Topic report No 6

Towards agri-environmental indicators

Integrating statistical and administrative data with land cover information

Eurostat
DG Agriculture
DG Environment
Joint Research Center
European Environment Agency

Joint publication

European Environment Agency

Layout: Brandenborg a/s

Note
The contents of this publication do not necessarily reflect the official opinions of the European Commission or other European Communities institutions. Neither the European Environment Agency nor any person or company acting on the behalf of the Agency is responsible for the use that may be made of the information contained in this report.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2001

ISBN

European Environment Agency
Kongens Nytorv 6
DK-1050 Copenhagen K
Denmark
Tel: (45) 33 36 71 00
Fax: (45) 33 36 71 99
E-mail: eea@eea.eu.int
Internet: http://www.eea.eu.int

Contents

Introduction:
The territorial dimension of the European agriculture **6**
Abstracts **9**
1. Geographical use of statistical data **11**
1.1. Data collection and retrieval units **11**
1.1.1. Collection units 11
1.1.2. Geographical retrieval/aggregation/reporting units 12
1.2. Spatial units for processing **13**
1.2.1. Normative regions 13
1.2.2. Analytical units: catchment areas 17
1.3. Methods & definitions **17**
1.3.1. Spatial disaggregation 17
1.3.2. Spatial aggregation 18
1.3.3. Spatial intersection 18
1.3.4. Processes for localising data more precisely in space 19
1.3.5. Processes for transferring data between different geographical zonings 20
1.4. Conclusion **22**
2. Dealing with the modifiable areal unit problem **25**
2.1. Background **25**
2.2. The limitations of the NUTS nomenclature **26**
2.3. The modifiable areal unit problem (MAUP) **29**
2.4. Methods **32**
2.5. Conclusion **32**
3. Comparison of CORINE land cover data with IACS data in Belgium and Italy and with land use in Slovenia **39**
3.1. Introduction **39**
3.2. Building Land Cover Diversity Index: Data Issue **39**
3.2.1. Explanation of the Belgian IACS treatments of data 39
3.2.2. Share of land covered by IACS 40
3.2.3. Computing the Agricultural Landuse Diversity Index using IACS information 41
3.2.4. Shannon diversity indicator in Belgium from IACS data 43
3.2.5. Comparison of IACS and CORINE land cover diversity indexes 45
3.3. Building Land Cover Diversity Index: Classes Issue **46**
3.3.1. Treatment of Italian IACS data 46
3.3.2. Localisation of the IACS parcels and breakdown of IACS & CORINE . 47
3.3.3. Agricultural areas in IACS and CORINE 48
3.4. Land Cover Diversity: The Structure Issue **54**
3.4.1. Methodology 54
3.4.2. Results 54
3.5. Conclusions **56**

4. Land cover in the context of the Natura 2000 network . . . 60
4.1. General background of the Natura 2000 program . . . 60
4.2. Use of the NATURA2000 geo-layer within a GIS environment . . . 60
4.2.1. Overview of 'use-cases' of the Natura 2000 data within DG ENV . . . 61
4.2.2. Some examples of complaints . . . 61
4.2.3. Other users . . . 67
4.3. Natura 2000 sites and CORINE land cover data, the case of Greece . . . 67
4.3.1. Land cover classes in the Natura 2000 sites of Greece . . . 67
4.3.2. Recorded land use in the Natura 2000 sites against CORINE land cover data . . . 70
4.4. Conclusions . . . 73
5. Reassignment of the Farm Structure Survey's data . . . 75
5.1. Introduction . . . 75
5.2. Characteristics of the Farm Structure Surveys . . . 76
5.2.1. Scope of the survey and contents . . . 76
5.2.2. Date of the surveys and Geographical coverage . . . 78
5.2.3. Spatial observation unit . . . 78
5.3. Relocalisation of the structure survey's data . . . 78
5.3.1. Objective . . . 78
5.3.2. Prospects . . . 79
5.4. Comparison of nomenclatures . . . 79
5.4.1. Data used . . . 79
5.4.2. Highest aggregated level of the nomenclatures . . . 79
5.4.3. Linking nomenclatures at a finer level . . . 81
5.4.4. Proposals for improvement . . . 85
5.5. Possible work . . . 85
5.5.1. Integration of time factor . . . 91
5.5.2. Integration of spatial units . . . 91
5.5.3. Cases where CORINE land cover underestimates AA . . . 91
5.5.4. Cases where AA is lower than agricultural areas in CORINE land cover . . . 91
6. Using CORINE land cover to map population density. . . . 94
6.1. Introduction . . . 94
6.2. Available Data. . . . 94
6.2.1. Communes without urban area in CORINE land cover. . . . 94
6.3. Modified areal weighting with given coefficients.. . . . 95
6.4. Search of weighting coefficients. . . . 96
6.4.1. Disaggregation of regional data to assess the validity of weighting coefficients.. . . . 96
6.4.2. Application with 16 grouped classes. . . . 97
6.4.3. Stratification and further grouping of CORINE land cover classes. . . . 98
6.5. Suggested disaggregation. . . . 100
6.5.1. Further coefficient tuning. . . . 100
6.6. Quality assessment of the disaggregation in Arezzo (Italy). . . . 101
6.6.1. Disagreement at pixel level and at census section level. . . . 101
6.7. Conclusions . . . 104

7. **Agricultural statistics spatialisation by means of CORINE land cover to model nutrient surpluses** **106**
7.1. Introduction **106**
7.2. Methodology **106**
7.2.1. Definition of nutrient surplus 106
7.2.2. The surplus model used 106
7.2.3. The core role of CORINE land cover 107
7.2.4. Data requirements 108
7.2.5. Modelling scenarios 108
7.2.6. Spatial levels of calculation and reporting 108
7.3. Results in terms of nitrogen surplus **109**
7.3.1. Main characteristics of source data 109
7.3.2. Results on the Loire-Bretagne Water Agency district 110
7.3.3. Results on the Elbe basin 112
7.3.4. Relations between the surpluses and the type of occupation of the land 113
7.4. discussion of Main findings **114**
7.4.1. The positive input of CORINE *land cover* 114
7.4.2. Sources of uncertainty related to input data and constants 116
7.5. Conclusions **116**
8. **Comparing CORINE land cover with a more detailed database in Arezzo (Italy)** **120**
8.1. Introduction **120**
8.1.1. The Arezzo ILC database and CORINE land cover 120
8.2. Comparing total area per class **121**
8.3. Statistical analysis of the differences between the two maps **122**
8.3.1. Pixelwise (naïf) commission and omission disagreement. 123
8.3.2. Recoding into 9 classes. 124
8.3.3. Scale-corrected thematic disagreement by CORINE land cover polygon. 125
9. **Conclusions** **128**

Introduction: The territorial dimension of the European agriculture

Landscape definitions

There is no fixed definition of landscape. For some people it is a purely visual concept, comprising the geomorphological features such as fields, forests, hills and rivers that characterise the earth's surface. For others, landscape represents the complex interaction between human societies and the natural environment.

To complicate matters, the word has different connotations in different languages. Piecing the various concepts together, we can conclude that landscape is not only about what we see but also about the processes that have created what we see. It is not simply about the naturally created beauty of the Vosges or the Odenwald, but also about the processes of human settlement, agriculture and other activities that have accentuated or been detrimental to that beauty.

Public concern about the landscape is on the increase. While they may find it hard to understand notions such as natural resource protection or biodiversity, people feel very strongly about landscape issues. This may be at local level – which is understandable because it touches them directly – or at national or transnational level, where landscape is sometimes an important identifying element (the Alpine arc, the Atlantic front).

To summarise, landscape is linked to nature and to culture. It is about natural features, about farmland and forests and patterns of human settlement. It is about what is produced when nature and man interact.

Taking the environment into account at a political level

Confirmed in the European Treaties, the political direction is clear: community policies must integrate environmental concerns.

The Cardiff European Council in June 1998 invited the Council to establish its own strategies for "giving effect to environmental integration and sustainable development" in various policy areas. The Agriculture Council was one of those invited to start this process.

The Vienna European Council in December 1998 reaffirmed this commitment and moreover requested the European Commission to provide a coordinated report on indicators. The Agriculture Council was invited to continue the work it had started, with a view to submitting a comprehensive strategy, including a timetable for further measures and a set of indicators, to the Helsinki European Council. In July 1999 the Agriculture Council asked the Commission to prepare a report on agri-environmental indicators.

In its communication Directions towards sustainable agriculture'[1], the Commission underlined that the Agenda 2000 reforms provided a new impetus for the integration of environmental concerns into agricultural policy. The Commission, Member States, local authorities and agricultural and rural communities now had a considerable range of policy instruments at their disposal to encourage a move towards more sustainable farming practices.

The Communication Indicators for the integration of environmental concerns into the CAP'[2] (Common Agricultural Policy) reasserted the Commission's interest in establishing agri-environmental indicators and presented a timetable for their development.

At the same time, following the York seminar of September 1998, the Organisation for Economic Co-operation and Development (OECD) had started working on agri-environmental indicators at an international level. . Results of this work will be published soon.

(1) COM(1999) 22
(2) COM(2000) 20

In summer 2000 From land cover to landscape diversity in the European Union was published. This report was the fruit of a collaboration between the Commission's Directorate General for Agriculture, Joint Research Centre (JRC) and Eurostat, and the European Environment Agency.

Agenda 2000

The polluter pays' principle of course applies to agricultural production. According to the principle, farmers bear compliance costs up to a reference level of good farming practice' reflected in property rights. Many rural areas, however, have more ambitious environmental objectives than good farming practice' alone can achieve. In this case, the objectives can only be met by offering appropriate remuneration. To protect or improve the environment, it may be judged appropriate to pay farmers to use their privately owned resources or production factors for activities that go beyond good farming practice.

Reforms under Agenda 2000 represented a significant step towards realising this approach. In the common rules regulation[3], Member States are required to take measures necessary for the protection of the environment. Three courses of action are provided for. First is the application of compulsory restrictions. Such measures are already applied in Member States in relation, for example, to the pollution of water by nitrates. Secondly, Member States may apply cross-compliance by attaching specific environmental conditions to the granting of direct CAP payments. Thirdly, Member States may use agri-environment programmes to protect or enhance the environment beyond good farming practice. Furthermore, while the CAP is a common European policy, Agenda 2000 recognises that the diverse nature of the farmed environment across Europe means that the policy has to be applied in a decentralised way.

It is necessary to define what constitutes good agricultural practice. Quite clearly, beyond defining a very basic level of requirement, it is difficult for the Commission to determine the rules to be respected by a farmer in Tuscany or in the Massif Central. These should be defined at regional or local level.

Local agri-environmental systems

Every human activity has, and always has had, an impact on the environment. This is nothing new. It is also true that humans have an increasing capacity to intervene and to influence nature, both positively by protecting and building, and negatively by damaging and destroying. This also applies to agriculture. But whereas the relationship between agriculture and the environment is difficult to define at a general level except in extremely vague terms, it is relatively easy to analyse at local level. It depends on and is determined by local conditions, it is site specific'.

It is useful to refer to the concept of agri-environment, a notion which embraces the complex relationship between agriculture and environment and takes into account both positive and negative environmental impact. Indeed, a historical inter-relationship exists between agriculture and the environment. Modification of production systems has resulted in intensification, specialisation and concentration of agricultural production. The impact of these changes has been obvious as much for the environment as for the landscape.

The territorial dimension therefore appears to be a key element in understanding the development of agriculture. First of all, it incorporates the complex that includes land cover, land use and agri-environmental features. It encourages debate in spatial terms as well as allowing us to refer to the mosaic' of landscapes that we encounter. In the long term, it offers us the possibility of developing tools for the spatial description of agri-environmental features.

Secondly, rather than concentrating exclusively on individual characteristics, it may be more appropriate to refer to clusters of characteristics. Finally, these constituents of EU territorial information – land cover and/or land use, environmental characteristics and other geomorphological and socio-cultural data – can be aggregated, presented and analysed using a geographical information system (GIS).

Integrating statistical and geographical information

A well-founded and consistent system of agri-environmental indicators, which includes landscape, will contribute to the detection of

(3) Commission Regulation (EC) No 1750/1999 of 23 July 1999

environmental problems. It will help the European Union explain to its citizens what it is doing and what remains to be done to promote sustainable agriculture in Europe and at an international level. This will also help the EU's trading partners understand the importance that Europe attaches to the environmental functions of its agriculture.

In this way the CAP no longer treats EU agriculture as merely an economic system but more and more as an integrated territorial complex. This integration depends on internal balances of various kinds.

Data, which in fact are usually collected through a bottom-up process (although only made available at country or large region level) are now needed at local level. This more precise geographical positioning has to be provided. The method of collection may mean the data are readily available. However, when the detailed location is not available (problems of confidentiality) or simply not provided, an alternative is to recreate (estimate) a more precise spatial breakdown using techniques such as those illustrated in this publication. Increased spatial precision enhances statistical relevance. Geographical information is now on a par with statistical and administrative information. It is a necessary complement which ensures relevance.

Taking good account of geo-referenced (localised) data is a tool at the service of agri-environmental policies. Bringing out and putting to use the territorial information in statistical data (Farm Structure Surveys) or administrative data (Integrated Administration and Control System) by remapping them to relate to territorial units (zones) relevant to policy issues (landscape units, river basins) is an important example. That is what this publication aims to show.

It is for all these reasons that an in-depth study was necessary. A pleasing feature has been the constructive collaboration between the European Environment Agency, Eurostat, the Joint Research Centre, and the Agriculture and Environment Directorates General. This second publication is a further demonstration both of the need for such collaboration and of the results it can achieve (see http://europa.eu.int/comm/agriculture/publi/landscape/index.htm).

This publication offers a set of eight papers presenting and analysing methodological approaches for integrating statistical and administrative data with land cover information for the development of agri-environmental indicators.

Abstracts

Towards Agri-environmental Indicators:
Integrating statistical and administrative data with land cover information

Geogra hical use of statistical data
Administrative (NUTS) regions are mostly designed to express political will. However, to answer technical questions, analytical zonings enable a better approach of the territory. It is therefore necessary to transfer public statistics from NUTS regions to other zoning systems. A practical method consists in combining spatial aggregation and spatial disaggregation operations.

Dealing with the modifiable areal unit roblem
The Eurostat article deals with the problems generated by the heterogeneity of the units making up the NUTS nomenclature and of possible ways to tackle them. It is based on research work carried out in 1997 in the framework of the SUPCOM programme. The final report of this study proposes more than ten methods of aggregation, disaggregation or creation of areas from points. Conclusions underline the fact that there is no generic solution to the problems encountered. The selection of a method will highly depend on the nature of the problem to be solved and on the quality of the data available.

Com arison of CORINE land cover data with IACS data in Belgium and Italy and with land use in Slovenia
The DG AGRI article highlights the richness of the Integrated Administration and Control System (IACS) for land cover analysis. An agricultural landscape diversity index based on IACS data is calculated and commented for Belgium. Then, a comparison is carried out between IACS and CORINE (agricultural classes) inventories for an Italian region — Perugia. Finally for Slovenia, results of the land use census and the CORINE inventory are compared at commune level.

Land cover in the context of the Natura 2000 network
The European Commission is constructing a medium-scale geographic database for the European network of protected sites, called the Natura 2000 network. To fully exploit this database, it will be used in relation with other geographic datasets. This article gives an overview of the GIS for Natura 2000 project and resulting database, including the potential use of it within the daily work of staff from the DG ENV. A more detailed analysis is carried out for Greece for which the Natura 2000 data is compared with the CORINE land cover data.

