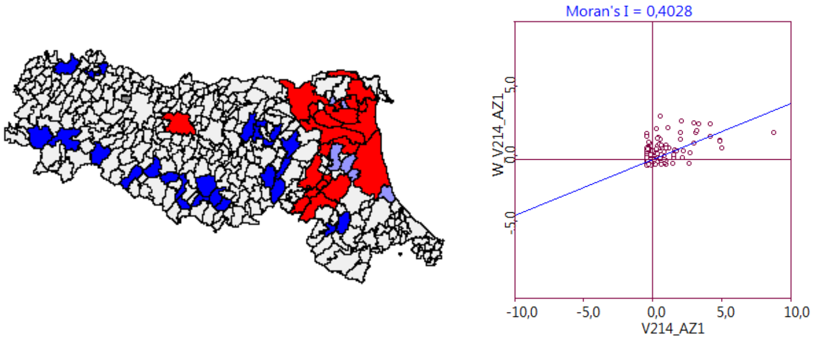


Figure 13: LISA Cluster Map and Moran I for measure 214 (sub-measure 2)

Eliminato: *12*

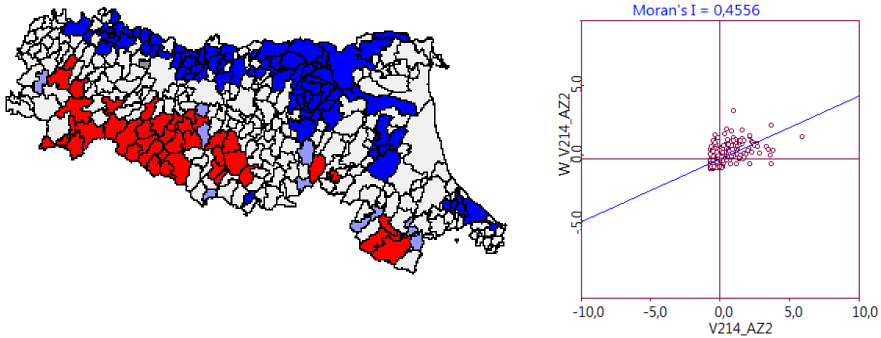


Figure 14: LISA Cluster Map and Moran I for measure 214 (sub-measure 8)

Eliminato: *13*

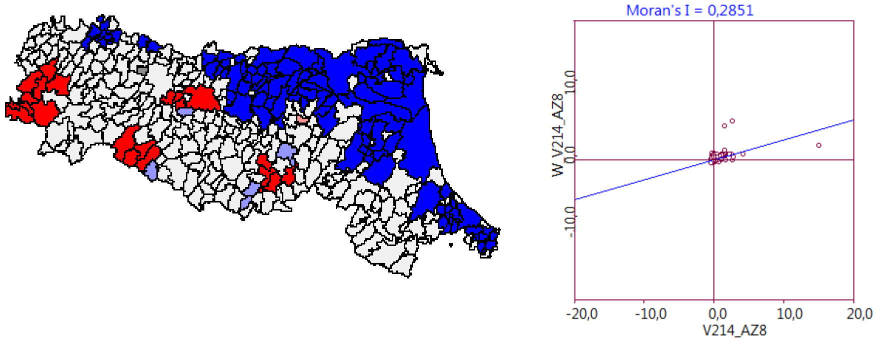


Figure 15: LISA Cluster Map and Moran I for measure 214 (sub-measure 9)

Eliminato: *14*

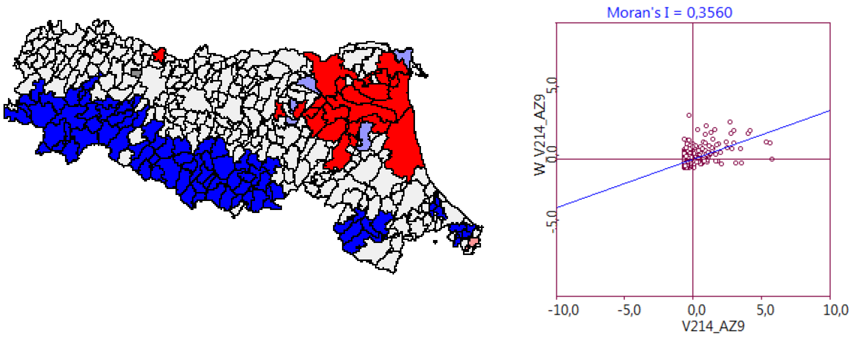
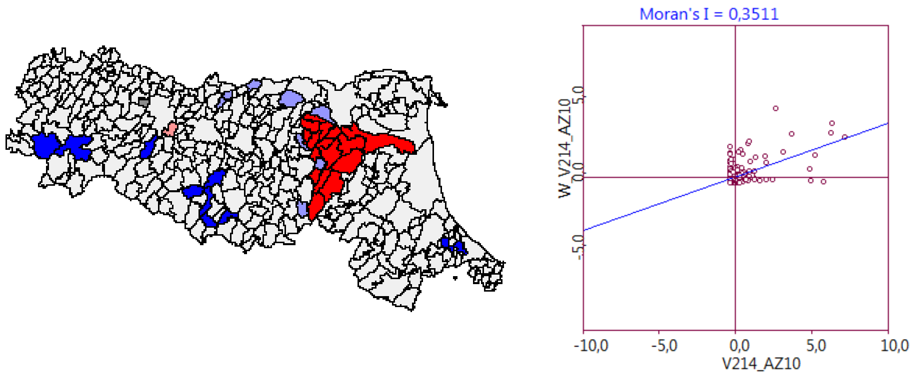


