

Spatial Analysis of Rural Development Measures Contract No. 244944

Work Package No. 4

September 2011

D4.2.1

Exploratory Spatial Data Analysis: Why, How and What it shows us.

Martijn Smit, Eveline van Leeuwen (VU University Amsterdam), Sandra Uthes, Ingo Zasada (ZALF)

Document status

Public use	x
Confidential use	
Draft No.1	14 9 2011
Final	1 10 2011
Submitted for internal review	1 10 2011





Table of contents

1. Spatial Patterns	2
2. What is ESDA?	3
Why?	3
How?	3
3. Results	8
Economic Development of Primary Sector	
% of GVA in the primary sector (09b2007)	11
% of GVA in the primary sector (09b2007); at NUTS1 level	12
% of GVA in the primary sector (09b2007), at NUTS0 level	13
GVA in the primary sector (09a2007)	14
Unemployment	
% unemployed (in active population) (03a2009)	15
Life-Long Learning in Rural Areas	
% of 25-64 year olds participating in education and training (035a2009)	16
Development of Services Sector	
% of GVA in services (033a2007)	16
Tourism Infrastructure in Rural Areas	
Total number of bed places (031a2008)	17
Self-Employment Development	
% self employed (030b2009)	18
Self-employed persons (030a2009)	19
Employment Rate	
% employed (in total population, ages 15-64) (02a2007)	20
Economic Development of Non-Agricultural Sector	
% of secondary plus tertiary sectors in total GVA (029b2007)	21
GVA in secondary and tertiary sectors (029a2007)	22
Economic Development	
GDP per capita (in PPS; EU-27=100) (O1aaverage)	23
Farm Structure	
Total labour force on all farms (C472007)	24
% of holdings larger than 100 ESU (C46c2007)	25
% of holdings smaller than 100 ESU (C46b2007)	26
% of holdings smaller than 2 ESU (C46a2007)	27
Average economic farm size in ESU (C452007)	28



% of holdings larger than 50 ha (C44c2007)29
% of holdings smaller than 50 ha (C44b2007)30
% of holdings smaller than 5 ha (C44a2007)
Average farm size in ha (C432007)32
% utilized agricultural area (C422007)33
Number of farms (C412007)34
Agricultural Land Use
% permanent crops (C3c2007)35
% grass (C3b2007)36
% arable (C3a2007)37
Educational Attainment
% adults (ages 25-64) with medium or high educational attainment (C220a2009)
Long-Term Unemployment
% active population (C210a2009)
Structure of the Economy
% GVA in tertiary sector (C191c2007)40
% GVA in secondary sector (C191b2007)41
% GVA in primary sector (C191a2007)42
Population Density
inhabitants/km² (C17a2007)43
4. Conclusion
5. References



1. Spatial Patterns

Patterns in space abound, both because many things are organized in space by human design, and because of mutual influences. However, the human mind and eye may easily misjudge where patterns deviate from randomness, as "they are subjective, likely to tire, and so to err" (Ripley 2004). As ESRI writes, "In many ways, [spatial statistics] extend what the eyes and mind do intuitively to assess spatial patterns, trends, and relationships." (ESRI, *Introduction to Spatial Pattern Analysis*, online training seminar).

What may seem to be a pattern, and can be proven to be statistically significant, at one scale level, will surely look different at other scale levels. In the extreme case, all spatial relations disappear when we look at the European Union as a whole. At the other extreme, the number of relationships and connections between individual citizens and farmers, farms and households, animals and habitats is so large that it does most often not make sense to apply rigid exploratory spatial data analysis to it. For this report, we have chosen an intermediate scale, the NUTS2 level. This guarantees an acceptable availability of data, which is often not available at lower levels, while still giving us the opportunity to see potentially meaningful cross-regional clusters. Moreover, the NUTS2 level is in a number of European countries exactly the level at which Rural Development Programmes operate.

Two types of spatial patterns are generally of interest. The first is obvious: clustering, both of high and of low values, can indicate underlying factors for farms, firms or people to agglomerate. These agglomeration patterns are well known especially in urban economics, but are also studied in agricultural economics (Larue & Latruffe 2009). The second type of spatial patterns is that where high values and low values occur side-by-side. This can be the case for example where an urban region is surrounded by rural areas, and rural-to-urban migration leads to positive population growth in the city, but negative population growth around it.