Reassignment of the Farm structure Survey's data
Potentialities of the Farm Structure Survey's data can be increased by improving their location in space on the basis of the geographical information provided by CORINE land cover. As a first step the definition of links between nomenclature of both databases as well as the one of spatial distribution methods specific to each class are presented.

Using CORINE land cover to ma o ulation density
An algorithm is presented to estimate coefficients to disaggregate statistical data by modified areal weighting with the help of a more detailed geo-referenced information. The procedure is applied to combine population data at commune level with CORINE land cover. The accuracy is assessed by comparison with census section data for a test site.

Agricultural statistics s atialisation by means of CORINE land cover to model nutrient sur luses
A pilot study carried out over catchments draining a total of approx. 300 000 km^2 (France, Germany, Czech Republic) shows that the use of the CORINE land cover makes it possible to use NUTS3 level agricultural statistics to compute surplus and apportion them to the relevant environmental units, such as river catchments. The findings suggest that the major uncertainties come from agronomic constants, with special mention to the feeding source of the livestock.

Com aring CORINE land cover with a more detailed database in Arezzo (Italy)
A simple confusion matrix gives a pessimistic view of the agreement between two land cover maps at different scales. A procedure is

illustrated to separate the effects of scale and co-location accuracy from the thematic disagreement. The results show very good agreement between CORINE land cover and a more detailed database for the main land cover classes.

1. Geographical use of statistical data

Methodological overview

Claude Vidal*, Javier Gallego, Maxime Kayadjanian*****
(*Eurostat/F1, **JRC, ***LANDSIS)

1.1. Data collection and retrieval units

1.1.1. Collection units
Zoning is a division of space into homogeneous areas. A zoning is **relevant** for a given topic if the basic units are homogeneous with regard to this topic. For example, in an agricultural context, one can define zones according to the proportion of the area used for agriculture. The more specific the zoning the more relevant it is for the topic concerned. The agricultural areas defined previously will probably not be very relevant for problems of employment. Employment' areas are spread around urban areas while agricultural' ones are rather found between them. There is no zoning which is relevant for all purposes.

Administrative zoning (NUTS (4) for the European Union) consists of administrative units structured according to the hierarchy of the territorial decision making units. Public statistics are recorded and are aggregated according to NUTS. However, this zoning is not the most relevant one for answering a precise technical question. The use of information has its effectiveness improved by the relevance of zoning. Thus limiting the use of a polluting product with the aim of reducing the pollution of a watercourse is effective when the restriction is applied to the area which feeds that watercourse (drainage basin).

The cost-effectiveness of statistics and of public policies can be improved by working with data, not in the (not very relevant) zoning for which they are directly available but by restructuring them into another zoning system. When it is possible, this spatial redistribution of statistical data is faster and more effective than setting up a new tool for gathering information. Since most regional statistical data is collected and aggregated in administrative zonings, the task is to redistribute them into various other relevant zones: drainage basins (water), landscape units (environment), coalfields, employment areas, or any other *ad hoc* zoning.

From a purely statistical point of view, the ideal would be to have geo-referenced data and to consolidate them at a later date in a variety of ways according to requirements. But this is hardly realistic. Moreover statistics mostly uses sample surveys, the representativity of which depends on the zonings which were used as basis. Only in a very few cases is restructuring based on geo-referenced information possible: when the data collection units, agricultural plots (IACS (5)) or survey points (LUCAS (6)) are very small in relation to the aggregation zones.

The use of units built on a systematic geometrical division of space (squares, sampling by point grids) makes it possible to escape zonings based on a specific topic. This type of approach remains rare because it is expensive from the point of view of survey costs. However there are two British examples: population census, Countryside Survey, which use this method.

Figure 1. illustrates by some examples the possible cases of combination of zonings during the collection and during the retrieval (remapped) of the statistical data.

(4) NUTS: Nomenclature of Territorial Units for Statistics (Ref. ISBN 92-829-7275-0)
(5) IACS: Integrated Administration and Control System
(6) LUCAS: Land Use-Cover Area Statistics

Table 1. **Collection and retrieval units examples**

		Collection units		Use
	Zoning	**Geographical**	**Geometrical**	
Retrieval units	Administrative zoning e.g. NUTS	FSS	Britain population census	Political decision-making
	Ad-hoc zoning e.g. drainage basins, landscape units		Soil map	Decision-making aid
			Countryside survey	
	No zoning e.g. grids	IACS	CORINE land cover	Statistics compilation
			LUCAS	

1.1.2. Geogra hical retrieval/aggregation/ re orting units

According to the problem considered and the nature of the data available, the appropriate geographical level for data processing analysis and reporting can vary according to the geographical unit for which the data was collected. Data then has to be remapped from one zoning system to another one (see below **3**).

The way information is consolidated should take account of the territorial decision units. Thus, a policy applied to a given area requires information concerning this area.

At present most regional data is available for administrative regions (i.e. NUTS units). Now, for many applications, this aggregation is inappropriate for analysis, in particular when geographical and spatial issues are important. Moreover, alternative zoning systems to the NUTS regions increasingly constitute a new geographic frame for regional planning, economic development or environmental monitoring policies at European level. We can mention the example of the Natura 2000 areas, which aim at maintaining a certain level of biodiversity throughout Europe, or catchment areas, which constitute the natural' frame for managing water resources efficiently.

Three types of retrieve units for improved processing and reporting can be defined (Figure 1):

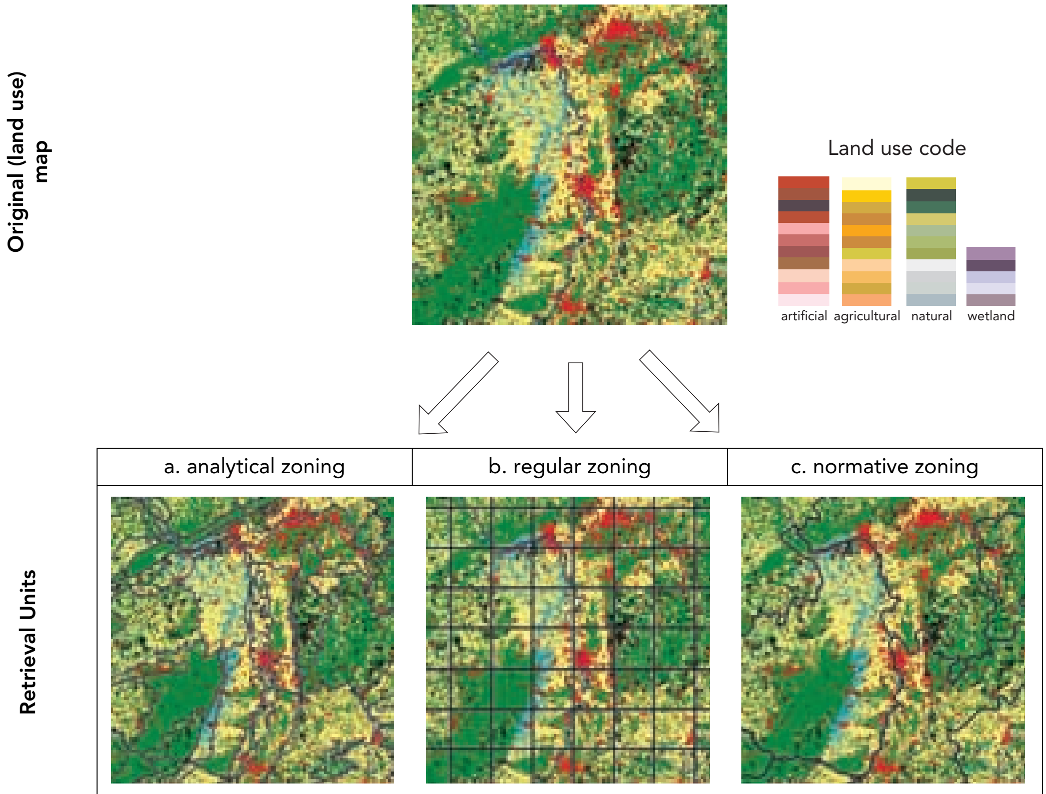
a) Geographic or **analytic regions** are defined according to the topics concerned. They are defined on the basis of geographical features such as topography (drainage basins), biodiversity (e.g. Natura 2000), landscape (e.g. French régions agricoles') or socio-economic criteria (employment areas, urban centres, cross-border areas).

   This brings special advantages to **functional** analyses (economic, environmental). Some of these divisions are frequently used in the Member States. A harmonised definition of these regions would ensure their international comparability. Unfortunately, there is still a long way to go. Moreover, there are almost as many potential divisions as there are subjects of analysis.

b) Grid or **regular zoning**, in which units have the same size, and are randomly built. This aims at avoiding any hypothesis on the spatial distribution of information. This type of zoning system is particularly used when data has to be processed with specific operators borrowing from image processing (i.e. filtering, neighbourhood analysis). In this case, regular zoning is essentially an intermediate stage for further processing. It is also a way to increase the comparability between units (all have the same area), no matter what the topic.

c) Administrative or **normative regions** are the expression of political will. Their limits are defined according to the tasks assigned to the local authorities and to the volume of population that seems to correspond to an optimum size for their successful achievement. It is to be noted that historical elements frequently underlie these territorial divisions.

Different zoning and statistical consequences — Figure 1.



	a. analytical zoning	b. regular zoning	c. normative zoning
Data diversity	low	high	high
Unit size diversity	high	null	medium
Relevance	High for topic analysed	Medium for all	High with decision
Statistical quality	Data homogeneity	Unit size homogeneity Systematic sampling	Correspondence with administrative data

From a **statistical** point of view, normative and analytical regions present advantages and disadvantages. Generally, the normative regions have a legal existence and correspond to an administrative reality in the country concerned. These are clearly defined places under the control of a specific part of the public administration, in particular implementing official authority, regional policy. Traditionally, these regions have been recognised by the national statistical offices as a framework for collecting, processing and disseminating information. The administrative and historical reasons, which led to the definition of these regions, vary widely from one country to the other. The result is a remarkable heterogeneity of the territorial breakdown even if we consider only the area or the population of the units concerned.

1.2. Spatial units for processing

1.2.1. Normative regions

1.2.1.1. NUTS

The Nomenclature of Territorial Units for Statistics (NUTS, *Nomenclature des Unités Territoriales Statistiques*) was created in the beginning of the 1970s. At that time, it was set up to define coherent territorial divisions for the compilation of comparable regional statistics at European level. This exercise also led to the creation of Eurostat's REGIO (regional statistics), which contains a broad range of regional ([7]) indicators.

The designers of the NUTS had to make a difficult choice between two major types of zones. This involves, on the one hand, the normative regions and, on the other hand, the analytical regions. The decision was taken to base the NUTS nomenclature on institutional divisions in force in the Member States. This is a pragmatic choice linked to

(7) REGIO database: User Guide - ISBN 92-828-8758-8

the availability of statistical data in the countries and to the implementation of EU's regional policies.

The first three NUTS levels (regional levels)
NUTS is a hierarchical nomenclature (see table 2.). It is defined **for the EU Member States**. The territory is subdivided into level 1 regions. Each of these is then split up into a level 2 regions, and so on. If the local level is excluded, the administrative structure of the Member States is generally based on two main regional levels (for example: regions and departments in France). The problem of delimiting economically comparable territories at each level led to the definition of an additional regional level. This additional level is derived from one or the other of the two main levels, depending on the particular Member States.

Levels 4 and 5 (local levels)
The design of Community policies has become an increasingly complex process where national, regional and local interests have to be taken into account. The result is a growing demand for more detailed information by Commission departments. In order to satisfy this demand, Eurostat created two additional NUTS local levels (NUTS 4 and 5) at the beginning of the 1990s. These two local levels were defined on the basis of the choices and principles outlined above for all the Member States.

NUTS level 5 corresponds to the *'commune'* or its equivalent. It provides the framework for the European infra-regional database (SIRE).

NUTS units constitute spatial reference units. The most detailed level (NUTS 5) allows the description of a variety of areas regardless of the notion of level provided these areas are defined in terms of groupings of NUTS 5 areas. However, the availability of data at local level is poor (problems of statistical confidentiality and representativity of information).

1.2.1.2. Statistical regions in the EFTA countries and the Central European Countries (CECs)

There is an ever-increasing demand for statistical information at a regional level for the countries of Central Europe and the EFTA countries. In order to achieve common definitions, Eurostat and the National Statistical Institutes of the CECs and EFTA have agreed to define regional levels according to principles similar to those used in the establishment of the NUTS (see table 3.). However, these classifications do not preclude any decision on the NUTS that will be taken if some of the countries join the EU.

Statistical regions are defined at three different levels. For levels 4 and 5, codes were assigned for the EFTA countries. For the CEC it is planned to do this exercise at a later stage. Table 3 shows which units might be used for these countries.

Table 2.

	NUTS 1		NUTS 2		NUTS 3		NUTS 4		NUTS 5	
BE	Regions	3	Provinces	11	Districts	43	–		Communes	589
DK	–	1	–	1	Amter	15	–		Kommuner	276
DE	Länder	16	Regierungsbezirke	40	Kreise	441	–		Gemeinden	16 176
GR	Groups of development regions	4	Development regions	13	Nomoi	51	Eparchies	150	Demoi/Koinotites	5 921
ES	Agrupacion de comunidades	7	Comunidades	17	Provincias	50			Municipios	8 077
	autonomas		Autonomas				–			
			+ Ceuta Mellila there	1	+ Ceuta Mellila there	2				
FR	Z.E.A.T	8	Regions	22	Departments	96	–		Communes	36 664
	+ DOM	1	+ DOM	4	+ DOM	4				
IE	–	1	Regions	2	Regional Authority	8	Counties/County	34	DEDS/Wards	3 445
					Regions		Boroughs			
IT	Gruppi di regioni	11	Regioni	20	Provincie	103	–		Comuni	8 100
LU	–	1	–	1	–	1	Cantons	12	Communes	118
NL	Landsdelen	4	Provincies	12	COROP regio 's	40	–		Gemeenten	672
AT	Gruppen von Bundesländern	3	Länder	9	Gruppen von	35	–		Gemeinden	2 351
					Politischen Bezirken					
PT	Unintermitting	1	Comissaoes of	5	Grupos de Concelhos	30	Concelhos-	305	Freguesias	4 208
	+ Regioes autonomas	2	coordena1cao regional				Munícipios			
			+ Regioes autonomas	2						
FIN	Manner-Suomi/Åland	2	Suuralueet	6	Maakunnat	20	Seutukunnat	85	Kunnat	455
SW	–	1	Riksområden	8	Lan	21	–		Kommuner	286
U.K.:		12		37		133		443		11 095
England	Government Office Regions	9	Counties (some grouped); Inner and Outer London	30	Upper tier authorities or groups of lower tier authorities (unitary authorities or districts)	93	Lower tier authori-ties (districts) or individual unitary authorities	354	Wards	8 619
Wales	Country	1	Groups of unitary authorities	2	Groups of unitary authorities	12	Individual unitary authorities	22	Wards	908
Scotland	Country	1	Groups of unitary authorities or LECS	4	Groups of unitary authorities or LECS (or parts thereof)	23	Individual unitary authorities or LECS (or parts thereof)	41	Wards (or parts thereof)	1 002
N. Ireland	Country	1	Country	1	Groups of districts	5	Districts	26	Wards	566
EU-15		**78**		**211**		**1093**				**98 433**

Table 3.