Figure 16: LISA Cluster Map and Moran I for measure 214 (sub-measure 10)

Eliminato: *15*



4.2.3 Modelling strategy

The models considered are an ordinary least square regression, a spatial lag model and a spatial error model.

The models have been applied both to the whole of measure 214 and for the selected individual sub-measures.

All models use Queen 1st order contiguity.

The model is built at municipality level (1 municipality=1 observation). The dependent variable is represented by participation and measured through the ratio between the number of participating farms and the total number of farms in the municipality.

The explanatory variables have been selected based on the preliminary analysis of expected determinants and spill-over mechanisms (see SPARD deliverable 5.1). The detail and descriptive of the explanatory variables is available in annex 1 of this document.

4.2.4 Model results

Results are presented in three tables. For each table a model is applied to data considering all 214 participation and separately for each sub-measure.

Representing in the same table the same model with same variables it is possible to compare explanatory variables for different sub-measures.

The dependent variable is for all models and all measures the participation expressed as proportion of uptake farms on the total of farms in the municipality. The number of the municipalities is 341 in the Emilia Romagna region.

In the OLS model (Table 1) Adjusted R^2 are basically satisfactory, though not very high, except in the case of sub-measure 10. No variable is significant for all sub-measures and the aggregate, but several variables are consistently relevant across several measures.

Table 1: Results of the participation models, considering an Ordinary least square model

variable	**All214**	**214-sub-measure 1**	**214-sub-measure 2**	**214-sub-measure 8**	**214-sub-measure 9**	**214-sub-measure 10**
CONS	13,0174	4,2897	**12,4616***	-2,9120	0,6110	1,4198
PIANURA	-9,8394	-0,2432	-8,4111	-1,3251	0,0328	-0,0602
COLLINA	-8,1707	0,0043	-7,0536	-1,2984	-0,1285	0,0143
MONTAGNA	-6,7816	-1,1080	-6,1824	0,3295	-0,1271	-0,0123
DENS_AB	**-0,0029***	**-0,0010****	-0,0011	-0,0004	**-0,0002*****	**-0,0004****
COND_DIR	-0,0879	0,0054	-0,0691	-0,0309	-0,0036	-0,0029
ONLY_HHLAB	**0,1050***	0,0172	0,0570	0,0279	**0,0051****	0,0049
ARABLE	**-0,0796****	**-0,0221****	-0,0084	**-0,0514****	**0,0023***	0,0053
FRUIT	0,0303	0,0022	**0,0381****	-0,0043	-0,0006	-0,0018
GRAZING	-0,0018	0,0023	**-0,0325***	0,0283	0,0001	0,0026
FOREST	**0,0852*****	-0,0085	**0,0979*****	-0,0026	**-0,0025*****	-0,0026
LIVESTOCK	**0,1245***	**-0,0583*****	**0,1204*****	0,0769	**-0,0054***	**-0,0155****
YOUNG	0,2521	0,0661	-0,0760	0,1887	-0,0030	-0,0360*
AGE_MORE65	0,0195	0,0147	-0,0446	0,0480	-0,0023	-0,0157

PARTIME	0,0072	**-0,0533*** **	-0,0078	0,0727	-0,0037	-0,0050
PREFASS	**4,3047*** **	**0,9297*** **	0,4710	**2,4660*** **	0,0460	-0,1119
PREFIDRO	-1,8020	-0,5414	-0,6447	-0,2928	**-0,1082** **	0,0099
PREFNAT	-0,7995	-0,4340	**-2,4242*** **	**2,2028*** **	-0,0008	0,1549
PREFPAE	1,4905	-0,0455	**2,0256** **	-0,4516	0,0709	0,0985
PREFSUOLO	-0,2204	0,4867	-0,2576	-0,1884	0,0267	-0,0127
R2 adjusted	0,3680	0,2150	0,4469	0,1907	0,2344	0,0419
Log likelihood	-1189,2700	-759,8040	-1038,2500	-1074,4300	-98,6045	-450,4130

The constant is significant only for sub-measure 2, the only measure most uniformly distributed across the region. Density of inhabitants is always negative, showing that participation tends to be higher in most remote areas. The presence of only household labour is relevant only on the aggregate and sub-measure 9, both with a positive effect. Share of different crops have marked different behaviour across sub-measures. The same applies to livestock.

Variables related to different preferential areas are relevant for all sub-measures, except sub-measure 10, though different preferential features apply to each sub-measure.