Once spatial patterns have been discerned, we can and will ask ourselves whether these can be further analyzed with spatial econometrics, or indeed if these *should* be taken into account in any analysis. This will be taken up in Deliverable 4.3.



2. What is ESDA?

Why?

Exploratory Spatial Data Analysis is a first step to check whether spatial patterns exist, or, in other words, whether high and low values are suspiciously sorted in space. Moreover, it can show how strong these associations are, compared to other patterns. Finally, although we will not go down that road here, it can be used in a multivariate setting, to show how the spatial patterns of two variables interact.

We will perform ESDA on a large set of baseline and context indicators relevant to Rural Development. These indicators come from the report *Rural Development in the European Union* (European Union 2009). For each indicator, the report gives data for the latest year available – usually, 2007, 2008 or 2009. Not all indicators cover the complete European Union, and we have left out those which don't.

The indicator variables we chose could be used as dependents in analyses of Rural Development Programmes, as performed within the project. However, LISA can and should also be used to investigate independent variables to look for possible sources of bias. Finally, it can be extremely useful to perform a LISA analysis on the residuals from a regression, to see whether any spatial pattern has remained undetected by the variables already in the regression.

How?

The analysis is always done for one specific year – spatio-temporal panel analysis is still in its infancy, and although three-dimensional visualizations (two in space, one in time) do exist, formal analysis tools have yet to be developed. As a simple workaround, the analysis can be repeated for all available years, or for the first and last available year. However, if spatial patterns exist for one year, that is already enough to merit the inclusion of spatial econometrics in a model.

Apart from the subjects and the years, we also have to choose a weight matrix. This topic is covered extensively in Deliverable 4.1. In this case, we use k-10 nearest neighbours; the instructions below briefly repeat some of the considerations in this respect.



To perform ESDA, we will use the OpenGeoDa software. (Open)GeoDa is a free software package that conducts spatial data analysis, geovisualization, spatial autocorrelation and spatial modeling. With GeoDa comes a free workbook entitled *Exploring Spatial Data with GeoDa: A Workbook* (Anselin, 2005). Together with this, we use a GISCO shapefile as provided by Eurostat, from which we have cleaned all unnecessary statistical levels.

We will now give a step-by-step instruction how to perform ESDA.

- First, we have to prepare the data. Merge the .dbf-file that goes with the shapefile with the data of interest. It is recommended to use OpenOffice/LibreOffice Calc for that, but Access will do too. Excel won't work; recent versions (from Office 2007 onwards) no longer save as .dbf.
- Start OpenGeoDa. It opens just the toolbar (see image).



- Open the shapefile. Note that the projection can not be changed inside GeoDa. If the map looks ugly, load the shapefile in ArcMap, change the projection, and save the shapefile again with the new projection. Also, unwanted polygons can be changed there, for example to remove French Guyana or the Azores.
- Create weights (easiest from the toolbar, fourth icon, or alternatively through the Tools menu). A numeric unique ID for every polygon is needed. Usually, such an ID is already present in any shapefile that has been created with ArcMap, but GeoDa can also create one. Check the table by clicking on the third







icon on the toolbar.

- Now there are four types listed, but not all of these are useful (see Spard Deliverable 4.1). A weights file is a matrix which for every region A-region B combination tells GeoDa whether they are to be considered neighbours or not (0/1).
 - k-nearest neighbours is the safest bet. This assigns as neighbours to region A all its nearest neighbours, whether they actually share borders or not. The number of neighbours is the same for all regions; this is the number k you choose. Depending on the size and number of regions, settings vary; 10 is quite tractable in the NUTS2 setting. The main advantage is that islands have no problems; but a disadvantage is that distances between 'neighbours' can vary widely across the map (e.g. North Sweden vs. the Netherlands).
 - A distance cutoff is similar, but here all regions within a certain distance range are considered neighbours. Some regions that are far off (Cyprus, Azores, Iceland) may end up without neighbours, frustrating GeoDa.
 - Queen (and the similar rook) contiguity is the most basic: whoever touches your region is a neighbour. Disadvantage is that some models don't work with this type; especially, LISA will not work when there are regions without neighbours – Cyprus, Sicily, etc. etc.
- Now there are two types of analysis that are of interest: univariate and multivariate. Both are accessed through the LISA command; the Moran's I options are included there as well.