	LEVEL 1		LEVEL 2		LEVEL 3		LEVEL 4		LEVEL 5	
BG	Geographski Zoni	3	To be defined		Oblasti	28	Obshtini	255	Naseleni Mestza	5340
CZ	Ceska Republika	1	Groups of Kraje	8	Kraje	14	Okresy	77	Obce	6233
EE	Eesti	1	Eesti	1	Groups of Maakond	5	Maakond	15	Vald+Alev+Linn	254
HU	Magyarorszag	1	Tervezesi-Statisztika Regio	7	Megyek+Budapest	20	Statistikai Kisterseg	150	Telepules	3130
LT	Lietuva	1	Lietuva	1	Apskritis	10	Savivaldybes	56	Seniunija	446
LV	Latvija	1	Latvija	1	Regions	5	Rajoni+Pilsetas	33	Pagast+Pilsetas	
PL	To be defined		Wojewodztwa	16	To be defined				Gminy+Miasta	3356
RO	Romania	1	Regions	8	Judet+Bucuresti	42			Communes+Municipiu+ Orajse	2948
SI	Slovenija	1	Slovenija	1	Statisticne Regije	12			Obcinah	147
SK	Slovenska Republika	1	Zoskupenia Krajov	4	Kraje	8	Okresy	79	Obce a Mesta	2871
CEC		11		47		144		665		24725
IS	Island	1	Island	1	Island	1	Landsvaedi	8	Sveitarfélög	124
LI	Liechtenstein	1	Liechtenstein	1	Liechtenstein	1	Landschaften	2	Gemeinden	11
NO	Norge	1	Landsdeller	7	Fylker	19	Handelsdistrikt	90	Kommuner	435
CH	Schweiz/Suisse/Svizzera	1	Gross Regionen/Grandes Régions/Grandi Regioni	7	Kantone/Cantons/ Cantoni	26	Bezirke/Districts/ Distretti	184	Gemeiden/Communes/ Comuni	2929
EFTA		**4**		**16**		**47**		**284**		**3499**

1.2.2. Analytical units: catchment areas
The systems described for **agricultural areas** are rather open, i.e. they are directly related to other areas. Economic conditions may induce their evolution or their change (if existing structures have come to function badly). In the long term, their limits can therefore change. Other systems (described below) are much more closed, and their limits are stable.

The natural geographic entity for environmental studies when they involve hydrological processes is the catchment area. A catchment (watershed' in the USA) is an area drained by a single river. It is close to the geographer's river basin'. Catchments can be defined hierarchically. A large catchment area associated with a major river can be divided into sub-catchment areas associated with tributary rivers.

In **hydrological modelling** a territory should be systematically divided into structured catchment areas. For this purpose, one can define several levels. Rivers that flow into the sea define the first level. Their size can be very heterogeneous. In coastal areas there are generally very small catchment areas, that are difficult to determine.

For the second level we can give an indicative size, for example 50 000 km^2; first level catchment areas which are significantly larger than this have to be divided. Step by step, each catchment area of a given level can be divided into a set of catchment areas of the following level.

The catchment perimeter (named watershed' in the UK and divide' in USA) can be determined from a digital terrain model (DTM). A DTM is a geographic data layer containing the altitude of points of a grid. Its main characteristics are the resolution of the grid (for example the DTM might give one value every 100 m) and the accuracy (the altitude can be given for example with an error of ± 5m). If the resolution of the available DTM or its accuracy is insufficient, determining catchment limits can be problematic. Problems can appear even if we assume that the DTM is perfectly accurate and sufficiently dense. The line separating two catchment areas can be a sharply defined ridge, but in many cases the limits can be difficult to determine. Underground flow creates supplementary difficulty for defining precisely the border of a catchment area.

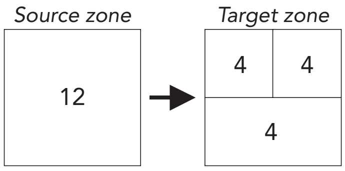
For landscape analysis, the relevant spatial units can usefully be homogeneous in terms of landscape. Landscape units can also be defined at different levels. For biodiversity issues, habitat areas are used, their nature being determined by the life style of the species considered. From a cultural point of view, influence areas for various items (languages, practices, art.) can also be defined. With regard to commercial distribution, trade areas are more relevant. There are also employment e.g. travel-to-work' areas. So zoning is linked to an issue and to a broad geographical level.

1.3. Methods and definitions

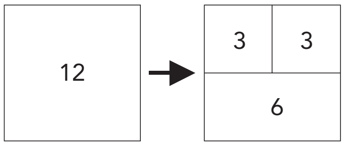
Two basic spatial operations are currently used to modify the distribution or the representation of quantitative geographical data; *spatial disaggregation* and *spatial aggregation*. When the geographical extent and/or observation units of data do not fit, layers of information have to be overlapped in order to perform aggregation or disaggregation. To facilitate explanation, we call *source zone* the zone in which data is originally attributed and *target zone* the zone to which they are remapped.

1.3.1. S atial disaggregation
Disaggregation consists of transferring the data of the source units to the target units, when target units are subdivisions of source units. Three types of disaggregation can be carried out:

Equal: the split is calculated by dividing the value of the source unit (i.e. 12) by the number of units (i.e. 3) of the target zone.

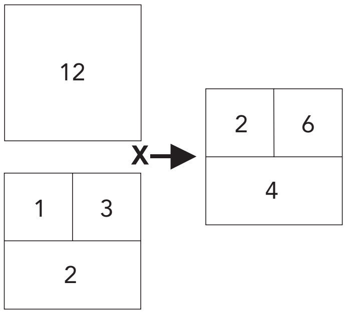


Pro ortional to area: the split is calculated by multiplying the value by the relative area of each target unit. This process is called **areal weighting.** Data is considered as homogeneously spread over the area of source units.

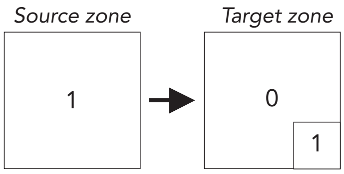


According to another variable (co-variable): the value of the source unit is distributed between the target units following the distribution of co-variable.

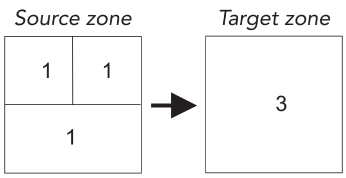
It can be done directly if the relation between the data and the co-variable is known (mathematical function or categorical relation) or indirectly, if the relation has to be statistically defined (regression). In general this relation links densities. So the area of one sub-unit is combined with the co-variable to calculate the target values.



Particular co-variable: control variable allows a part of initial target zone to be excluded. Its effect is the same as a mask, which restricts the field of disaggregation. When target units cover complete source units, density remains the same. When target units cover only part of the source units, density increases.

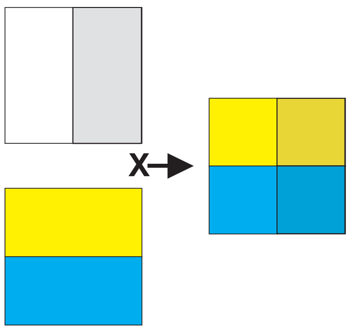


1.3.2. S atial aggregation
Aggregation is the reverse of the process of disaggregation. It consists in summarising data of the source units that fit the target units.



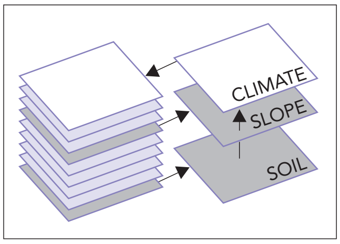
In general, adjacent units of the source zone are merged into target units putting as far as possible like with like.

1.3.3. S atial intersection
When units of a target zone are not embedded within the units of the source zone, layers have to be intersected spatially. Conversely to NUTS regions, of which units at a certain level are subdivisions of units of the above level, administrative units and catchment areas for example do not fit. They have first to be overlaid to define which unit(s) of one layer compose the unit(s) of the other one and in which proportion.

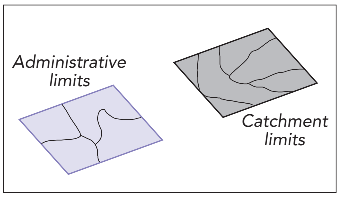


1.3.4. Processes for localising data more recisely in s ace

Spatial processing of statistical data generally aims at carrying out two basic operations:

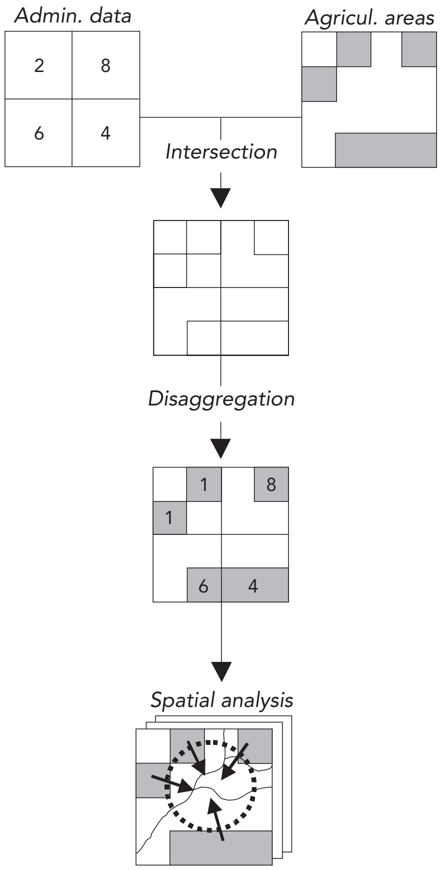


- The usefulness of statistical data attached to administrative units is severely limited in terms of spatial analysis. Reallocating data which does not have a very exact geographical reference data onto land cover units should give a more realistic geographical positioning and thus facilitate their combination with other geographical variables (e.g. climate, soil, slope).



- The administrative frame often not well adapted to carrying out specific studies on spatially distributed variables, especially environmental studies. The population or the quantity of a given type of agricultural production has to be examined for more relevant zoning system, such as landscape units and catchment areas. Data then has to be transferred from administrative units to these geographical units.

A more exact location of statistical data within administrative units can be made by means of ancillary data such as land cover/land use information. For this one seeks a strong link between the data and land cover units (8). For example, if statistical data concerns quantity of pesticides used by NUTS region level 3, there is a high probability they are related mainly to agricultural areas.



The process involves two steps. Firstly, layers of information are intersected in order to define which agricultural sub-units (target units) compose administrative units (source units). Secondly, within each administrative unit, data is disaggregated into target units.

At this step, as data is localised more precisely in space, they can be used more efficiently in spatial analysis studies, involving the combination of different geographical variables (e.g. estimation of pesticide quantity on agricultural fields and their distance to waterstreams).

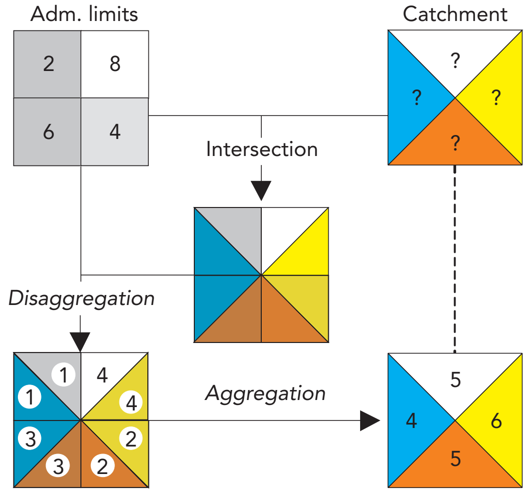
(8) See papers of M. Kayadjanian & C. Vidal in the present publication.

1.3.5. Processes for transferring data between different geogra hical zonings

Simple areal weighting

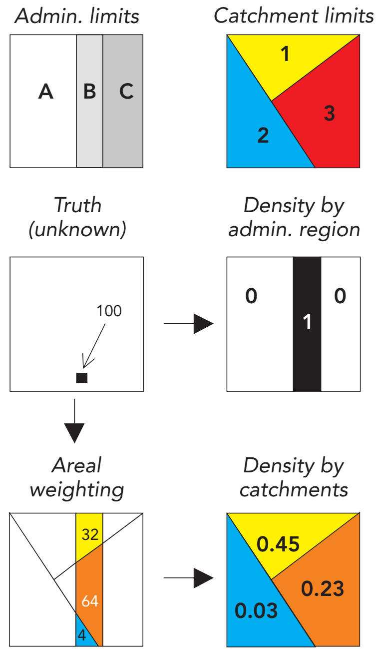
Transferring quantitative data from one geographical zoning to another one (e.g. from administrative units to catchment units) is done in three steps.

- The first step consists in calculating the intersection of both zoning systems in order to define intermediate units, embedded within both source and target units (e.g. administrative units).
- The second step consists in disaggregating the source data within the intermediate units by areal weighting. In the trivial example opposite, all source units quantities are divided by 2, because intermediate units represent half of the source units
- Finally, data is aggregated within target units.



If no additional information is integrated, this method of data transfer is the so-called ’simple areal weighting’ (Flowerdew et al., 1991). It is known to yield very poor results (Fisher and Langford, 1995). A simple example to illustrate the effects of this method on the representation of the spatial distribution of a variable follows.

We now consider a variable which happen to be concentrated at one point in space. This might be cropped areas in a narrow valley or very intensive livestock e.g. a beeflot.



In the adjacent diagram a value of 100 is noted in the administrative region B. The phenomenon happens in fact also to be localised in the catchment area 2. Official statistics give a density (assumed uniform) of say 1 which in the absence of other information is assumed to apply uniformly over the whole area of B. Applying simple areal weighting as indicated in the adjacent diagram one gets densities of 0.45, 0.03 and 0.23 in the catchment areas 1, 2 and 3 respectively.

In this example it is the catchment area in which the phenomenon concerned is actually found, which is assessed as having the lowest density. In this case of simple areal weighting the calculated densities are determined by the relative size of the intersections of zones (administrative regions and catchment areas). They have no direct link with the actual spatial distribution of the variable.

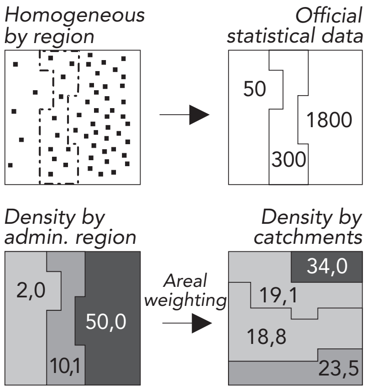
This follows from the non relevance of the spatial breakdown. The hypothesis of a homogeneous spatial distribution was wrong.

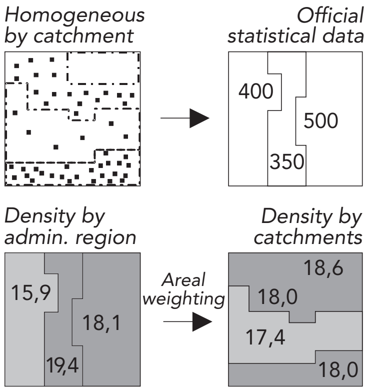
If a variable has a homogeneous distribution in each administrative region, official statistics will give a good density mapping, which will be lost (smoothed away) if simple areal weighting is performed.

If the data is really homogeneous within catchment area (rather than by administrative region) then clearly they are better represented under catchment areas.

However, much information will be lost if the density by catchment is computed by simple areal weighting. If there is a strong belief that a particular variable has a homogeneous behaviour by catchment, the possibility of computing official statistics per catchment in the future can be considered.