From Table 2, the Rho parameter (coefficient of spatial dependence) has positive value and is highly significant in all models. Also, except for the sub-measure 8, R^2 increases.

Table 2: Results of the participation models, considering a spatial lag model

variable	**All214**	**214-sub-measure 1**	**214-sub-measure 2**	**214-sub-measure 8**	**214-sub-measure 9**	**214-sub-measure 10**
CONSTANT	14,0261	3,6916	**11,4725***	0,6823	0,4450	0,6952
PIANURA	-9,0208	-0,2611	-6,6975	-2,4023	0,0377	-0,0983
COLLINA	-7,2658	0,1415	-5,4062	-2,2714	-0,1078	-0,0960
MONTAGNA	-5,9619	-0,6989	-4,6015	-0,8809	-0,1174	-0,1487
DENS_AB	-0,0024	**-0,0006** **	-0,0009	-0,0003	**-0,0002*** **	**-0,0002***
COND_DIR	-0,0816	0,0006	-0,0687	-0,0134	-0,0035	0,0024
ONLY_HHLAB	**0,0955***	0,0038	0,0514	0,0285	**0,0043***	0,0055
ARABLE	**-0,0888*** **	**-0,0187** **	-0,0136	**-0,0605*** **	0,0018	0,0025
FRUIT	0,0174	-0,0014	**0,0275***	-0,0092	-0,0005	-0,0021
GRAZING	-0,0143	0,0009	**-0,0333** **	0,0102	0,0001	0,0022
FOREST	**0,0656*** **	-0,0031	**0,0782*** **	-0,0113	**-0,0016***	-0,0016
LIVESTOCK	**0,1190***	**-0,0364** **	**0,1012** **	0,0750*	-0,0037	**-0,0108***

YOUNG	0,2353	0,0406	-0,0255	0,1131	-0,0026	**-0,0297***
AGE_MORE65	-0,0164	0,0154	-0,0453	-0,0169	-0,0016	**-0,0152***
PARTIME	0,0117	**-0,0381****	-0,0112	0,0733	-0,0027	-0,0032
PREFASS	**3,4612*****	0,4070	0,2353	**1,8907*****	0,0150	-0,0840
PREFIDRO	-1,5096	-0,1493	-0,6430	-0,1669	**-0,0781***	0,0052
PREFNAT	-0,5706	-0,2032	**-1,9573*****	**1,7176****	0,0005	-0,0065
PREFPAE	1,2625	-0,0528	**1,7565****	-0,4359	0,0662	0,1304
PREFSUOLO	-0,6406	0,3600	-0,6497	-0,2694	0,0374	0,0070
R2	0,4256	0,3985	0,5070	0,3125	0,3293	0,3046
Log likelihood	-1184,7800	-726,9450	-1031,1100	-1062,2900	-89,6772	-417,9370
Rho	**0,2389*****	**0,5607*****	**0,2757*****	**0,3996*****	**0,3276*****	**0,5575*****

Results in term of significant variables are altogether consistent with the OLS model, with a few exceptions, notably Inhabitant density that lose significativity in the overall measure model, arable crop and livestock that loses significativity, young and age above 65 that become significative for sub-measure 10 (with negative sign). Preferential area variables maintain the same effect, except for PREFASS (greater part of the municipality is under Natura2000 or nitrate vulnerable areas) in the case of integrated production.

In the spatial error model (Table 3) the spatial variable (Lambda) is highly significant, in analogy to the Rho. The resulting R2 is very similar to the one of the previous model. On the contrary the significant variables change somehow more substantially. In particular, the constant becomes significant for integrated production (again one of the more uniformly diffused measures), but loses significativity in sub-measure 2.

Table 3: Results of the participation models, considering a spatial error model

variable	All214	214-sub-measure 1	214-sub-measure 2	214-sub-measure 8	214-sub-measure 9	214-sub-measure 10
CONSTANT	16,5613	**4,3760***	10,8363	5,2333	0,6387	0,6732
PIANURA	-6,8951	0,0898	-4,8672	-3,0182	0,0182	-0,0881
COLLINA	-4,7557	0,5566	-3,5826	-2,3294	-0,1453	-0,1785
MONTAGNA	-3,2362	-0,2251	-2,5001	-1,2522	-0,1637	-0,2347
DENS_AB	-0,0021	-0,0006	-0,0010	0,0003	**-0,0002*****	-0,0003
COND_DIR	-0,0261	0,0076	-0,0557	0,0471	-0,0040	0,0030
ONLY_HHLAB	0,0788	-0,0073	0,0492	0,0224	**0,0042***	0,0109
ARABLE	**-0,1341*****	**-0,0156***	-0,0274	**-0,1089*****	0,0020	0,0027
FRUIT	0,0037	0,0016	0,0244	-0,0294	-0,0006	-0,0030