What windows to op	ien? 🛛 🔀
Significance Map	ОК
Box Plot	Close
Moran Scatter Plot	

• Export results from a map or graph by right-clicking and choosing Save Image. Here you can also select Zoom, which can be handy on a large map. Unfortunately, the results that come out are not high-quality vector images.

Analyses

The analyses performed in a standard ESDA are threefold. First of all, there is the regular plot of observed values. (This is not discussed above, but GeoDa provides for this as well, as do all other mapping tools, including the GIS viewer in MetaBase, ArcMap, and many websites of statistical offices.)



Secondly, there is the LISA cluster map. In a LISA analysis (Local Indicators of Spatial Autocorrelation; see Anselin, 2005:140), clustering of similar and contrasting values is analyzed through the use of the so-called Local Moran's I. This indicator takes on values between -1 and +1, where 0 stands for a random spatial pattern, and the two extremes indicate two types of spatial association: as the Local Moran's I approaches +1, a cluster of similar values (either high or low) is present, but as it goes down to -1, it indicates high and low values are 'suspiciously' mixed.

The third tool is a Global Moran value. This value indicates whether in the whole area spatial clustering or mixed patterns, as opposed to random patterns, exist. This value is the correlation between the values in each region and the (average) value of the surrounding region(s), and it is thus easiest to deduce from a scatter plot of those two.

With GeoDa, a LISA cluster map can produced, and in the following section this map will always be shown. GeoDa offers a separate significance map with the cluster map, but the cluster map itself also indicates which regions have insignificant values; the significance map would only be needed to see exactly at what level specific values on the cluster map are significant. GeoDa also offers the scatter plot mentioned above, which we will not normally reproduce below, except in isolated places as a demonstration.

Univariate vs. multivariate

The difference between univariate and multivariate is of course that in the univariate model, productivity in agriculture (as a random example) is related to productivity in agriculture in neighbouring regions. In the multivariate model, productivity in agriculture can still be your dependant, but the explanatory variable is





something else in the neighbouring regions. See for example the map to the right: the cluster map shows that especially around Paris the amount of investments in agriculture is high, and so is employment in the food sector. Only the region south of Ile de France still has the same high investments, but now in an area where employment in the food sector is actually quite low. The map shows you, simply put, where the extremes are, and the dark colours are clusters in the traditional sense, of high (red) or low (blue) values, whereas the lighter shades are the strange cases.

OpenGeoDa also performs linear regression, but it is quite limited. For our analyses (see upcoming deliverable 4.3), we use Stata with a spatial analysis package. This we also use for normal scatter plots and histograms, but OpenGeoDa in theory produces those as well.



3. Results

The following pages show the results of the ESDA. For each indicator, a descriptive map and a LISA cluster map is shown. The descriptive text comments on the Global Moran's I, where relevant. Note that regions outside the EU, as well as EU regions outside Europe proper, have been omitted.

Legend for LISA maps

The LISA maps share a common legend, which is as follows:



The option 'Neighborless' never occurs in the current analyses, as we chose a weight matrix of k-10 nearest neighbours.

Other plots

A significance map for indicator O9b2007 (the % of GVA in the primary sector) is given on the right. Darker green colours indicate more significant results, ranging from the 5% level (light green) via 1% (medium green) to 0,1% (dark green). Note that the grey regions form no





significant clusters, and these areas are grey also on the LISA map presented below on p. 10.

A Moran scatterplot of the same variable is given below. On the horizontal axis are standardized values for each region (i.e. their mean has been set to 0, and the horizontal scale is in standard deviations rather than units), and on the vertical axis the weighted values for their neighbours. One region at the bottom right has a very high value, whereas its neighbours on average have rather low values. If the difference is significant, this region will show up in a LISA map as a high-low value (pink).



Economic Development of Primary Sector % of GVA in the primary sector (O9b2007=C191a2007)

Quantile: 09B2007

[0,02603:0,7195] (44) [0,7209:1,398] (44) [1,422:2,116] (44) [2,15:3,358] (44) [3,377:5,155] (44) [5,211:12,3] (45)



.... P



Object indicator 9b – Economic development of the primary sector (share of total GVA) – shows especially high values at the periphery of the EU (Finland, Poland, Romania, Greece and Spain) and low ones in the serviceoriented economies of the UK or Germany and some capital regions. A Moran's I coefficient of 0.48 indicates some positive spatial clustering effects. Low-low relationships are found on the British Islands, Belgium and Germany, high-high relationships on the Eastern periphery.