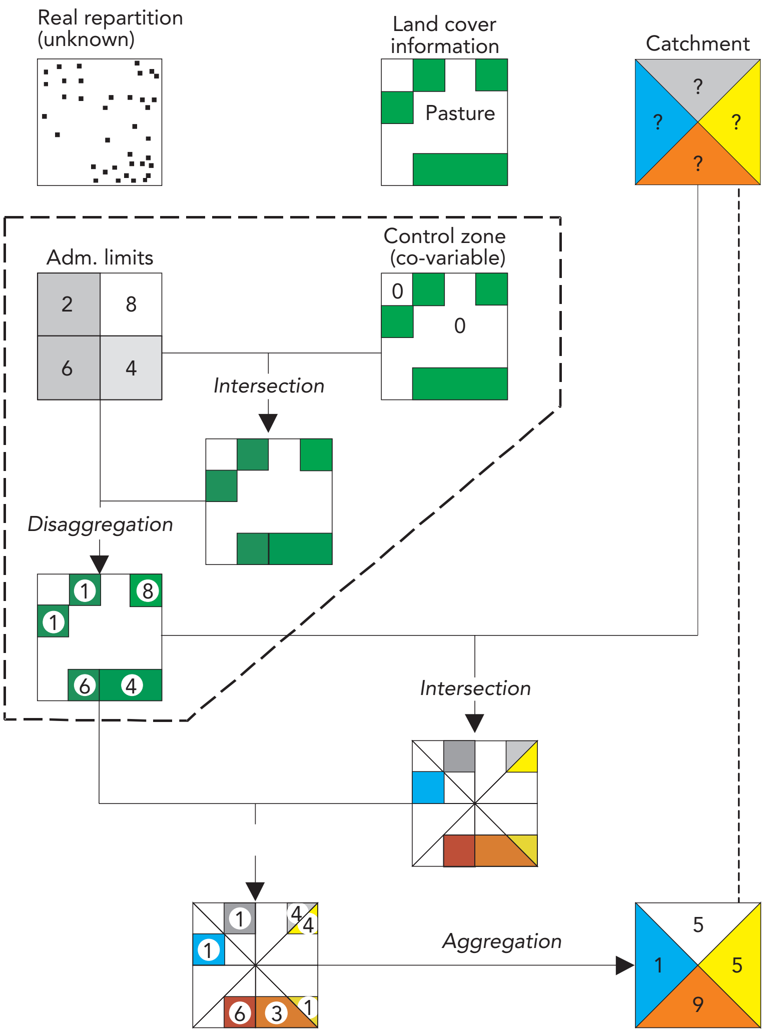
However, there would be a risk of getting wrong conclusions in a change analysis if densities for past dates for which original data is available for the old administrative region have been computed by areal weighting.





Areal weighting using control zones

To take better account of the real location of original information, data can be localised more precisely in space using control' zones before their transfer to another zoning system. The aim is to improve the way in which data is disaggregated during the transfer process in order to reduce the previously shown bias inherent to the simple areal weighting method.



We can have some external information (i.e. land cover) for which we can assume that a strong link exists with the observed phenomena (e.g. grazing livestock location and pastures).

From statistics, we only know the grazing livestock number by administrative units. In principle, the process is the same as for simple areal weighting method. But, here data is first disaggregated into control' zones (co-variable, i.e. pastures).

The application of control' zones will improve the reallocation, if the control' zones match reality, whereas simple areal weighting would degrade the quality of the data allocated to each catchment.

Using this method, we introduce additional spatial information in the second step based on the hypothesis that there is no grazing livestock if there is no pasture. The distribution of pasture is only a proxy for the distribution of livestock, but it brings useful additional spatial information.

1.4. Conclusion

Many topics need spatial analysis and statistics for administrative regions are often inadequate. This first introductory paper tries to give an overview of various sophisticated methodologies used in this publication.

History and strategic choices about statistics or territory have led to the data available relating to spatial delimitations, which are inappropriate for analysing certain issues. This can be offset by restructuring the data into more appropriate spatial groupings. In this restructuring we seek to:

- use efficient methods which limit the number of spatial operations and so also the loss of information;
- capitalise on other spatial information available using co-variables.

Subsequent papers show this approach in operation.

Annex: CORINE land cover

Conce ts:
land cover and land use are efficient territory describers. A nomenclature's quality depends on its use. Nomenclatures and the topics they were developed to cover vary greatly, e.g. urban planning and agriculture need very different nomenclatures, because the interesting variations operate at different geographical scales.

Zoning is partitioning a space into different parts, each one taken as homogeneous for a given aspect. So in a given climate area, a single climate variable is taken as applicable everywhere. In another area, a different one is better. This simplification allows a first level of understanding of complex realities. The representative value is not exact in each point, but it is near enough'.

The relevance of zoning thus **de ends on the to ic**, i. e. a zoning, which is relevant for agriculture, may not be relevant for other topics. So the agricultural basins are drawn between the urban areas, but the employment areas around them.

In the agricultural example, the zoning is based on crop homogeneity inside each area. So areas under cereals are distinguished from vineyard areas or livestock (pasture) areas.

The natural environment (soil, relief, climate) favours one crop rather than another. To take an extreme case no olive production areas can be found in Northern Europe. In addition capital stock and farmers' know-how help maintain this geographical specialisation. Finally the networks for distribution or collection of goods and services reinforce this fact, agriculture production is structured on a long term basis within its geographical environment and any redirection is difficult.

CORINE land cover (CLC) is a geographic land cover/land use database encompassing most of the countries of the European Community (except Sweden and Finland for which the inventory will be finalised in 2000–2002) and the majority of the Central and East European countries and parts of the Maghreb.

The component land cover of the CORINE program (**Coor**dination of **In**formation on the **E**nvironment) aims to gather information relating to environment on certain priority topics for the European Community (other programs are Corine Air, Corine Coastal Erosion, Corine Biotope).

Proposed in 1985 by the European Commission, CLC was initiated to satisfy the need for precise and easy accessible information on land cover in Europe. An update of the CLC database has been launched in January 2000. CLC describes land cover (and partly land use) according to a nomenclature of 44 classes organised hierarchically in three levels (table 2.). The first level (5 classes) corresponds to the main categories of the land cover/land use (artificial areas, agricultural land, forests and semi-natural areas, wetlands, water surfaces). The second level (15 classes) covers physical and physiognomic entities at a higher level of detail (urban zones, forests, lakes), finally level 3 is composed of 44 classes.

CLC was elaborated based on the visual interpretation of satellite images (Spot, Landsat TM and MSS). Ancillary data (aerial photographs, topographic or vegetation maps, statistics, local knowledge) were used to refine interpretation and the assignment of the territory into the categories of the CORINE land cover nomenclature.

The smallest surfaces mapped (mapping units) correspond to 25 hectares. Linear features less than 100 m in width are not considered. The scale of the output product was fixed at 1:100 000. Thus, the location precision of the CLC database is 100 m.

These technical (cartographic) specifications are based on the three basic requirements:

- legibility of the printed map,
- a representation of the essential features of the terrain,
- a reasonable trade-off between project operating costs and provision of land cover information requirements within overall project budgetary constraints.

The CORINE land cover database is made available in a vector and raster based format. The original vector land cover database managed by the European Environment Agency's Topic Centre on land cover (ETC/LC) is converted into a grid format for public dissemination. The grid data is available at two different resolutions: (1) a dataset with 100 m grid cell size and (2) a dataset with 250 m grid cell size. The vector to raster data conversion is performed by assigning to each grid cell the class code according to the land cover polygon(s) it overlays. If a cell has more than one possible code (i.e. it overlays more than one polygon), the most dominant land cover class (covering most of the area) within the new grid cell is assigned as the new class value (see Eurostat/GISCO Database Manual).

Table 3. **Corine land cover nomenclature**

Level 1	Level 2	Level 3
1. Artificial surfaces	1.1. Urban fabric	1.1.1. Continuous urban fabric
		1.1.2. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units
		1.2.2. Road and rail networks and associated land
		1.2.3. Port areas
		1.2.4. Airports
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites
		1.3.2. Dump sites
		1.3.3. Construction sites
	1.4. Artificial, non-agricultural vegetated areas	1.4.1. Green urban areas
		1.4.2. Sport and leisure facilities
2. Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land
		2.1.2. Permanently irrigated land
		2.1.3. Rice fields
	2.2. Permanent crops	2.2.1. Vineyards
		2.2.2. Fruit trees and berry plantations
		2.2.3. Olive groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous agricultural areas	2.4.1. Annual crops associated with permanent crops
		2.4.2. Complex cultivation patterns
		2.4.3. Land principally occupied by agriculture with significant areas of natural vegetation
		2.4.4. Agro-forestry areas
3. Forests and semi-natural areas	3.1. Forests	3.1.1. Broad-leaved forest
		3.1.2. Coniferous forest
		3.1.3. Mixed forest
	3.2. Shrub and/or herbaceous vegetation associations	3.2.1. Natural grassland
		3.2.2. Moors and heathland
		3.2.3. Sclerophyllous vegetation
		3.2.4. Transitional woodland scrub
	3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes, sand plains
		3.3.2. Bare rock
		3.3.3. Sparsely vegetated areas
		3.3.4. Burnt areas
		3.3.5. Glaciers and perpetual snow
4. Wetlands	4.1. Inland wetlands	4.1.1. Inland marshes
		4.1.2. Peat bogs
	4.2. Coastal wetlands	4.2.1. Salt marshes
		4.2.2. Salines
		4.2.3. Intertidal flats
5. Water bodies	5.1. Continental waters	5.1.1. Water courses
		5.1.2. Water bodies
	5.2. Marine waters	5.2.1. Coastal lagoons
		5.2.2. Estuaries
		5.2.3. Sea and ocean

2. Dealing with the modifiable areal unit problem

Spatial transformation methods for the analysis of geographic data

Daniel Rase (Eurostat/E4, GISCO)

2.1. Background

At the beginning of the 1990s, Eurostat created the SIRE database (European Infra-Regional information System) as a response to an increasing demand for detailed information emanating from the Commission's departments. This database contains about thirty variables from the 1981 and 1991 population censuses. The link between SIRE and the Geographical Information System of the Commission (GISCO) was established via the level 5 of the NUTS (Nomenclature of Territorial Units for Statistics). This level corresponds to the commune level or its equivalent.

Using these two observation dates, the GISCO team had the idea, in 1995, of producing a map representing the total population variation at communal level between 1981 and 1991. This map was immediately criticised by the cartographers. Along the Franco-Belgian border the reader could observe a very strong spatial discontinuity caused by the big difference in area between the communes of both countries. The map shows very significant differences in France where the communes are small whereas these disparities are made less obvious by the large size of the Belgian communes.

Convinced by the arguments of the specialists, Eurostat launched a call for tenders (under the SUPCOM research project programme). The latter invited specialists in spatial analysis and cartography to study solutions to the problem encountered. The question raised was simple: how can a correct European map at level 5 be created? The answers, for their part, appeared to be much more complex.

The Nene Centre for Research (Northampton) made a successful bid. From the beginning, Eurostat proposed to extend the scope of the study in order to include the problems generated by the transfer of information from one type of territorial division (for example: administrative units) to another type of division (for example: drainage basins). This question was raised by agricultural and environmental specialists wishing to combine socio-economic data with environmental data in their analyses. This environmental data is generally collected on the basis of points (e.g. meteorological stations) or on the basis of specific zones (e.g. Natura 2000 protected areas).

Eurostat also insisted that the contractor only proposed methods that could be put into practice with standard tools. The GISCO team wanted to apply the solutions directly to its production environment. The final report of this study reviews the various methods providing ways to tackle the problems caused by the heterogeneity of the territorial breakdown and apply them to case studies mostly based on Eurostat data.

Other researchers were interested in the subject. The Hypercarte' [9] project, which was set up in response to Eurostat's call for tenders, deserves being especially mentioned in this respect. Although it was not selected, this team continued to work on the problem and published several papers or working papers on the subject. The Hypercarte' team followed another direction than the one proposed in the SUPCOM report. What it mainly challenges in Eurostat's approach is the use of existing GIS tools. It considers that the solutions proposed in the report are very expensive in terms of processing as well as ineffective from a theoretical and practical point of view. According to these researchers, the principal usefulness of the report is the demonstration of the inadequacy of existing tools to solve the problem. Instead, they prefer a conceptual and theoretical approach which leads to the creation of new spatial analysis and map-making tools. They also proposed different empirical applications of those new methods on various scales (world population distribution, European Union, Belgium, Franco-Belgian border, etc.). Their

(9) http://www.parisgeo.cnrs.fr/cg/hyperc/index.htm

methods were adopted and developed by various public administrations and statistical services in France (INSEE, IFEN) and Europe (CORILIS Project from EEA, Study Programme on European Spatial Planning).

The debate continues (10). It is not up to the GISCO team to make a final assessment of these questions. Nevertheless, Eurostat has followed up the SUPCOM project. GISCO asked a team from JRC (11) in ISPRA to apply one of the simple methods proposed in the report. They tried to disaggregate population data at NUTS level 5 by using the CORINE land cover layer in order to distribute the data in a more realistic way.

2.2. The limitations of the NUTS nomenclature

The heterogeneity of the territorial breakdown
The pragmatic choice made by Eurostat to base the NUTS nomenclature on institutional divisions has an inpact on the homogeneity of the territorial breakdown.

Tables 1. and 2. demonstrate the disparities existing for the same level within the countries and between the countries.

Table 1. **Area of the regions (km²)**

	NUTS 1	NUTS 1	NUTS 1	NUTS 2	NUTS 2	NUTS 2	NUTS 3	NUTS 3	NUTS 3
	Ave	Min	Max	Ave	Min	Max	Ave	Min	Max
EU-15	41 676	161	410 934	15 406	31	154 312	2 974	12	98 911
BE	10 173	161	16 844	2 774	161	4 440	710	101	2 016
DK	43 094	43 094	43 094	43 094	43 094	43 094	2 873	97	6 173
DE	22 295	404	70 554	8 918	404	29 480	809	36	3 058
GR	32 906	3 808	56 457	10 125	2 307	18 811	2 581	356	5 461
ES	72 113	7 242	215 025	28 044	31	94 193	9 708	12	21 657
FR*	70 361	12 012	145 645	24 356	1 128	83 934	6 333	105	83 934
IE	70 273	70 273	70 273	35 137	33 276	36 997	8 784	922	14 283
IT	27 392	13 595	44 430	15 066	3 264	25 707	2 925	212	7 520
LU	2 586	2 586	2 586	2 586	2 586	2 586	2 586	2 586	2 586
NL	10 382	7 291	11 871	3 461	1 434	5 741	1 038	128	3 429
AT	27 953	23 554	34 384	9 318	415	19 173	2 396	415	4 615
PT	30 635	779	88 797	13 129	779	26 931	3 064	779	8 503
FI	152 265	1 526	304 529	50 755	1 526	128 294	15 226	1 526	93 003
SE	410 934	410 934	410 934	51 367	6 490	154 312	19 568	2 941	98 911
UK	20 318	1 584	78 132	6 590	320	39 777	1 833	35	14 295

* incl. DOM

(10) Grasland C., Mathian H., Vincent J.M.: 2000, Multiscalar Analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences, *Statistical Journal of the United Nations ECE*, 17, IOS Press, 1-32.
(11) JRC: Joint Research Centre

Population of the regions 1996 (1000) Table 2.