GRAZING	-0,0340	-0,0039	**-0,0340****	-0,0181	-0,0001	0,0006
FOREST	**0,1003*****	**-0,0056****	**0,1055*****	0,0104	**-0,0022****	-0,0006
LIVESTOCK	**0,1444****	-0,0422	**0,1003****	**0,1234****	-0,0042	-0,0137*
YOUNG	0,1297	0,0338	-0,0254	-0,0707	-0,0012	-0,0275
AGE_MORE65	-0,0854	0,0112	-0,0410	**-0,1233****	-0,0019	-0,0193
PARTIME	0,0081	-0,0383	-0,0274	**0,0920****	-0,0029	-0,0039
PREFASS	**2,4189****	0,2443	0,0548	0,8757	0,0000	-0,0898
PREFIDRO	-0,7178	-0,0506	-0,4862	0,5896	**-0,0917***	-0,0669
PREFNAT	-0,5491	-0,1589	-1,7659**	0,9194	0,0029	-0,1110
PREFPAE	1,2325	-0,0272	**1,8242****	-0,4529	0,0706	0,1600
PREFSUOLO	-1,3801	0,2867	-0,9645	-0,6753	0,0422	-0,0074
LAMBDA	**0,3572*****	**0,5954*****	**0,3082*****	**0,5752*****	**0,2900*****	**0,5888*****
R2	0,4359	0,3936	0,5064	0,3548	0,3102	0,3108
Log likelihood	-1184,3116	-730,208	-1032,01	-1058,89	-93,5801	-418,104

Inhabitant density and household labour remain relevant only in sub-measure 9. Land use variables change slightly. Of the socio-economic variables age over 65 and part-time farmers remain the same with some significance, but they are now only relevant for sub-measure 8. Except for PREFPAE, the role of preferences shows some change in all cases.

4.2.5 Discussion

The main limitation of the work performed in this task relates to the data limitations in terms of lack of flexibility about the scale of analysis (the only feasible scale was on the municipality level). This has implications in terms of:

- Consistency with potential spill-over effects, that are mainly on the farm level;
- Effectiveness of the non-spatial explanatory variables, that in most cases are limited to a few information related to secondary data about crops, age and population in a municipality.

This on the one hand leaves space to the possibility that relevant spillover are not taken into account by the by the model, while, on the other hand, the spatial variable also incorporate spatial differentiation that is explained by other variables not available for the model.

In spite of this limitations the main lessons coming from the analysis are the following:

- It was possible to estimate models for measure 214, with a relevant ability to explain participation;

- Within this models the spatial component was highly significant and important;
- The explanatory variables are sharply differentiated by sub-measures;
- The regional priorities affect the results probably as a mixed effect of environmental characterisation and of priority in awarding of the funding (which could also be incorporated in expectations, hence affecting the decision to apply since the beginning);
- Socio-economic indicators appear as less often significant and less stable across models, hinting at a process strongly affected by "harder" structural and agricultural variables.

4.3 Diversification in non-farming activities (Measure 311)

4.3.1 General features and participation

Measure 311 "Diversification in non-farming activities" includes 3 sub-measure, respectively: 1) agritourism (restructuring of rural buildings to be used for agritourism); 2) touristic hospitality (restructuring of rural buildings to be used for touristic hospitality); 3) energy production from alternative sources. The measure provides a contribution on the capital cost of investment, with priority in areas having overall development problems and intermediate areas.

The measure was designed since the beginning to have a low target of participation (501 farms on the whole region), and the measure is target to only rural area. Such target was reached for about half already in 2010. The ratio between application and approve demand was almost 2. One could then expect that budget constraint and priority have had an effect on location of final participants. The large majority of participants have been concentrated on sub-measure 1. The less numerous on sub-measure 2.

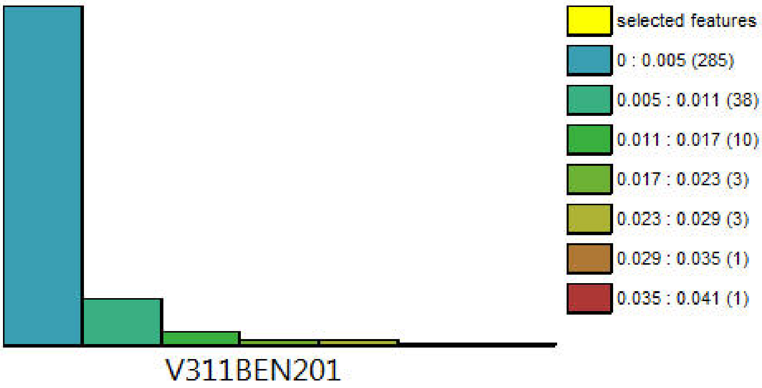
.

The distribution among municipalities was very skewed, with two thirds roughly of municipalities having no participants (Figure 17).