% of GVA in the primary sector (O9b2007); at NUTS1 level







The same objective as the previous map, but now at NUTS1 level. The number of regions is now much smaller, and that is probably a first reason why no significant clusters are found in the LISA analysis (left). However, differences within countries are still visible, for example in Romania, where the share of agriculture is very high in the western region, and low in the east of the country, or within Germany. (Due to a number of zero values, a regular Quantile map could not be produced by GeoDa, which is not so good with handling so-called ties. Therefore, we show a Box Map instead, where values have been standardized.

% of GVA in the primary sector (O9b2007), at NUTS0 level





Once more the same objective as previously, but now at NUTSO (country level). All spatial effects have disappeared, and the map above just plots country averages.

GVA in the primary sector (O9a2007)

Quantile: O9A2007 [3,5:216,6] (44) [220,2:362,4] (44) [363:491,5] (44) [496,1:710,3] (44) [713,6:1153] (44) [1168:5790] (45)



.000



Object indicator 9a – Economic development of the primary sector (GVA in the primary sector, i.e. in absolute figures) – shows high values in France, Spain and Italy, as well as parts of Finland and Romania. Southern French and a string of Iberian regions form a cluster of high values. Spatial clustering of low values in neighbouring regions exist in the UK and some regions in Sweden, Eastern Germany, Belgium, the Czech Republic and Slovakia. However, the Moran's I is low at 0.17, indicating a low level of spatial autocorrelation.

Unemployment % unemployed (in active population) (O3a2009)

Quantile: O3A2009 [2,108:5,155] (44) [5,158:6,567] (44) [6,601:7,58] (44) [7,692:9,39] (44) [9,403:11,34] (44) [11,59:26,19] (45)



1.4.1



The object indicator 3a – Unemployment – shows particular high values in Spain, Eastern Germany, Ireland and the Baltic countries. Hardly surprising, in the Netherlands, Southern Germany, Austria and Northern Italy low values are found. There the low-low neighbourhood effects are most prevailing. The spatial distribution appears rather heterogeneous (Moran's I coefficient of 0.49), with high-high clusters in Spain and Portugal, but not in the Baltic countries. note the low 'island' value for the Lisbon area in a 'sea' of high values.

Life-Long Learning in Rural Areas % of 25-64 year olds participating in education and training (O35a2009)

Quantile: O35A2009 [0,4184:3,988] (44) [4,055:6,09] (44) [6,1:7,387] (44) [7,473:10,45] (44) [10,54:18,58] (44) [18,68:35,14] (45)



: . . 1



.... B

Object indicator 35a – life-long learning – is characterised by very strong regional variation. High values are found in Scandinavian countries, the Netherlands and the UK, whereas low values prevail in Eastern and South-Eastern Europe. Note the absence of a significant cluster in Austria/Slovenia, which stand out on the map above, but disappear on the cluster map on the left. With a Moran's I coefficient of 0.73, the spatial pattern of the variable values is characterised by high spatial autocorrelation.

Tourism Infrastructure in Rural Areas Total number of bed places (O31a2008)

Quantile: 031A2008

[848:2,685e+004] (44) [2,727e+004:4,448e+1 [4,521e+004:6,561e+1 [6,583e+004:9,567e+1 [9,616e+004:1,54e+01 [1,545e+005:7,066e+1



: . . 1



°.00 \$

Object indicator 31a – Tourism infrastructure in rural areas – seems to reflect the presence of natural and cultural amenities with high values in France, Italy, Austria and coastal regions. The Eastern periphery of Europe, from Finland down to Greece, is characterized by low values. With a Moran's I coefficient of 0.35, there is some spatial autocorrelation: significant clusters of high values persist in Southern France and Northern Italy. The low amount of bed places in Greece is puzzling, but local spikes of high values occur in Crete and other Greek islands, as well as in Rumania along the Black Sea.

Self-Employment Development % self employed (O30b2009)





.....



.... B

Object indicator 30b – Self-employment development (share of the total employment) – is characterised by a highly clustered spatial distribution (Moran's I coefficient amounts 0.62): low clusters in Swedish, Danish, German and French regions. High values are found in Southern and Eastern Europe, but not all areas with high values form part of significant clusters (e.g. Galicia).