	NUTS 1	NUTS 1	NUTS 1	NUTS 2	NUTS 2	NUTS 2	NUTS 3	NUTS 3	NUTS 3
	Ave	Min	Max	Ave	Min	Max	Ave	Min	Max
EU-15	4 811	25	17 920	1 787	25	11 044	345	20	5 025
BE	3 386	949	5 890	923	242	1 633	236	39	949
DK	5 263	5 263	5 263	5 263	5 263	5 263	351	45	627
DE	5 120	679	17 920	2 048	507	5 291	186	36	2 166
GR	2 619	1 011	3 449	806	184	3 449	205	21	3 449
ES	5 671	1 610	10 867	2 205	129	7 244	763	60	5 025
FR*	6 669	1 644	11 044	2 308	160	11 044	600	73	2 563
IE	3 626	3 626	3 626	1 813	965	2 661	453	206	1 058
IT	5 218	1 603	8 942	2 870	119	8 942	557	92	3 778
LU	416	416	416	416	416	416	416	416	416
NL	3 883	1 631	7 253	1 294	277	3 339	388	54	1 311
AT	2 686	1 770	3 395	895	275	1 595	230	21	1 595
PT	3 309	242	9 428	1 418	242	3 538	331	47	1 834
FI	2 562	25	5 099	854	25	1 813	256	25	1 232
SE	8 845	8 845	8 845	1 106	391	1 755	481	58	1 744
UK	4 900	1 663	7 895	1 589	372	4 366	442	20	1 728

* incl. DOM

In terms of area, we can observe enormous differences within the same country. Spain, for example, has a minimum size of 12 km^2 and a maximum of 21 657 km^2 at level 3. The problem is identical between the countries: we can observe at NUTS 3 an average of 9 708 km^2 in Spain and 809 km^2 in Germany.

Similar disparities exist with regard to the population: we can see NUTS 3 regions ranging from 60 000 to 5 025 000 inhabitants in Spain; an average at NUTS 3 ranging from 763 000 inhabitants in Spain to 186 000 inhabitants in Germany.

It should be noted that the majority of the maps produced today (at levels 1, 2 or 3) as well as the statistical analyses based on the NUTS do not take these disparities into account. In general, the geographical component of the problem is disregarded. Certain researchers have used a mixture of NUTS 2 and NUTS 3 in their work. Indeed, the average area of NUTS level 3 in France approaches the average area of NUTS level 2 in Germany. This type of approach makes it possible to obtain more homogeneous territorial units in terms of area but could create problems when decisions are taken on the basis of results combining different political levels.

The problem is even more obvious at commune' level (NUTS 5). Table 3. provides the same information as the preceding ones but at local level.

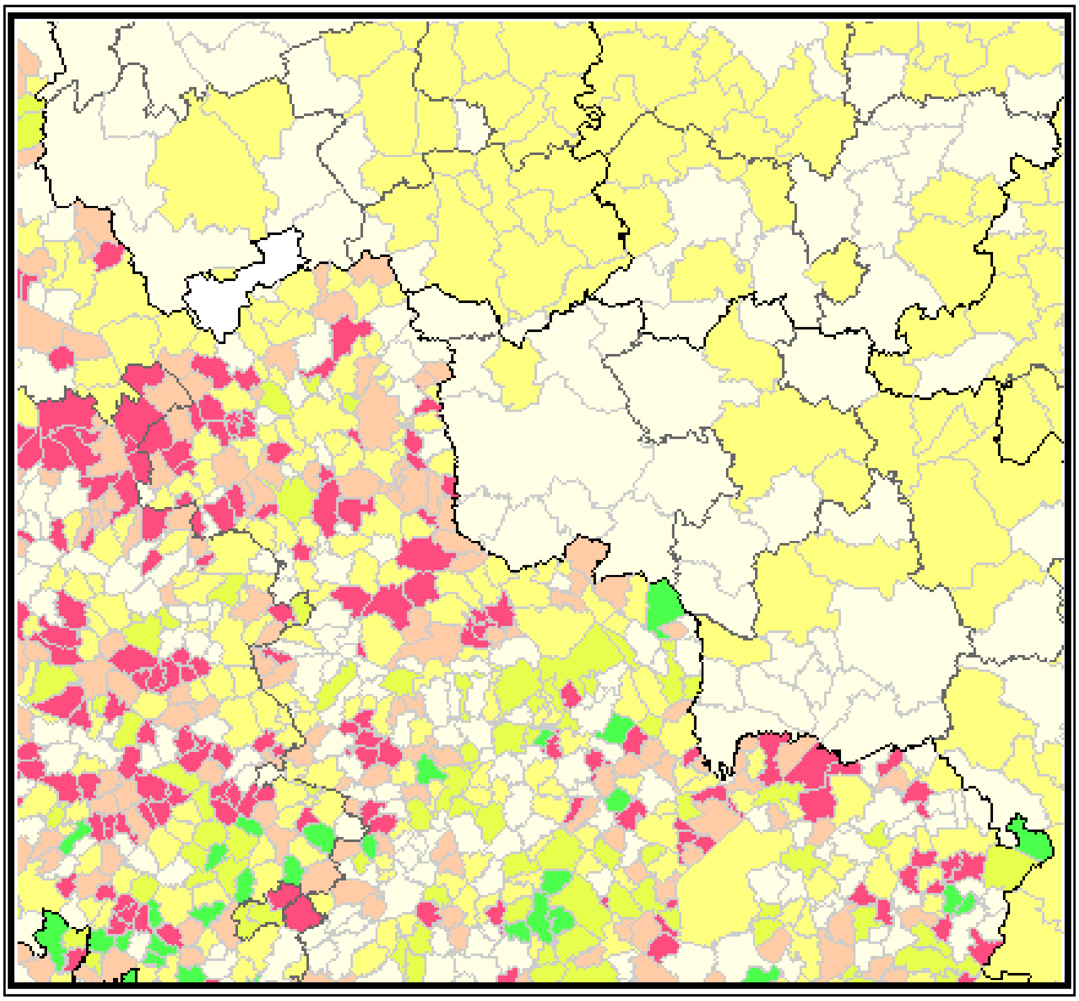
Table 3. **Area and population of the communes (1991)**

codcoun	min (area in hectares)	max (area in hectares)	ave (area in hectares)	min (total population)	max (total population)	ave (total population)
EU-15	0	1944678	3296	0	3010492	3704
BE	114	21375	5182	93	467518	16942
DK	40	56364	15614	111	464773	18647
DE	0	75531	2205	0	2164904	4962
GR	43	57717	2228	0	772072	1732
ES	10	175030	6271	0	3010492	4813
FR	4	1836000	1726	0	2152329	1584
IE	5	12333	2045	0	25843	1023
IT	10	149874	3720	31	2775250	7010
LU	528	11336	2192	183	75833	3260
NL	171	46571	5049	234	698917	22216
AT	11	46688	3594	50	1539848	3341
PT	5	46177	2184	47	79801	2344
FI	590	1517280	66216	129	497542	10933
UK	0	281041	2188	0	31609	4911
SE	883	1944678	144695	2924	679364	30437

To convince ourselves of the difficulties generated by these disparities, we refer to figure 1. This map represents the variation of the population at NUTS level 5 between the censuses of 1981 and 1991 along the Franco-Belgian border ([12]). We observe a border effect caused by the difference in size of the territorial units between France and Belgium. This phenomenon is linked to the merging of the communes carried out by Belgium in the 1970s.

(12) see also: Grasland C., March 2000, 'Facing the MAUP: The example of variation of population at the Franc-Belgian border', The Hypercarte Project, Working Paper n°3: http://www.parisgeo.cnrs.fr/cg/hyperc/wp3/wp3.htm

Figure 1.



It can be summarised that the disparities observed at the first three levels of the NUTS are even more emphasized at NUTS level 5. When we look at maps, there is a visual dominance of the large units masking the fact that the smallest units are actually generally the most populated. Another phenomenon to mention is the border effect appearing when the units of two countries are of very different size. These disparities also make it difficult to use NUTS level 5 as a building block for the definition of aggregates such as the urban areas. By virtue of its small size, the French NUTS 5 would be ideal in this context. This is not the case for Spain and even less for Sweden.

Do we really need to worry about this problem? The following paragraph which deals with the Modifiable Areal Unit Problem' will try to provide a (provisional) answer.

2.3. The modifiable areal unit problem (MAUP)

The cartographic representation of data and the results of spatial analysis do not produce an objective output. The result is largely influenced by the way micro-data was aggregated. What the reader perceives is a combination of basic data, the zoning system used and cartographic aspects (choice of the classes, colours, symbols, etc.). Cartograms (Dorling, 1995) try to overcome the problem by allocating equal areas of the map to equal numbers of people instead of equal areas of space allocated to equal areas of land. Each region is then represented by a circle with its area proportional to the population of the region. It is positioned as accurately as possible at its geographical location.

Stan Openshaw formalised the MAUP in a small document published in 1984.

The MAUP is subdivided in two closely related sub-problems. First of all, there is the scale problem generating a variation of the results when the number of analysed territorial units is gradually reduced (for example while going from NUTS level 3 to NUTS 2 and then to NUTS 1). The second component of the problem is linked to the choice of the division used at the scale under consideration. It is indeed possible to define a great number of aggregates from a set of basic units, which are likely to produce different results.

Figure 2. is an illustration of the impact of both aspects of the MAUP.

In this example, 100 individuals are distributed over a square of 100x100. The assumption is that the space occupied by each individual equals a square of 1x1. Thus, the first map (2.a) represents the finest (atomistic) level: the maximum information available on the distribution of the population of interest.

It is assumed that the practical problem to be solved is the realisation of a map of the population density in the observed area.

Figure 2.d shows a first possible map of population density based on a triangulation of space, according to minimum distance between each individual. In this case, it is possible to measure the density of population in every point, according to the inverse of the area of the Thiessen's polygon built around each inhabited point. But in the general case, the individual information is not available and one starts from a territorial division of space in cells where more than one individual may be located.

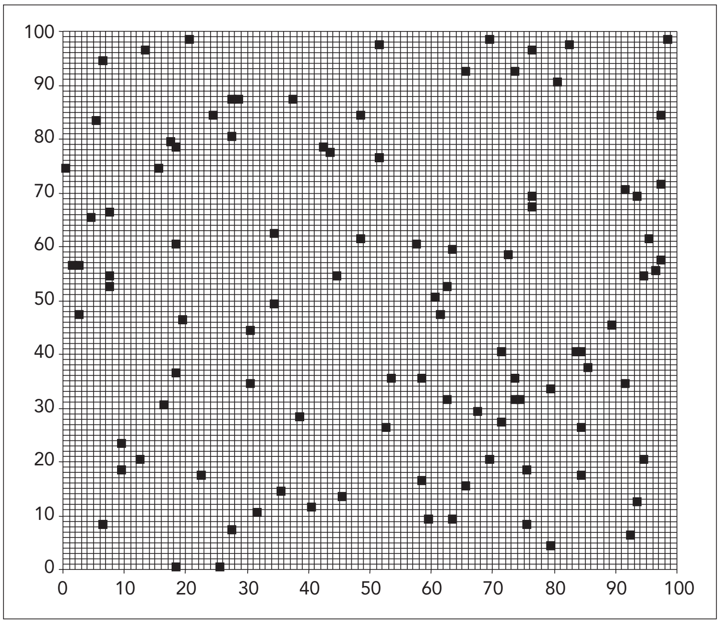
Figure 2.b shows a territorial partition based on a regular grid where a network of 20x20 square cells was chosen. The configuration of density in a grid map of population density (fig 2.e) is related to the scale of the aggregation process but also to the shape of areal units (squares, hexagons or triangles) and their reference coordinates.

Most of the time, the territorial partition is based on heterogeneous administrative units. This is what is shown in fig. 2.c.

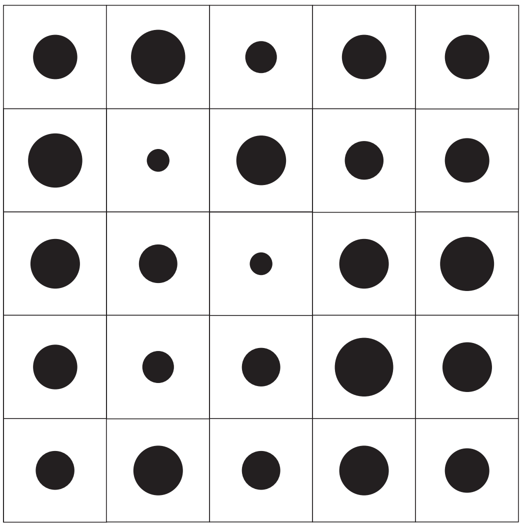
The number of territorial divisions in fig. 2.b and 2.c are approximately equal (25 and 27). But the loss of information on the location of individuals is more important in the case of administrative division than for a regular grid. This is due to the heterogeneity of the administrative divisions.

We will see in the following paragraphs that the problem was the subject of numerous methodological studies, none of which lead to a standardised solution. Several methods exist. Depending on the nature of the problem considered, it will be necessary to opt for one or other of these. In any case, the decision should be made very carefully.

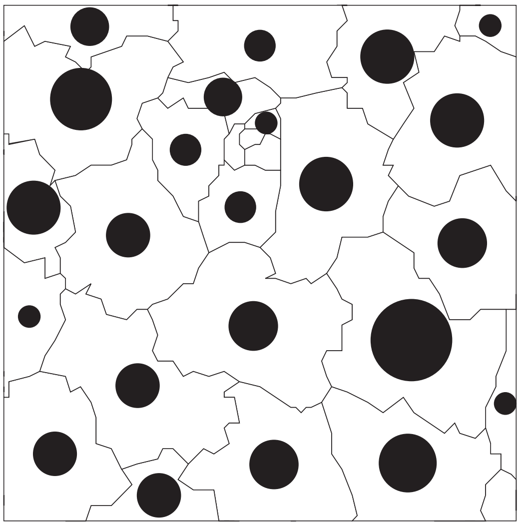
Figure 2 **Variation of density according to territorial divisions**



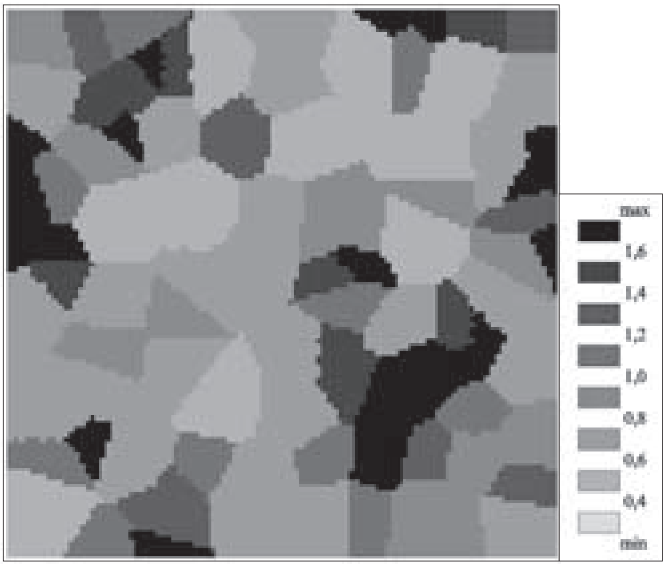
2.a)Information level 1 (100 individual position: maximum information)
Random sample of 100 points distributed on a grid with size 100x100



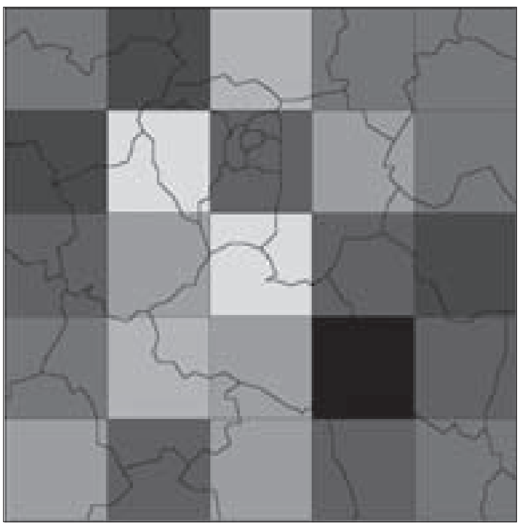
2.b) Information level 2: regular aggregation
Projection of a regular grid size 20x20 on the random sample of points



2.c) Information level 3: irregular aggregation
Projection of French administrative divisions on the random sample of points



2.d)Density derived from individual information
Based on the inverse of the area of Thiessen's polygons



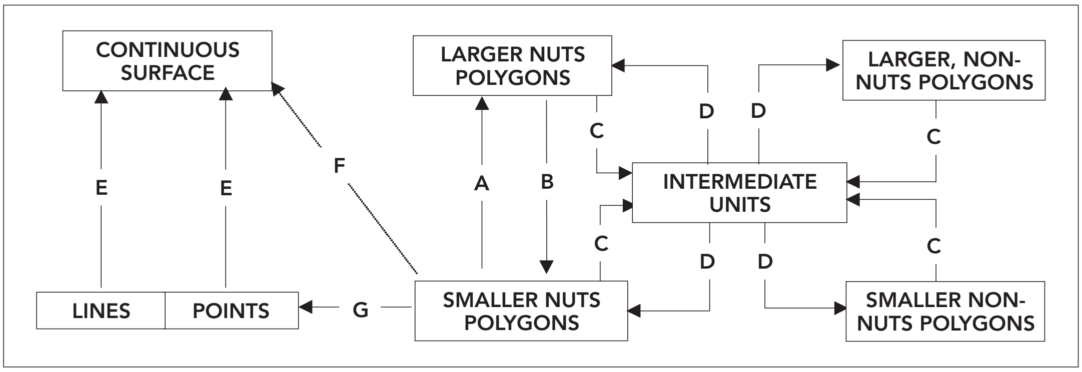
2.e) Density derived from grid information

2.f) Density derived from administrative information

2.4. Methods

Figure 3 represents a generic model of the questions raised by Eurostat under the SUPCOM project of 1997.