Eliminato: Figure 16

Figure 17: Distribution of participation across municipalities (share of farms in the municipality)

Eliminato: *16*



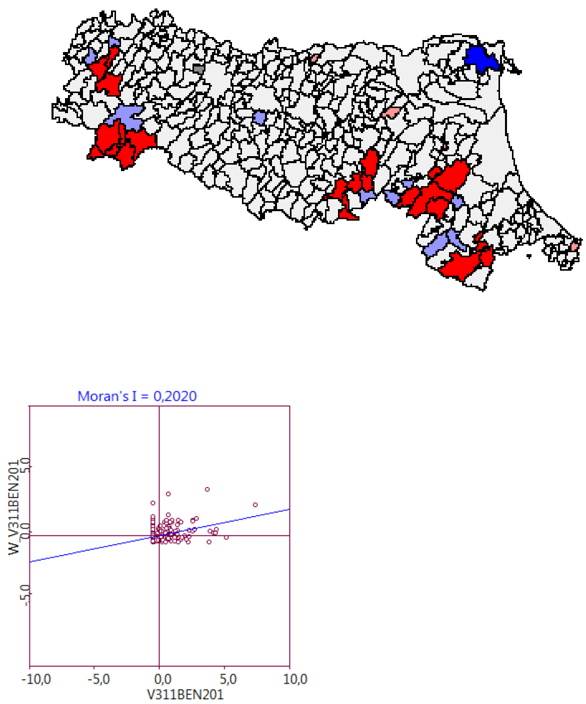
4.3.2 Spatial analysis

Figure 18 shows the results of the LISA analysis for measure 311 as a whole.

Eliminato: Figure 17

Figure 18 LISA Cluster Map and Moran I for measure 311 (all submeasures)

Eliminato: *17*

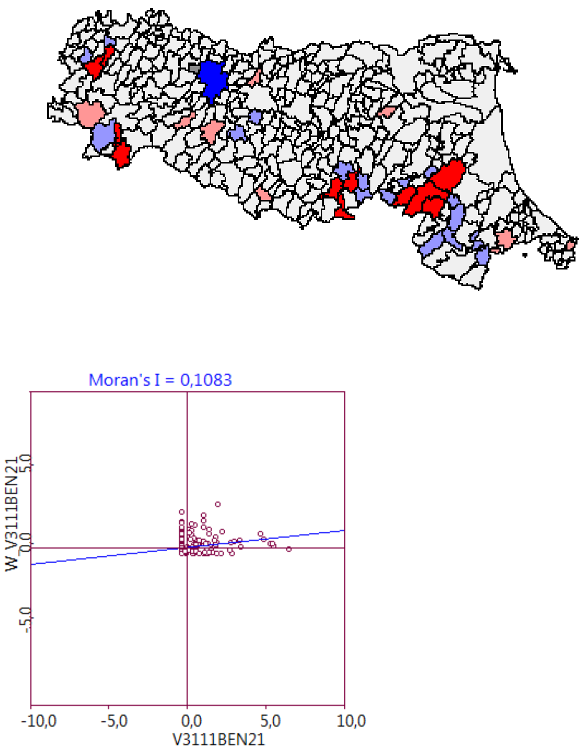


There are only a very few clusters in the region, while the Moran I index is relatively low. The same results happens if we consider the individual sub-measures. However, in addition to this, the location of the clusters changes sharply in sub-measure 1 (Figure 19), in which the low-low cluster in the plain moves from a very different area. On the contrary, the high-high cluster more or less reflect the aggregated ones, consistent with the fact that sub-measure 1 is the one with more participation, hence affecting the overall performance of the measure. Not surprising that higher participation is located at the border with the Tuscany region, located in the municipality crossed by the main street which connect the main city of Emilia Romagna and Tuscany.

Eliminato: Figure 18

Figure 19 LISA Cluster Map and Moran I for measure 311 (sub-measure 1)

Eliminato: *18*

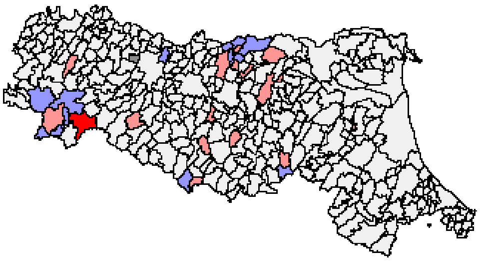


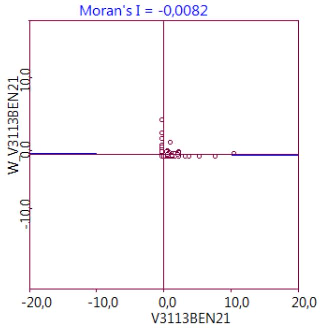
Sub-measure 3 shows hardly any concentration of similar municipalities, except for one municipality with high participation, close to other municipalities with high participation (Figure 20).

Eliminato: Figure 19

Figure 20 LISA Cluster Map and Moran I for measure 311 (sub-measure 3)

Eliminato: *19*





4.3.3 Modelling strategy

The modelling strategy was tentatively the same as measure 121 and 214.

The models considered are an ordinary least square regression, a spatial lag model and a spatial error model.

The models have been applied both to the whole of measure 311. This choice was driven by the already low number of participants, that would not allow for a meaningful treatment of disaggregated information. The models use Queen 1st order contiguity.