Self-employed persons (O30a2009)





:....

Object indicator 30a – Self-employment development (self-employed persons, in thousands) – shows a very low autocorrelation (Moran's I coefficient of 0.14). High and low values are spread across the EU with some concentrations of high ones in Spain, Italy, Romania and Poland. Lower values persist in the North. Clusters are mainly found in Sweden, Finland and the German-French and Czech-Austrian borders (low-low), as well as in Italy (high-high).

Employment Rate % employed (in total population, ages 15-64) (O2a2007)

Quantile: O2A2007 [44,17:59,72] (44) [59,95:65,1] (44) [65,11:67,75] (44) [67,78:71,3] (44) [71,31:75,06] (44) [75,22:82,75] (45)



.000



Object indicator 2a – the Employment rate – shows high values in Scandinavia, the UK, the Netherlands, Southern Germany and Austria, low ones in Eastern and Southern Europe. Accordingly significant clusters exist in these areas (Moran's I coefficient of 0.56). Low-low groupings of regions are found in Poland, Hungary, Romania, Bulgaria, Greece and Italy. Note local Greek, Bulgarian and Cypriot outliers in the Eastern cluster of low values.

Economic Development of Non-Agricultural Sector % of secondary plus tertiary sectors in total GVA (O29b2007)

Quantile: O29B2007 [87,7:94,7] (44) [94,79:96,54] (44) [96,62:97,82] (44) [97,85:98,58] (44) [98,58:99,26] (44) [99,28:99,97] (45)



.....



Object indicator 29b – Economic development of non-agricultural sector (Share of total GVA) – reflects the relative importance of the tertiary and secondary sector. As such, it is the reverse of indicator 9b, covered above. Whereas metropolitan regions, the UK, the Benelux countries and Western Germany show high values, remote rural regions around Europe show rather low ones. Pronounced spatial clustering of high values exists in almost all German, Danish, British and Belgian regions. Again, Eastern Europe as well as Greece stands out as a cluster of low values. The Moran' I coefficient amounts to 0.48.

GVA in secondary and tertiary sectors (O29a2007)





. . . 1



Object indicator 29a – Economic development of non-agricultural sector (GVA in secondary and tertiary sectors, i.e. in absolute figures) – shows high values in metropolitan and "Blue banana" regions as well as in around the Western Mediterranean. Low values and strong neighbourhood effects are found in Eastern and South-eastern Europe. The regions around London, Paris and the areas along the French and Italian Riviera are influenced by high neighbouring values. The spatial clustering is rather weak, with a Moran's I coefficient of 0.10.

Economic Development GDP per capita (in PPS; EU-27=100) (O1aaverage)

Quantile: O1AAVERAGE [24,87:63,27] (44) [64,63:83,12] (44) [83,57:97,08] (44) [97,33:107,2] (44) [107,9:122,4] (44) [123,5:336,4] (45)



.... P



Object indicator 1a – Economic development (GDP in pps per capita) reveals a rather scattered spatial distribution pattern across Europe with weak neighbourhood spill-over effects (Moran's I coefficient of 0.17). High values are found in metropolitan region hinterlands and rural areas in Sweden, Poland and Spain. High-high clusters exist between Southern Sweden and Northern Germany, in the West of England and Wales as well as in the South of the Netherlands and adjoining Flemish regions. Note hoewever how the Swedo-Danish-German cluster has a number of low outliers. Low-low areas are found in Portugal, Bulgaria and parts of Poland.

Farm Structure Total labour force on all farms (C472007)

Quantile: C472007

[0:6260] (44) [6360:1,14e+004] (44) [1,148e+004:1,869e+([1,935e+004:3,567e+([3,658e+004:7,343e+([7,344e+004:4,623e+(



2.00 1



.... B

Context indicator 47 – Farm structure (Labour force on all farms combined, in absolute figures) – reveals the highest values in Ireland, the Iberian and Baltic countries, Poland, Romania and Bulgaria. Low values are found particularly in the service and production oriented regions of Sweden, the UK, the Benelux countries and Germany. The Moran's I coefficient of 0.48 indicates an average level of spatial autocorrelation. High-high values are found at the Eastern periphery of the EU, low-low ones around the North Sea.