Figure 3. **A generic model of spatial transformations for use with Eurostat data**



Key: A = nested aggregation; B = nested disaggregation; C = non-nested disaggregation; D = non-nested aggregation; E = area modelling (interpolation) based on-line but at all dated; F = area modelling (interpolation) based on polygon dated; G = centroid creation.

'A' represents small NUTS units aggregated towards larger NUTS units (for example to aggregate units of NUTS level 5 to build units of NUTS level 3). 'B' represents the opposite operation.

'C'+'D' illustrate the use of an intermediate unit to aggregate or disaggregate information and then to transfer it to another type of unit. This involves for example transferring NUTS level 3 data to drainage basins by using a transitional unit.

'E' represents the creation of areas from points or lines. In 'F' the same operation is carried out starting with polygons. Finally, 'G' represents the shift from polygons to their centroids.

Methods can be classified in five main categories:

- Methods for disaggregation of increasing complexity;
- Methods for transferring information from one type of unit to another;
- Methods for aggregating information to form more homogeneous units;
- Smoothing methods of varying degrees of complexity;
- Interpolation methods which make it possible to create areas from points.

The annex contains a short description of the various methods considered in the study.

The complete report of the 1997 SUPCOM project contains case studies illustrating some of the methods briefly decribed in this document. In most of them, Eurostat data was used. The report also contains an extensive annotated bibliography. A paper copy of the complete report can be obtained by addressing a mail to the GISCO bureau. ([13])

2.5. Conclusion

During the SUPCOM project, a model for the processing of statistical data was developed. It shows the relationship between the three possible processes that were identified: aggregation, disaggregation and generation of surfaces. More than ten different techniques were identified (see annex).

There is unfortunately no generic solution to the questions raised in this paper. The choice of one of the methods will highly depend on the nature of the problem to be tackled as well as on data availability and data quality.

(13) daniel.rase@cec.eu.int

References

Eurostat, (1999), 'Nomenclature of territorial units for statistics — NUTS'

Eurostat, (1999), 'Statistical regions in the EFTA countries and the Central European Countries (CEC)'

Dorling, D., (1995), 'A new social atlas of Britain', John Wiley & Sons

Grasland C., March 2000, 'Facing the MAUP: The example of variation of population at the France-Belgium border', The Hypercarte Project, Working Paper n°3: http://www.parisgeo.cnrs.fr/cg/hyperc/wp3/wp3.htm

Grasland C., Mathian H., Vincent J.M.: 2000, Multiscalar Analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences, Statistical Journal of the United Nations ECE, 17, IOS Press, 1-32.

Hunting Technical Services in association with Nene Centre for Research, GIS Application Development (SUPCOM 1997 — LOT 3) — Final Report.

Openshaw, S., (1984), The modifiable areal unit problem, CATMOG (Concepts and Techniques in Modern Geography) N 38.

Annex

TECHNIQUE (1)	Simple areal weighting/weighted overlay

PROCESS	Disaggregation (B, C + D)
POTENTIAL APPLICATION	Disaggregation of NUTS units to a grid Disaggregation of NUTS 3 units to NUTS 5 Disaggregation of NUTS units to non-NUTS units e.g. catchment areas
DESCRIPTION Areal weighting is probably the simplest and most widely used method for redistributing spatial data to new polygons. It is achieved by overlay of target zones and source zones and the subsequent determination of the area of intersection. This process assumes that the attribute of interest is distributed uniformly within all zones. It can be applied to either intensive (rate or proportion) or extensive (count or total) variables.	
PROS Easy to compute Data requirement is low Technique and algorithm is well developed and adopted in many commercial GIS packages.	**CONS** The assumption of uniform density of the attribute (e.g. population) is a simplistic model. This approach does not do justice to the additional information that is available in GIS and capabilities offered by digital processing

TECHNIQUE (2)0	Modified areal weighting using statistical/regression methods

PROCESS	Disaggregation (B, C + D)
POTENTIAL APPLICATION	Disaggregation of NUTS 3 units to NUTS 5 Transfer information from NUTS units to non-NUTS using ancillary information.
DESCRIPTION Modified areal weighting builds on the principles of areal weighting, using ancillary information (and not area) in the form of a predictor in order to 'inform' the distribution. The precise method used depends on the nature of the predictor variable used and the way in which it is applied. In the case of the regression methodology, the predictor is provided by one or more covariates, which have a definable statistical relationship with the variable of interest. This relationship can be derived in a number of ways: by analysis of data at the coarser level of aggregation; from more detailed data within a representative subregion; or from more detailed data for a sample of zones.	
PROS Acknowledges the complex relationships that may exist between variables. Auxiliary information available for target zones is incorporated in a formal and statistically appropriate way.	**CONS** Difficulty of finding suitable ancillary variables which are strongly related to the variable of interest. Target zone figures are estimates and cannot be guaranteed to be accurate.

TECHNIQUE (3)	Modified Areal Weighting (control zones)

PROCESS	Disaggregation (B, C + D)
POTENTIAL APPLICATION	Transfer information from NUTS units to non-NUTS using ancillary information.
DESCRIPTION Areal weighting uses the known data for a zone (source zone) to construct estimates for another zone (target zone). When they do not nest hierarchically and are assumed not to be homogenous, a third set of zones, referred to as control zones, can be used to aid the interpolation. Control zones are zones that are believed to have constant densities, such as urbanised areas or agricultural regions. They can come from a variety of different sources, either as digitised land use maps, dasymetric maps derived from remote sensing data or the expert knowledge of an area from the analyst	
PROS Acknowledges that the within polygon density distribution is unlikely to be uniform. Allows user input of prior knowledge. Small amount of subjective information can lead to greatly improved accuracy.	**CONS** Although goes some way to solving problem of non-uniform density, it still relies on a simple Boolean model. Difficulties in finding appropriate ancillary data to use area control zones. May rely on classified remote sensing data, which may only be 85-95 % accurate with unknown error distribution.

Optimisation (Redistricting/Regionalisation)	TECHNIQUE (4)

PROCESS	Aggregation (A, C + D)
POTENTIAL APPLICATION	Aggregation of NUTS 5 units to new units.
DESCRIPTION Criteria are set for the aggregation of existing units or regions, optimising one or more characteristics. The criteria can be based on heuristic rules, strict constraints or loose characteristics. The most common criteria are homogeneity, equality and compactness. To achieve homogeneity, units are merged into regions similar in terms of one or more attribute. To achieve equality, units have equal values of some attribute (often population) and compactness is a shape constraint. An Arc/Info based Zone DEsign System (ZDES) based on these principles has been developed by Openshaw. Another system called SAGE (Spatial Analysis in a GIS Environment) was also developed in the Sheffield Centre for Geographic Information and Spatial Analysis. It is freely available on the Web.	
PROS Fast processing time. Creates new units better suited to the analysis, especially if the data in question are small counts.	**CONS** Locally optimal solution may not be the best globally. Can be sensitive to the initial allocation of polygons to regions, different initial allocations producing different results. Shape constraint is difficult to define, sometimes resulting in strange linear areas. Number of posible solutions is very large. User input is required to assess whether the solution is suitable.

Simulated annealing/combinatorial optimisation	TECHNIQUE (5)

PROCESS	Aggregation (A, C + D)
POTENTIAL APPLICATION	Aggregation of NUTS 5 units to new units.
DESCRIPTION The procedure of simulated annealing is based on the physical process of annealing. When a metal or crystal in a liquid form is cooled slowly, and the system allowed reaching thermal equilibrium after each temperature drop, the atoms are able to lose energy at a rate slow enough so that an ordered lattice is formed. If the system is cooled too quickly thermal equilibrium is not reached and the atoms lock in an irregular structure. Simulated annealing differs from regionalisation in that the process cannot get stuck in local optima and so strange shapes are avoided.	
PROS Used random elements to overcome the dependence to initial conditions. Does not become trapped in local optima. Can solve optimisation problems that are otherwise hard to solve.	**CONS** Long time processing Does not produce a singleoptimum solution. Reiterations will produce different solutions.

Bayesian Map Smoothing	TECHNIQUE (6)

PROCESS	F
POTENTIAL APPLICATION	Mapping of rare event data (e.g. mortality rates, ethnic minority populations)
DESCRIPTION Rare event data is often highly unstable, due to the marked effect of relatively small changes in the numerator, and differences in the denominator from one area to another. When such data is mapped at the small area level (e.g. NUTS 5), they thus often show complex patterns of variations that may, in part, be artefacts of the data. With population based data, the most unstable estimates often tend to occur in more sparsely populated regions (e.g. rural areas) which are also defined by larger map units. These tend to dominate the map, and draw the reader' s eye to the very areas for which the data is least reliable. Bayesian 'map smoothing' techniques provide means of reducing these effects, by adjusting the estimates in each area to take account of the reliability of the estimate. In areas with a small denominator, the estimate is adjusted towards the mean; in areas with a larger denominator, the rate is left largely unadjusted. The overall effect of this method is to reduce the range of variation in the data, and produce a smoother map.	
PROS Provides best estimates of actual rates for rare event data Reduces effects of artefacts in data Tends to remove much random variation in small area data and produce simpler, more easily interpretable maps.	**CONS** Not available as part of proprietary GIS or most statistical packages Computationally complex Not yet widely used, other than for research applications Resulting map may be difficult to explain to non-specialist.

TECHNIQUE (7)	Polygon filtering

PROCESS	F
POTENTIAL APPLICATION	Create a statistical surface (e.g. percentage elderly population) from NUTS 5 units
DESCRIPTION Polygon filtering is an adaptive filtering process based on the simple concept of low pass filtering and is applied to intensive (ratio or rate) variables. The value of the filtered area is normally calculated by a weighted mean of its preceding value and the values of all its neighbours. The neighbourhood definition of an area can be set in different ways, for example topological neighbours (of first, second or higher order), polygons within a certain distance or only adjacent polygons having a traffic link.	
PROS May reveal underlying continuous phenomena or processes hidden by representation of administrative units. The cartographic influence of extreme values in areas with low population or small area is smoothed out. Only a few iterations are required to achieve the result.	**CONS** Assumed spatial autocorrelation. Results may need further classification, as the large number of different values will create a 'noisy' map.

TECHNIQUE (8)	Pycnophylactic (mass-preserving) interpolation

PROCESS	F
POTENTIAL APPLICATION	Create a surface of GDP from NUTS 3 data
DESCRIPTION Data available in zones is smoothed by an iterative algorithm that takes into account the data values of the neighbouring zones, whilst preserving the count number inside each zone. A high-resolution lattice or grid is laid on top of the study area. Each grid cell falling into a particular administrative unit is initially assigned the same proportion of the total district population. Then a moving filter operates on the grid, which replaces each grid value with a weighted average of its neighbouring values. After each iteration, the pycnophylactic constraint is enforced by adjusting the new population total to match the initial total. Since the filter will operate across district boundaries, the cell values will be modified from district boundaries inwards. The process is complete when any further adjustment would be smaller than a specified tolerance level and the area is very smooth.	
PROS Raster form can be incorporated into a GIS model Within zone distribution is revealed Useful if no ancillary information is available Can be applied to total population or any other count at zone level. Can be used to model change over time and show up patterns	**CONS** Simplistic model Population in the source zone is assumed to be relatively evenly distributed. Steep gradients (e.g. between rural regions and urbanised areas) are not handled well because of the enforcement of maximal smoothness May be over smoothed Computationally intensive

TECHNIQUE (9)	Focal functions

PROCESS	(E, G + E)
POTENTIAL APPLICATION	Create a population density 'surface' from NUTS 5 data, expressed as centroids of NUTS 5 areas
DESCRIPTION Focal functions comprise a suite of methods for computing values at new (unsampled) location based on the values of surrounding or neighbouring (but not necessarily adjacent) points. Values from the contributory points may be analysed in various ways, to provide different measures, using different focal functions. Examples include: — Focal gravitation computes the inverse-square distance-weighted average of all values within the neighbourhood — Focal maximum compute the maximum value of the contributory points within the neighbourhood The neighbourhoods used to compute these focal functions may be defined in various ways, for example: — In terms of distances (by defining a circular window of fixed radius around each location) — In terms of number of points, of intervisibility or 'cost of distances' like travel time Focal functions are applied to an entire map by visiting each new location in turn, and computing the values according to the specific function.	
PROS Focal functions are provided in many proprietary GIS. Highly flexible — users can define the function and neighbourhood according to need Different functions may be used in sequence to investigate different aspects of the data Computationally efficient	**CONS** Quality of the resulting map is dependent on the spacing and distribution of sample points Results depend upon the choice of function and the way in which the neighbourhood is defined

Weighted centroid smoothing | TECHNIQUE (10)

PROCESS	(G + E)
POTENTIAL APPLICATION	Create a population density 'area' from NUTS 5 units and their centre of gravity points

DESCRIPTION

The count data of each zone is distributed from the population-weighted centroid (gravity point) into cells of a fine resolution raster grid according to a basic distance decay model. It can be viewed as an adaptative kernel estimation technique.

A moving window (kernel) visits each centroid in turn, analysing the local centroid density to determine the size of the kernel. The count associated with the centroid is then distributed to the cells falling within the kernel, according to weightings derived from a distance decay function. The result is a matrix of cells containing population estimates, preserving the total population as represented by the centroid data. It is designed to preserve the population of each settlement and the unpopulated regions.