The model is built at municipality level (1 municipality=1 observation). The dependent variable is represented by participation and measured through the ratio between the number of participating farms and the total number of farms in the municipality.

The explanatory variables used are the same as in measure 214, except for the preferentiality variables, that are not included here.

4.3.4 Model results

The results of the models are given in Table 4

Table 4: Regression results

variable	OLS	Spatial lag	Spatial error
CONS	**0,0308***	**0,0298***	**0,0278*****
PIANURA	**-0,0232*****	**-0,0231*****	**-0,0226*****
COLLINA	**-0,0222*****	**-0,0223*****	**-0,0220*****
MONTAGNA	**-0,0231*****	**-0,0230*****	**-0,0225*****
DENS_AB	-0,000001	0,00000	0,00000
COND_DIR	-0,000028	-0,00003	-0,00003
ONLY_HHLAB	0,000053	0,00005	0,00004

ARABLE	**-0,000049*** **	**-0,00004** **	-0,00003
FRUIT	-0,000011	-0,00001	-0,00001
GRAZING	0,000002	0,00000	0,00000
FOREST	**0,000028***	0,00002	0,00003
LIVESTOCK	0,000001	0,00001	0,00003
YOUNG	0,000107	0,0001	0,00013
AGE_MORE65	-0,000073	-0,0001	-0,00005
PARTIME	-0,000043	0,0000	-0,00003
LAMBDA			**0,3094***
R2 adjusted	0,1707	0,2202	0,2188
Log likelihood	1332,0100	1339,4600	1338,7864
Rho		**0,2934***	

Due to the high number of zeros (67,7%), the regressions have very low explanatory capacity, with low adjusted R^2. This is only slightly higher in the spatial models (better in the spatial lag, but the difference is almost irrelevant), in which the spatial variables are also significant (though only at 10%).

The main significant variables are the same across models and are concentrated on the constant (always positive) and location variables (strangely all of which are negative).

In addition variables connected with density In addition arable land and forest are significant for the OLS, while arable land only is significant for the spatial lag model.

Altogether, also the significant variables provide little help in interpreting spatial location of this measure.

4.3.5 Discussion

The application of spatial to Measure 311 was not very satisfactory compared to other measure discussed above. In fact, it allows rather to illustrate the difficulty in spatial econometric analysis of measures with sparse and heterogeneous participation and several observations with participation close to zero.

The main message, here, besides the low overall model explanatory ability, is that also the results may be difficult to interpret and/or usually available space-related variables may be not a good support to understanding location of measures. Furthermore, using different spatial weights (e.g. distance to the main city of the region), could improve the analysis.

5 General discussion

The results of the case studies in Emilia Romagna show the relevance of spatial econometrics for the interpretation of the results of RDPs. Results support the expert expectation about spatial spill-over collected in the D5.1.

They also emphasises data limitations, that have effects, among others, on the ability to provide insightful interpretations of the outcome of the econometric models.

The results also emphasises difference in concept and problems across different measures, which remains one of the specific problems in overall evaluation of RDPs.

The specific case of Emilia Romagna, characterised by a strong targeting process in the measure design, also allows to highlight the importance of properly consider the policy design parameters in the econometric model, in order to attribute responsibility between factors affecting willingness to participate and policy factors affecting likelihood that the contract are awarded.

Also several issues remain unaddressed, in particular those related to the differential effects of the farm selection process in the cases in which the applications were higher than the budget vs. the case in which all eligible applicants where funded (eg. measure 121 vs 214).

With the exception of simulation of measure 121, we have applied only spatial weights based on the first order of contiguity. The spatial regime associated to such level of contiguity is quite obvious and reflect a neighbouring effects. Results shown that neighbouring effect are relevant in explaining participation. However further analysis wich are enabling to include higher order of contiguity could allow to consider other spatial regime associated with the participation.

Also improvement of the model could be realised trying alternative spatial weights. Even, if they contiguity weights have significance in all analysis, other spatial regime such as distance to the main market or distance to the main street could better explain the participation (see For example D5,1).

Finally the use of eligible vs. the total population as the reference population would further allow to refine the analysis.

6 Implications for further work

The work done leaves open possibility for further developments in at least two directions:

- Using area-based or payment-based proxies for uptake.
- Using impact variables as dependent variables;

The former seems more feasible based on available data, while impact variables can realistically be addressed only when a more systematic impact monitoring system will be in place. The Region is already developing a more advanced monitoring and evaluation system for the Farmland Bird Index and nitrogen balance.

Data availability remains a critical issue not only on the side of the dependent variable, but also of explanatory variables. In particular the lack of systematic individual information about non-participant seriously affects the possibility to econometrically estimate the determinants of participation. This study is based largely on agricultural census 2000 information, but potentially improved analyses are possible when the 2010 census information will be available.

The use of priority/preferential variables in the model allowed to test the treatment of planning/target variables. This seems a promising way of incorporating useful information for policy-makers, in the direction of supporting ex-ante analysis.