.00 0



1.4.1

Context indicator 46c – Farm structure (Economic farm size, share of holdings with 100 ESU or more) – reveals a strong spatial autocorrelation pattern across Europe (Moran's I coefficient of 0.74). High values are clustered in one supranational region from Denmark and Mecklenburg-Vorpommern all the way to Southwest France. A large number of regions with low value neighbourhood effects is found on the fringes, in Eastern Europe, Northern Scandinavia, Spain, Italy, Austria, Slovenia and Greece.

% of holdings with 2-100 ESU (C46b2007)





1.00 \$



.... B

Context indicator 46b – Farm structure (Economic farm size, share of holdings between 2 and 100 ESU) – shows a pattern similar to that of very large farms (>100 ESU) given above. However, the cluster of high values now runs from Denmark via Southern Germany and Eastern France to Eastern Spain, instead of via the Netherlands. Again, low values dominate in Eastern Europe, but not in Finland, which has very high values, and Greece, which has moderately high values. The Moran's I coefficient is again rather high, at 0.67.

% of holdings smaller than 2 ESU (C46a2007)





.00 \$



Context indicator 46a – Farm structure (Economic farm size, share of holdings with less than 2 ESU) – shows where very small farms occur. The map above shows that small farms abound in Eastern Europe as well as Sweden and Southern England plus Wales. This variable has strong spatial autocorelation, with a Moran's I coefficient of 0.75. Note that the high values in Sweden do not form a significant cluster, whereas the other two groups of high values do; low values (i.e. hardly any small farms) persist in a broad zone from Malmö to Barcelona.

Average economic farm size in ESU (C452007)





:. . . . A



.....

Context indicator 45 – Farm structure (Average Economic farm size in ESU) reveals the highest values in Eastern Germany and the Netherlands. Particularly low values are found in Austria, Southern Germany, Italy, Sweden and the Czech Republic. The spatial distribution is characterised by a strong positive autocorrelation between neighbouring regions (Moran's I coefficient of 0.65) with low-low values at the European periphery and high-high clusters in regions between Eastern Germany via Denmark and the Netherlands to Northern France. The map above shows that values diminish gradually from Central France towards Northern Spain, but the cluster map on the left indicates a significant break in Southern France, where two regions have significantly lower values than their neighbours.

% of holdings larger than 50 ha (C44c2007)





· · · · · /



Context indicator 44c – Farm structure (Physical farm size, share of holdings with 50 ha UAA and more) – shows high values and clusters in Northern Germany, Sweden and Denmark, Northern France and Scotland, all of which are areas of either extensive or agro-industrial production. Low values are extensively clustered in the Danube area, Italy, Greece, Lithuania, Poland and Slovakia. The Moran's I coëfficiënt is 0.60.

29

% of holdings between 5 and 50 ha (C44b2007)





°.00 \$



Context indicator 44b – Farm structure (Physical farm size, share of holdings with more than 5 ha but less than 50 ha UAA) - reveals the domination of high values in a broad area from the Netherlands down to Slovenia plus Scandinavia and Ireland. However, the Irish cluster is not significant; since we use a neighbour matrix with the 10 nearest neighbours, Ireland is compared to neigbours in England, Scotland and Wales, and values are low there. A lowlow regional cluster runs from southern Poland down to Greece. The Moran's I coefficient (0.63) indicates a general high level of spatial autocorrelation.

% of holdings smaller than 5 ha (C44a2007)





.....



Context indicator 44a – Farm structure (Physical farm size, share of holdings with less than 5 ha UAA) – is characterised by a strong autocorrelation of regional values, with a Moran's I of 0.72. Large shares of small holdings and high-high clustering are found in Eastern and Southern European countries. Low values and significant low-low clusters occur in Scandinavia and large parts of central Western European countries. Note the touching high-high and low-low clusters around Vienna.

1.001

Average farm size in ha (C432007)





°.00 \$



Context indicator 43 – Farm structure (Average physical farm size) – reveals strong differences within the EU. High values prevail in Denmark, Eastern Germany, Czech Republic, Scotland and Northern France, low values in Southeastern Europe. Accordingly spatial autocorrelation occurs in particular in those regions. Low-low clusters exist in Italy, Lithuania, Romania, Bulgaria, Greece as well as parts of Poland, Slovakia and Hungary. High-high clusters are found in Eastern Germany, Scotland and central regions in France. The Moran's I coefficient is 0.47.