PROS	CONS
Preserves population of each settlement and the unpopulated regions Within zone distribution is revealed Good approximation of residential geography at national level Can be applied to total population or any other data at the centroid location. Can be used to model change over time and show up patterns Overcomes the problems of changing administrative units.	Accuracy of the centroid placement and their weighting is hard to achieve; method works best if the centroid location is a good indictor of the centre of mass. Distance decay function used may have an influence on the results and perhaps different models of urbanisation will give different results Poor at urban centres because small zones may be sub-pixel sized. The estimated area is not independent of the choice of the cell size

Spline interpolation | TECHNIQUE (11)

PROCESS	(E, G + E)
POTENTIAL APPLICATION	Mapping of climate or air pollution data from point measurements (monitoring stations)

DESCRIPTION

Splines have been referred to as mathematical equivalents of the flexible ruler. They provide a means of building locally fitted polynomials to give a continuous area through a set of spatially distributed points. Local surfaces are joined at 'knots' or 'break points'. The method may be applied either as an exact interpolator (i.e. which forces the surface through the data points) or for smoothing. The degree of smoothing can be controlled by the user.

Different spline functions can used. The Laplacian spline method is one of the most widely used and best-developed, and permits smoothing of the data depending on the purpose of the analysis and error minimising criteria defined by the user.

PROS	CONS
Simple methods are available in proprietary GIS. Allows locally-optimised surfaces to be built into a continuous surface Retains local detail in the surface (unlike trend surface analysis) Provides error estimates for the data points Less affected by clustering and distribution of data points than some methods Surfaces can be adjusted locally without altering surface elsewhere	More sophisticated spline techniques not readily available in proprietary GIS Computationally impracticable with large datasets May be sensitive to method chosen to interleave local surfaces May produces local anomalies (e.g. at points of interleaving)

TECHNIQUE (12)	Trend surface analysis

PROCESS	(E, G + E)
POTENTIAL APPLICATION	Mapping of climate or air pollution from point measurements (monitoring sites)
DESCRIPTION Trend Surface Analysis is an extension of regression analysis to three dimensions, in which the two independent variables are spatial coordinates. It is a global interpolator, which fits a smoothed, mathematically defined area through the Z values. Commonly, relatively low order surfaces, such as linear, quadratic or cubic surfaces are fitted, using least squares regression techniques. Higher order surfaces can be applied, but these become difficult to interpret. Error estimates are provided at the data locations, and the r^2 value provides a measure of the goodness of fit of the surface. The statistical significance of the fitted surface can be tested by the F test. Differences in fit between different surfaces (e.g. between linear and quadratic surfaces) may also be tested using the F test	
PROS Available in many GIS and statistical packages Well-established and widely used techniques Computationally simple Provides error estimates and goodness of fit statistics for the fitted surface Useful for defining outliers	**CONS** May produce over-smoothed surfaces, especially in areas of complex geography Only works well with datasets which show clear spatial trends May suffer from severe edge effect at the limits of the data coverage Complex surfaces may be difficult to interpret and may be sensitive to extreme data values Assumes a random distribution of residuals around the trend — a condition which is often not satisfied May be sensitive to the spatial distribution of sample points

3. Comparison of CORINE land cover data with IACS data in Belgium and Italy and with land use in Slovenia

Eric Willems*, Jacques Delincé*, Amandine de le Court**, Paul Campling**, Bruno Buffaria* (*DG AGRI, **DG AGRI/GIS, *** Eurostat/F2)**

3.1. Introduction

In the first common publication From land cover to landscape diversity in the European Union', a DG AGRI article proposed a simple and easy way to compute land cover diversity index. This index was based on the aggregation at regional level of 18 classes of the CORINE inventory in cells of 3 km x 3 km.

In this article, we analyse the potential of using spatial information from the Integrated Administration and Control System (IACS) datasets for land cover analysis. In Belgium, a cartographic interpretation of the IACS inventory is carried out, analysed and contrasted with the CORINE land cover and the MARS *Agricultural Regions map*. The Agricultural Landscape Diversity Index and the Shannon index, based on IACS data, are calculated at regional level and compared. Both indices are then confronted with the CORINE land cover diversity index.

In Perugia Province, Italy, spatial analysis of land use typology is done by comparing CORINE agricultural classes with IACS classes in the area.And finally in Slovenia, results of the land use census and the CORINE inventory are compared at commune level.

Alongside the CAP reform of 1992 and the move towards a system of direct payments to farmers, rather than classical market interventions the Integrated Administration and Control System (IACS) was introduced on the basis of Council Regulation No 3508/92. The main aim was to have effective instruments to deal with the increased risks of irregularities and fraud inherent in the increased number of direct payments.

The IACS helped Member States to improve the way the farmers' aid applications are processed and to strengthen the checks of these payments that represent around 50 % of the overall CAP expenditure. IACS basically consists of alphanumeric identification systems for animals and arable land, which make it possible to run IT-based cross-checks avoiding, for example, that the same piece of land and/or animal is applied several times. Moreover, IACS provides for a certain percentage of on-the-spot controls of the aid applications (by physical visits by inspectors and/or, in the case of arable land, via remote-sensing methods), and for the selection methods based on risk analysis that need to be applied by the Member States.

The sectors covered by IACS include among others arable crops and set-aside land, forage-areas, rice, grain legumes and animal premia. However for this article, we were only interested in the land use (agricultural parcels). All parcels of holdings receiving subsidies have to be monitored by the IACS system. It is, therefore, an extremely detailed source of information about subsidised agricultural land use.

The Commission proposed the compulsory use of computerised geographical information system (GIS) techniques for the creation of spatial geographical (parcel) identification systems including also the compulsory use of orthophotography. In the present article, we have carried out comparisons for two Member States (Belgium and Italy) which have already adopted GIS technology for managing their IACS systems.

3.2. Building Land Cover Diversity Index: Data Issue

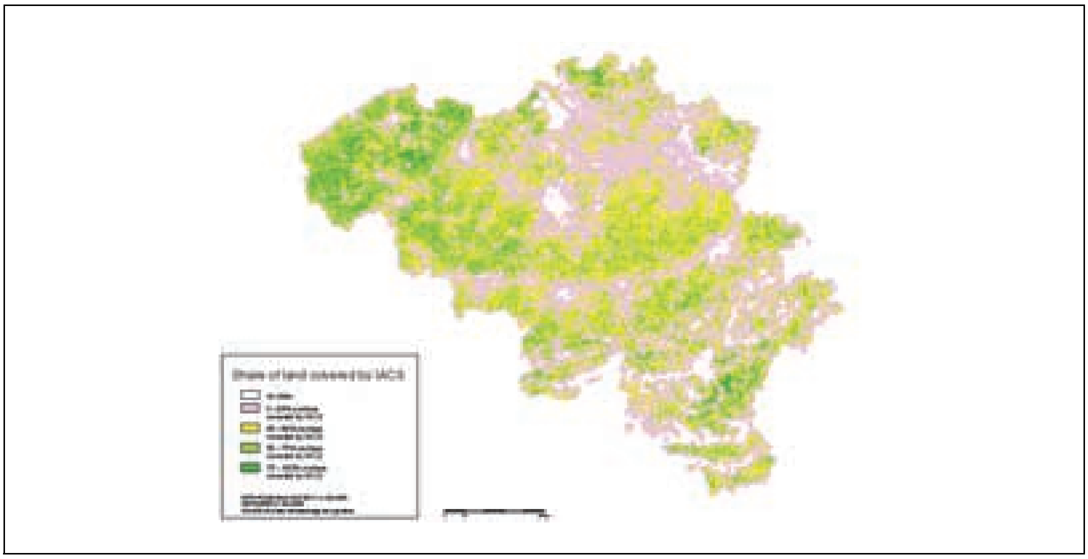
3.2.1. Ex lanation of the Belgian IACS treatments of data

The Belgian IACS database is a database at parcel level. However, we used data transferred to us by the Belgian control service that had been aggregated to 1 x 1 km^2 grid cells.

In each 1 x 1 km^2 grid cell, the share of area devoted to each agricultural use was given, as well as the non-agricultural area (14). The grid cells were grouped to obtain grid cells 3 x 3 km^2. For each square, the different areas were summed up. The data used refers to year 1999. (15)

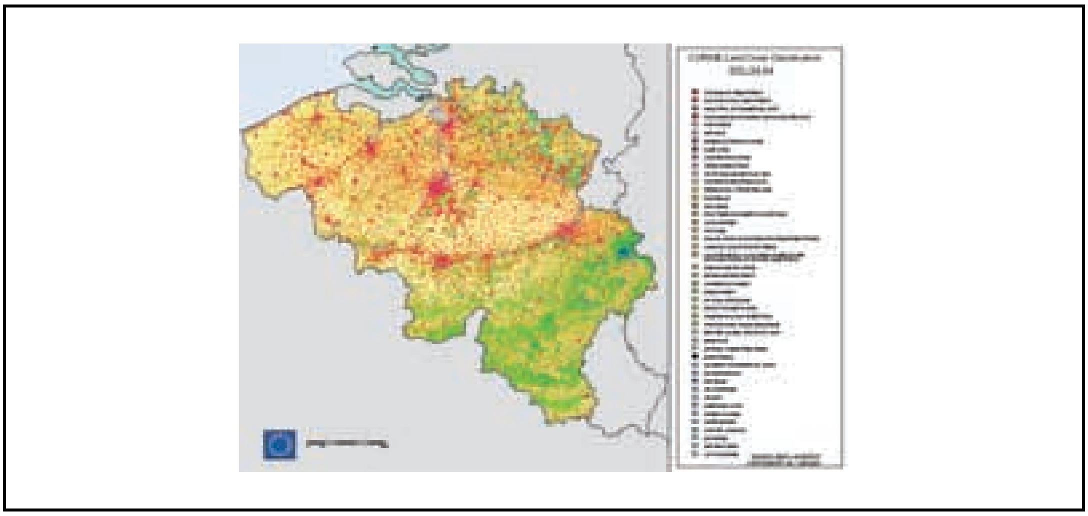
3.2.2. Share of land covered by IACS
IACS only overlays agricultural areas within the territory of Belgium Map 1 shows the area of land covered by the of aggregated IACS grid cells for Belgium, representing 36.6 % of the country.

Map 1. **Share of land covered by IACS**



Comparing the IACS map with the CORINE LC map (map 2) or with the *Agricultural Regions of Belgium: land use* map (16) (map 3) explains the variation in IACS overlay.

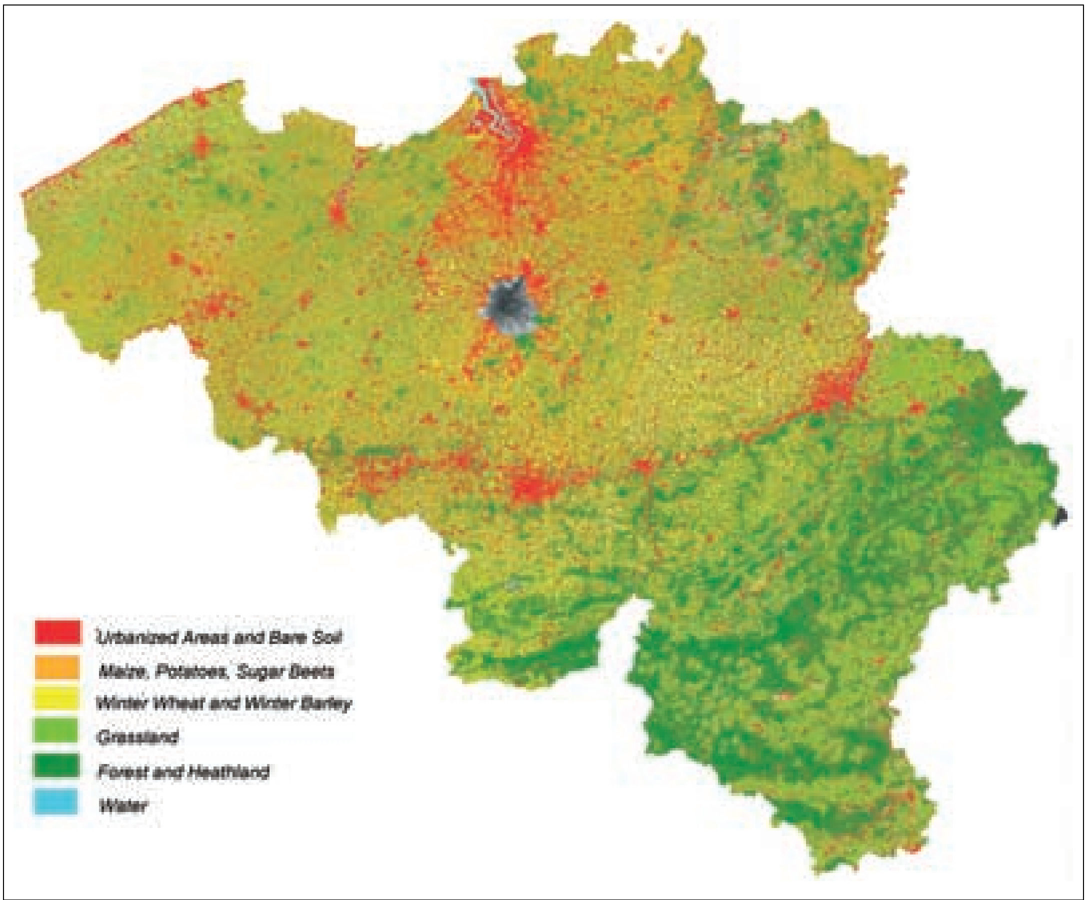
Map 2. **Belgium CORINE land cover**



(14) See Table 1 for the list of land use classes
(15) In IACS 1997 and 1998 are also available. This would allow us to carry out a temporal analysis, however this analysis is not presented in this article.
(16) Agricultural regions of Belgium: land use MARS project 1992. This map is part of the agricultural inventory of Belgium by remote sensing. This project is financed by the Belgian Science Policy Office (in the frame of the national research programme in remote sensing TELSAT of SSTC), the Belgian Ministry of Agriculture and the C.E.C. Image classification: Laboratory of Remote Sensing and Forest Management — University of Gent. Spatio-cartography: Laboratory SURFACES — University of Liège.

The 0 % IACS coverage (white) are mainly overlapping important urban areas: Brussels, Liège, Antwerp, Gent and Charleroi, in red on map 2. and 3. But also 0 % IACS is found in the forest and natural vegetation areas of southern Belgium, which are non-agricultural areas, and thus are not monitored by IACS. The white parts of map 1 fit perfectly with the forest and wetland regions (dark green) of maps 2. and 3. Urbanised areas alongside the 2 main Belgian rivers, the Schelde and the Meuse (in red in maps 2. and 3.) are, not surprisingly, lightly covered by IACS (class 0-25 %).

Agricultural regions of Belgium — land use Map 3.



3.2.3. Com uting the Agricultural Land Use Diversity Index using IACS information

3.2.3.1. Method of calculation

The IACS nomenclature comprises 62 different land cover types (see table 1.). As our interest is to shed light on the relationship between agriculture and landscape, we aggregated classes number and reduced it to 15.

We regrouped classes with similar landscape impact': green coverage whole year or part of the year, same or different heights, same or different colours, etc. For example, we regrouped winter wheat, winter barley and winter rye (group 2) or peas, beans and sweet lupins (group 6). Table 1 shows IACS classes aggregated into different groups to which different colours have been attributed.