7 Annex 1 - Descriptive statistics for measure 121

The data used in the analysis concern the three calls of the measure 121 which cover the years 2008-2010. The data used for the explanatory variables are obtained from the ISTAT and Census 2000.

The applicants and the participants are obtained dividing respectively the cumulative number of farms (at 2011) who were applicants to measure 121 and the cumulative number of farms (at 2011) who got the payments in each municipality, with the total number of farms in each municipality.

In table 1 and the statistical descriptives of the above mentioned variables are presented.

Table 1. Statistical descriptives of the applicants rate and participation rate per Province.

		Applicants				Participations			
Province	Municipality (#)	Mean (%)	Std. Dev.	Min (%)	Max (%)	Mean (%)	Std. Dev.	Min (%)	Max (%)
BOLOGNA	60	2.83	2.18	0	8.47	1.72	1.55	0	5.93
FERRARA	26	5.03	2.84	0.57	10.54	3.59	2.16	0.23	7.22
FORLI CESENA	30	3.08	2.45	0	9.46	1.63	1.49	0	6.25
MODENA	47	2.50	1.44	0	6.17	1.43	1.12	0	4.66
PARMA	47	3.25	1.87	0	9.23	1.13	1.05	0	3.68
PIACENZA	48	5.08	4.93	0	16.67	4.22	4.61	0	16.67
RAVENNA	18	5.17	2.46	1.05	10.44	3.93	2.29	0.83	8.79
REGGIO EMILIA	45	3.66	2.81	0	14.75	1.65	1.32	0	4.92
RIMINI	20	1.24	1.76	0	7.94	0.78	1.29	0	5.61
TOTAL	341	3.49	2.96	0	16.67	2.14	2.48	0	16.67

The average applicants rate for the entire Region is 3.49%, with minimum value of zero (it means no applicants in the municipality) and maximum value about 17% of the farm of a municipality. The rate of applicants are strongly diversified across the Emilia Romagna provinces. Three provinces (Ravenna, Piacenza and Ferrara) have high rate participation, about 2 % more than the average value of the region. Three Provinces (Rimini, Bologna and Modena) have very low rate of applicants compared to the others. Such Provinces have less than 3% of applicants rate, almost half value of the three province with the highest value. Finally, the remaining three provinces (Reggio Emilia, Forlì Cesena and Parma) have the value closer to the average of the region. The high standard deviation value shows high differences between the observation within Provinces (municipality). In particular, a strong heterogeneity is observed in Piacenza province. In fact in this Province there are several municipalities whit no applicants and the municipalities with the maximum percentage of applicants 17%. With exception of the provinces Ferrara and Ravenna there are several municipalities in which there has been no applicants

Due to the high competitiveness of the measure, the priority sets implemented by each local administration has resulted relevant in the ranking of the applicants and in the access to the payments. The average value of the participation in the entire region is 2.14%. Compared with the applicants rate, there is a reduction of the 1.50% of the number of farm. Such a strong reduction is due to the selection mechanism implemented in each Province. The ranking is realised using the priority criterion set by each Province. The average value of participation rate in each province is included between 0.78% in the Rimini province and 4.22% in the Piacenza Province. As explained in the paragraph 2, the calls of the measure 121 has been very competitive and the ranking of eligible farms has been established using the regional and local priority mechanism. Within the province there is strong heterogeneity. The higher participation rate is in one municipality in the Piacenza Province with all applicant that have received the payments.

The variable of participation rate has been used as dependent variable in the econometric models and the variables presented in table 2 are used as explanatory variables.

Table 2. Statistical descriptives of the data used.

Category	Variable	Descriptions	Obs	Mean	Std. Dev.	Min	Max
Policy design	Dlfa	Less favourable area (dummy)	341	0.41	0.49	0	1.00
	prob_crops	Regional and province priority	341	16.27	6.05	0	32.23
Characteristics of the farmers	potsuccess	Percentage of farm with potential successor	341	97.90	5.09	9.94	100.00
	Young	Percentage of young farmer (less than 40 years old)	341	8.87	3.41	0.80	21.25
	age_more65	Percentage of farmers older than 65 years old	341	38.26	8.08	18.06	63.41
	Agredu	Percentage of farms which have owner with anagricultural education	341	6.51	6.23	0	37.35
	high_edu	Percentage of farms which have the owner with university degree	341	4.09	3.11	0	26.44
Legal status	part_colle	Percentage of cooperative	341	0.13	0.35	0	3.36
	cond_dir	Percentage of farms which are conducted directly by the farm	341	91.02	8.55	8.85	100.00
Structure of the Farm	labcon_sup	Percentage farms for which the owner work full-time	341	9.52	8.90	0	50.59
	Partime	Percentage of part-time farmers	341	58.62	13.95	24.61	95.13
	only_hhlab	Percentage of farms which use only household labour	341	82.09	12.13	47.73	100.00
	lab_more2	Percentage of farm which use more than 2 FTE works	341	13.10	7.69	0	43.14
	prop_mach	Percentage of farm which are owned of the machinery used	341	85.90	8.99	21.38	100.00
	tractor_less 100cv	Percentage of farmers with tractor with low power than 100cv	341	72.67	18.77	0.07	98.40
	tractor_more 100 cv	Percentage of farmers with tractor with high power than 100cv	341	44.88	21.12	0	98.36
	ave_farmsize	Average dimension of the farm (ha)	341	11.68	6.77	1.77	47.88
	sau_sat	Percentage of UAA on the TAA	341	0.75	0.19	0.14	0.97
	corpi_av	Average number of plots in each municipality (#)	341	4.81	6.99	1.27	59.94
Farming systems	Arable	Percentage of farm with arable crops	341	73.63	20.63	0	100.00
	Fruit	Percentage of farm with fruit crops	341	22.51	22.36	0	94.12
	Forest	Percentage of farm with forest	341	38.92	39.22	0	100.00
	Pigs	Percentage of farm with pigs reared	341	4.54	4.26	0	35.09
	livestock	Percentage of farm with livestock	341	14.51	11.71	0	67.19