% utilized agricultural area (C422007)



°.00 0



Context indicator 42 – Farm structure (Utilized Agricultural Area) – contrasts the European urbanisation pattern. Low shares are found in England, the Benelux countries, Western and Southern Germany, as well as in many metropolitan regions. Therefore spatial autocorrelation is only weakly pronounced (Moran's I coefficient of 0.27). Low-low effects dominate in Central England and the Benelux countries as well as Greece. High-high effects exist in France, Spain, Latvia and Lithuania.

.....

Quantile: C422007

[0:1,411e+005] (44) [1,436e+005:2,662e+C [2,67e+005:4,35e+005 [4,354e+005:6,56e+005 [6,609e+005:1,031e+C [1,04e+006:5,471e+005]





2.00 B

Context indicator 41 – Farm structure (Number of farms) – is characterised by a medium level of spatial autocorrelation in Europe (Moran's I coefficient of 0.48). The highest values are found along the Eastern fringe of the EU and in the Mediterranean. Low values are present from Eastern Germany up to the UK. Therefore in regions of a broad belt around the North Sea low-low neighbourhood effects prevail. High-high relationships merely exist in Lithuania, Romania, Bulgaria and parts of Poland.

Agricultural Land Use % permanent crops (C3c2007)

Box Map (Hinge=1.5) : C3C2 Lower outlier (0) < 25% (47) 25% - 50% (83) 50% - 75% (68) > 75% (30) Upper outlier (37)



.....



Context indicator 3c – Agricultural land use (Share of permanent crops area) – shows a clear North-South-gradient, with very low values in Scotland and Scandinavia and high ones particularly in the Mediterranean basin. Spatial clustering of similar low values can be found on the British islands, around the Baltic Sea and in parts of France. Southern Spain and Italy as well as the whole of Greece feature strong highhigh neighbourhood effects. The Moran's I coefficient amounts to 0.55, indicating a moderate level of spatial autocorrelation.





:000



°. 0 0 \$

Context indicator 3b – Agricultural land use (Share of permanent grass area) – shows a rather scattered value distribution despite a few regional concentrations of high values on the British islands, in the Alps and in parts of Spain, as well as low values in Scandinavia and parts of the Balkans. Local clusters are found in Ireland, UK and the Alpine regions (high-high) and in Scandinavia, Western Poland, Slovakia, Hungary and Eastern Romania (low-low). However, the low-low regions have a number of high 'islands' (in pink) around them. The Moran's I coefficient is 0.38.





:. vo \$



Context indicator 3a – Agricultural land use (Share of arable area) – reveals clusters of high values in Northern and Eastern Europe and low values in the Alpine region, the British islands and Western parts of the Iberian peninsula. Accordingly, high-high clusters exist in a large area around the Baltic Sea, with minor groupings in Rumania and North-Western France. Low-low clusters exist around the Irish Sea, on the fringe of the Iberian peninsula (with higher values in the arid regions of central Spain) and around Northern Italy. The Moran's I coefficient amounts to 0.42.

Educational Attainment

% adults (ages 25-64) with medium or high educational attainment (C220a2009)





°.00 B



Context indicator 22a – Education attainment – shows the highest values in Central and Eastern European countries. Low values occur in regions all along the Mediterranean basin. With a Moran's I coefficient of 0.77, spatial clustering across Europe is particularly high. Denmark shows an interesting local exception, with some low values in an area where otherwise high values prevail.

2008

Labour participation % active population (C210a2009)

Quantile: C210A2009

[0,3526:1,103] (44)
[1,129:1,712] (44)
[1,718:2,436] (44)
[2,459:3,177] (44)
[3,179:4,287] (44)
[4,348:12,09] (45)



: . . \$



Context indicator 21a – Labour participation – reveals a spatial distribution pattern characterised by the existence of smaller regional clusters across Europe. For instance, in Southern Spain and Portugal, the Mezzogiorno, Eastern Germany, as well as in parts of Greece concentrations of high values are found. Low values again prevail in the North of the Netherlands, the South of England, almost all of Sweden and Finland, and a large region from Milan to Bratislava. The general degree of autocorrelation is rather low (Moran's I coefficient of 0.40).

Structure of the Economy % GVA in tertiary sector (C191c2007=O33a2007)

Quantile: C191C2007

[45,42:59,32] (44)
[59,65:64,83] (44)
[65,23:68,8] (44)
[68,84:73,13] (44)
[73,16:77,04] (44)
[77,23:92,99] (45)



:



.000

Context indicator 19c – Structure of the economy (GVA in the tertiary sector) – reveals high values in the metropolitan centres and tourist-oriented regions in the Mediterranean. Low values are found in the rural areas of Eastern Europe, Germany and Spain. Spatial clustering of regions with low values takes place in Eastern Europe. High-high clusters exist in Southern England, parts of France and Italy, Belgium and Denmark. The Moran's I coefficient amounts to a below average value of 0.34.

% GVA in secondary sector (C191b2007)



°.00 \$



14.1

Quantile: C191B2007

[6,983:20,48] (44) [20,49:24,93] (44) [24,93:28,29] (44) [28,42:31,75] (44) [31,78:36,47] (44) [36,49:50,33] (45)

> Context indicator 19b – Structure of the economy (GVA in the secondary sector) – reveals a dominance of high values in many regions of Central and Eastern European countries (CEC) and Northern Scandinavia. No clear "country" patterns are found as both low and high values co-exist in countries like Italy, Germany, UK or the Netherlands. The Moran's I coefficient of 0.34 indicates a rather low spatial autocorrelation, although significant clusters can be found in Central and Eastern European countries (high-high) as well as France and England (low-low).

Population Density inhabitants/km² (C17a2007)







....



.... B

Context indicator 17a – Population density - traces the pattern of European urbanisation. Hardly surprising, the metropolitan regions, as well as the regions along the "Blue Banana" between the British Midlands and Lombardy show highest values. Low population density is found especially at the European periphery. Only few regions in the Northern and Baltic countries, some Southern French regions (low-low) and regions around London and Rhine-Meuse delta (highhigh) show significant clusters. The Moran's I is extremely low, at 0.05; the scatter plot shows that although there are regions with higher population density, their neigbours are always close to the average.



4. Conclusion

This report gave an overview of the possibilities of Exploratory Spatial Data Analysis. These are not gratuitous possibilities – like any proper exploratory analysis, they should be taken advantage of whenever an analysis with possible spatial effects is undertaken, even if only to show spatial autocorrelations will *not* be a problem. For all variables shown in this report, there *was* an amount of spatial autocorrelation. Even if at the global level (the global Moran's I) the amount of autocorrelation was low, some individual regions still stood out as positive or negative clusters.

We looked at a wide range of indicators of which some were related to general characteristics of the European regions under research, such as life-long learning, or unemployment, and others more to agricultural focused issues, such as farm size and GVA in the primary sector. From all this, there are three important lessons we can learn:

- 1. It is important to carefully select the way an indicator is taken into account. GVA in the primary sector measured in absolute terms shows different levels of clustering than when measured in relative terms (0.17 against 0.48). This also holds for the number and share of self-employed persons (0.14 against 0.62).
- 2. Likewise, the spatial scale matters enormously. Looking at the share of agriculture in total value added, we found a moderate amount of spatial autocorrelation at the conventional NUTS2 level (a Moran's I of 0.48), but all spatial clusters disappeared when we looked at NUTS1 and NUTS0 regions. Unfortunately, analyses at lower levels are not feasible at a pan-European scale, but within Spard these will be undertaken in the case studies.
- 3. In general, the indicators related to agriculture show relatively strong spatial clustering effects. In particular when looking at the economic and physical size of farms, high values for Moran's I are found. Outside agriculture, strong spatial clustering effects are found for life-long learning and the level of higher educated adults (0.73 and 0.77 respectively).



5. References

- Anselin, L. (2005), *Exploring Spatial Data with GeoDa: A Workbook*, available at http://geodacenter.asu.edu/system/files/geodaworkbook.pdf
- European Union (2009), *Rural Development in the European Union: Statistical and Economic Information*, available at http://ec.europa.eu/agriculture/agrista/rurdev2009/RD_Report_2009.pdf

Larue, S. & Latruffe, L. (2009), *Agglomeration externalities and technical efficiency in French pig production*, Working Paper SMART – LERECO no. 09/10.

Ripley, B.D. (2004), Spatial Statistics, Hoboken: Wiley & Sons

All maps are $\ensuremath{\mathbb C}$ EuroGeographics for the administrative boundaries.