The explanatory variables are classified in five categories (table 2). The first category concerns the characteristics of the policy design variables, which are mainly connected with the location in less favourable areas (which is priority for the measure 121) and the variable prob_crops. Such variable represents the sum of the percentage of the farms by the relative priority in the area. Such variable have theoretically value between 0 and 1. The lower value represents the situation with all farmers in one municipality having priority equal to zero, at the contrary value one represent the situation where all farmers of a municipality have the maximum priority score.

The second category of variables concerns farm characteristics. These variables refer to the ageing, farm successor and the farm education. These variables are measured as percentage of the farmers in each municipality with potential successor (potsuccess), younger than 40 (young), older than 65 years old (age_more65), with agricultural education of the owner (agredu) and with high level of education of the owner (edu_high).

The third category (legal status) is composed by only two variables: the percentage of the in each municipality farms associated in cooperative (part_colle) and the percentage of the farmers in each municipality who conduct directly the farm (cond_dir).

The fourth category refers to the farm structure in each municipality. The variables considered are the amount of household and external labour used on the farm, the farm size and the capital used.

Concerning the labour several variables about the percentage of part-time farming (part-time) the percentage of farms where the owner allocate all the working time to agricultural activities (labcon_sup), the percentage of farms that use only household labour (only_hhlab) and the percentage of farm who use more than two full time equivalent (lab_more). The variables used as proxy of farm size are: the average farm size of the farm in the municipality (ave_farmsize), the ratio between UAA and total agricultural area (uaa_taa) and the average number of plots in the municipality (corpi_av). The capital used are measured using three variables: the percentage of farm who are owner of the machinery (prop_mach) and the percentage of farms with tractor with power more and less than 100cv (tractor_more100 and tractor_less100).

The last category are the farming system of the municipality and are measured as a percentage of farms with arable crops (arable), with fruit (fruit), with forest (forest) and which are rearing pigs or cows (pigs or livestock).

8 Annex 2 - Descriptive statistics for measure 214

Table 5: Descriptive statistics of independent variables

	Mean	sd	min	max	n
PIANURA	0,50	0,501	0	1	341
COLLINA	0,22	0,413	0	1	341
MONTAGNA	0,28	0,450	0	1	341
DENS_AB	219,3804	318,2207	3,9012	2793,8023	341
COND_DIR	91,0233	8,5491	8,8453	100	341
ONLY_HHLAB	82,0856	12,1332	47,7272	100	341
ARABLE	73,6342	20,6313	0	100	341
FRUIT	22,5124	22,3573	0	94,1176	341
GRAZING	26,5271	30,1852	0	100	341

FOREST	38,9227	39,2160	0	100	341
LIVESTOCK	14,5066	11,7109	0	67,1936	341
YOUNG	8,84506	3,4307	0,8	21,2500	341
AGE_MORE65	38,2556	8,0776	18,0555	63,4146	341
PARTIME	58,6209	13,9496	24,6061	95,1282	341
PREFASS	0,31	0,464	0	1	341
PREFIDRO	0,31	0,464	0	1	341
PREFNAT	0,25	0,433	0	1	341
PREFPAE	0,18	0,381	0	1	341
PREFSUOLO	0,36	0,481	0	1	341

Table 6: Descriptive statistics of dependent variables (participation=uptake/total per municipality)

	Mean	sd	min	max	n
All 214	8,4675	10,2610	0	100	341
Sub-measure 1	1,0234	2,5388	0	23,2560	341
Sub-measure 2	4,8410	7,0432	0	46,6256	341
Sub-measure 8	2,3380	6,4744	0	100	341
Sub-measure 9	0,2415	0,3806	0	2,4483	341
Sub-measure 10	0,3319	0,9545	0	7,1720	341

9 Annex 3 - Descriptive statistics for measure 311

Commento [FB1]: Aggiungerei descrittive per la misure 311

See measure 214.