Table 1. **IACS classes**

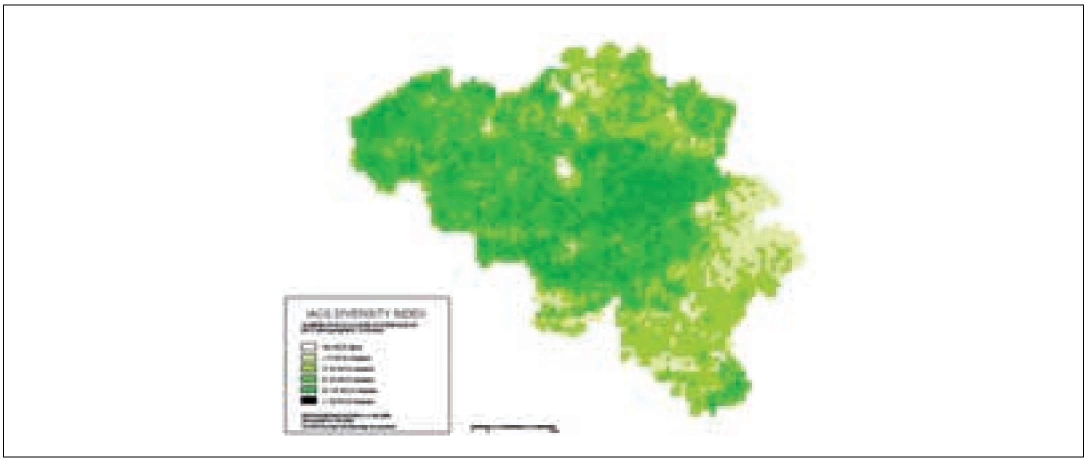
IACS class	groups	name	IACS class	groups	name
1	groupe 1	Silage maize	32	groupe 7	Graminae
2	groupe 1	Maize	33	groupe 6	Leguminious plant
3	groupe 2	Winter wheat	34	groupe 7	Mixture of graminae & leguminous
4	groupe 3	Spring wheat	35	groupe 7	Other coverage
5	groupe 2	Winter barley	36	groupe 7	Other coverage with min 20% certified seeds
6	groupe 3	Spring barley	37	groupe 4	Non food winter rapeseed
7	groupe 2	Winter rye	38	groupe 4	Non food spring rapeseed
8	groupe 3	Spring rye	39	groupe 5	Non food lineseed
9	groupe 3	Oats	40	groupe 8	Other annual non food crop
10	groupe 3	Triticale	41	groupe 8	Other pluriannual non food crop
11	groupe 3	Epautre	42	groupe 9	Afforestation
12	groupe 3	Buckwheat	43	groupe 10	Environmental set aside
13	groupe 3	Sorgho, millet, alpiste	44	groupe 8	Potatoes
14	groupe 3	Other cereals	45	groupe 8	Sugarbeet
15	groupe 4	Winter rapeseed	46	groupe 5	Fiber flax
16	groupe 4	Spring rapeseed	47	groupe 5	Hemp
17	groupe 4	Sunflower	48	groupe 6	Peas (not harvested dry)
18	groupe 4	Soy bean	49	groupe 6	Beans (not harvested dry)
19	groupe 5	Lineseeds	50	groupe 6	French bean
20	groupe 6	Peas (harvested dry)	51	groupe 6	Market garden
21	groupe 6	Beans (harvested dry)	52	groupe 6	Non edible horticultural crop
22	groupe 6	Sweet lupins	53	groupe 11	Fruits
23	groupe 7	Permanent grassland	54	groupe 8	Endive
24	groupe 7	Temporary grassland	55	groupe 12	Other crops
25	groupe 8	Fodder beets	56	groupe 12	Tobacco
26	groupe 8	Clover	57	groupe 12	Hops
27	groupe 8	Lucerne	58	groupe 12	Indetermined
28	groupe 8	Fodder rape	59	groupe 13	Broad-leaved trees
29	groupe 8	Fodder carrot	60	groupe 13	Coniferous trees
30	groupe 8	Other fodder crops	61	groupe 14	Water
31	groupe 10	Green cover	62	groupe 15	Buildings

The Agricultural Landcover Diversity Index was computed in two steps ([17]).

a) The base unit was a grid cell of 3 x 3 km² within which the different land cover groups were counted. The grid cell square of 3 x 3 km² was determined empirically; corresponding somehow to the human field of vision (what can be seen with the naked eye). The results are presented for whole Belgium on map 4.

The darker the square, the higher the IACS diversity of the agricultural land cover. It has to be recalled that results depend among other variables from the % of the IACS coverage (See map 1). Southern Belgium with a great deal forests and natural vegetation areas, which are not recorded in IACS, showed a high preponderance of light green grid cells.

Map 4. **IACS diversity index**



(17) See also 'Landscape and land cover diversity index' in *From land cover to landscape diversity index in the European Union* DG AGRI-Eurostat-JRC Ispra-EEA Report

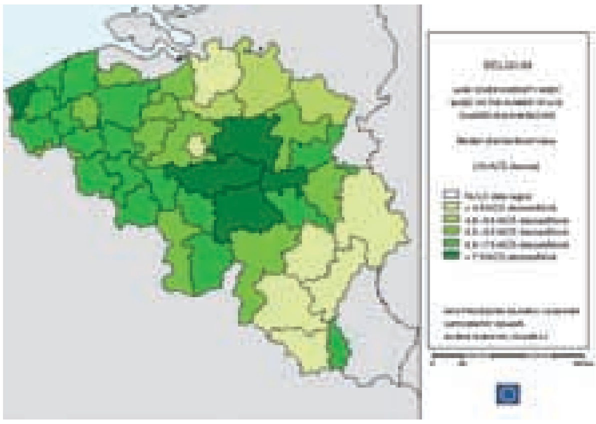
b) The number of land cover group classes, within each grid cell of 3 x 3 km², was summed up for each NUTS 3 region (Belgian arrondissements') The intrapolated median agricultural land diversity index class ([18]) was then assigned to each NUTS 3 region of Belgium, as the measure of agricultural diversity.

c) As mentioned before, the average IACS coverage for Belgium was 36.6 % with a standard deviation of 12.2 %, the lowest coverage being for Brussels with 6.5 % whereas the highest was located in Dijksmuide (North west). More than 60 % of the regions had IACS coverage that ranged between 30 and 50 %.

3.2.3.2. Results

The median for whole Belgium amounts for 6.1 and the standard deviation of the median to 1.4 or 23 %.

Median of IACS classes in NUTS 3 regions Map 5.



The Brussels region had the lowest median (1.7) and Waremme region in the Province of Liege had the highest median (8.5).

The regional distribution was as follows:

5 regions had a median above 7.5, of which four were located south-east of Brussels, in regions with large areas of agricultural land, and which one was located in the extreme North West of Belgium.

17 regions had a median ranged between 6.5 and 7.5.

11 regions had a median ranged between 5.5 and 6.5.

3 regions had a median ranged between 4.5 and 5.5.

7 regions had a median lower than 4.5 mostly in the South where the share of agricultural land is low, land cover is mostly forestry and natural vegetation areas.

3.2.4. Shannon diversity indicator in Belgium from IACS data

3.2.4.1. Computing Shannon index using IACS data

As for the agricultural land cover diversity index, the base unit for the regional Shannon diversity indicator was a grid cell of 3 x 3 km².

The Shannon indicator was first computed on each grid cell:

$$Sh\{s\} = -\sum_{c=1}^{16} p_i \ \ln(p_i)$$

(18) Preferred choice has been given to mediane, instead of averagefor statistical reasons. The distribution of groups percentages is not gaussian, and thus median is a more robust estimator. In order to have a finer picture, mediane was intrapolated.
The median of a set of measurements is the middle value when the measurements are arranged from smallest to largest. From this median definition, we derived the notion of intrapolated median. The intrapolated median is calculated as: (median-1) + (50 %- (cumulative frequencies in % at median –1 level)/ (cumulative frequencies in % at median level — cumulative frequencies in % at median — 1 level)). This illustrates better the diversity of a region. For example, a median of 6 may refer to two different situations and can cover 19 to 51 % of the population as well as 41 to 70 %, which is quite different. The intrapolated medians for the same distributions would be for the first 6.97 and 6.30 for the second.

where p_i is the proportion of each IACS class on the total area of the square.

Areas not registered in IACS wereconsidered as the 16th class. When a grid cell was on the border between two regions the whole grid cell was considered.

The Shannon indicator for a region was computed as a weighted average of the Shannon indicator for the squares. The weight is the area $A(s \cap r)$ of the squares in the region r:

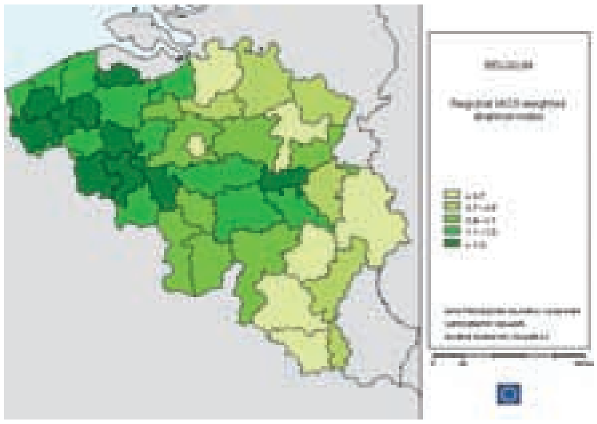
$$Sh\{r\} = \frac{\sum_s Sh\{s\} \times A(s \cap r)}{\sum_s A(s \cap r)}$$

The average number of classes per square in each region was a weighted average, using the same weights.

3.2.4.2. Results

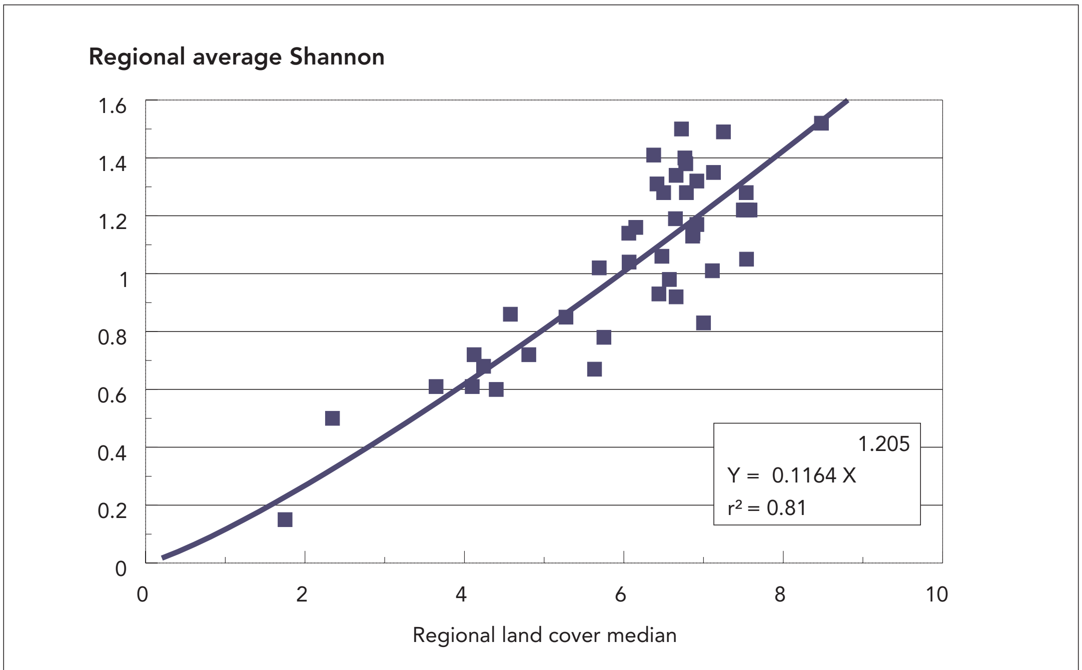
The average Shannon Index for Belgium was 1.05 with a standard deviation of 0.31.

Map 6. **Regional average Shannon**



As for agricultural land cover diversity index, Brussels region had the lowest Shannon Index (0.15) and Waremme the highest Shannon Index (1.52).

There was a strong correlation on a national basis between the computed land cover diversity index and the Shannon Index. The correlation between the regional average



Shannon Index and the regional land cover median ($r^2 = 0.81$) is shown in the graph.

3.2.5. Com arison of IACS and CORINE land cover diversity indexes

The land cover diversity Index was also computed using CORINE LC database. The number of CORINE LC classes was reduced from 44 to 22. We used the finest differentiation for agricultural classes (code 2) and shrub and/or herbaceous vegetation associations (code 3.2), the intermediate level for the code 4, and no differentiation for artificial surfaces (code 1) and for water bodies (code 5). The number of land cover classes per 3 * 3 km^2 grid cell was computed for each NUTS 3 region. The median number of land cover classes was then computed for each region.

The result is presented in map 7.

The median for Belgium was 4.8 and the standard deviation of the median was 0.65

What can be drawn from a comparison of the three indices?

A visual comparison of different maps does not evoke any evident relationship between the indices. Correlation between indices is not or very slightly significant. Moreover no conclusion may be derived about photo interpretation (bad or good quality, importance of subjective components) or possible bias originated by CORINE LC inventory.

The two different sources of information are only marginally equivalent/overlapping. IACS data concern exclusively agricultural areas whereas CORINE LC regards the whole land cover (agricultural and not). Furthermore, the two inventories are of different scales: polygons of minimum 25 ha for the CORINE LC, and in the order of meters for the IACS data.

Map 7.

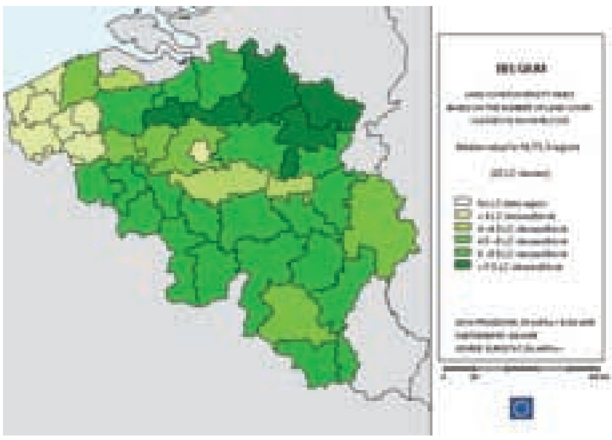


Table 2. **CORINE land cover nomenclature**

Level 1	Level 2	Level 3
1. Artificial surfaces	1.1. Urban fabric	1.1.1. Continuous urban fabric
		1.1.2. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units
		1.2.2. Road and rail networks and associated land
		1.2.3. Port areas
		1.2.4. Airports
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites
		1.3.2. Dump sites
		1.3.3. Construction sites
	1.4. Artificial non-agricultural vegetated areas	1.4.1. Green urban areas
		1.4.2. Sport and leisures facilities
2. Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land
		2.1.2. Permanently irrigated land
		2.1.3. Rice fields
	2.2. Permanent crops	2.2.1. Vineyards
		2.2.2. Fruit trees and berry plantations
		2.2.3. Olive groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous agricultural areas	2.4.1. Annual crops associated with permanent crops
		2.4.2. Complex cultivation patterns
		2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation
		2.4.4. Agro-forestry areas
3. Forests and semi-natural areas	3.1. Forests	3.1.1. Broad-leaved forest
		3.1.2. Coniferous forest
		3.1.3. Mixed forest
	3.2. Shrub and/or herbaceous vegetation associations	3.2.1. Natural grassland
		3.2.2. Moors and heathland
		3.2.3. Sclerophyllous vegetation
		3.2.4. Transitional woodland shrub
	3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes and sand plains
		3.3.2. Bare rock
		3.3.3. Sparsely vegetated areas
		3.3.4. Burnt areas
		3.3.5. Glaciers and perpetual snow
4. Wetlands	4.1. Inland wetlands	4.1.1. Inland marshes
		4.1.2. Peatbogs
	4.2. Coastal wetlands	4.2.1. Salt marshes
		4.2.2. Salines
		4.2.3. Intertidal flats
5. Water bodies	5.1. Inland waters	5.1.1. Water courses
		5.1.2. Water bodies
	5.2. Marine waters	5.2.1. Coastal lagoons
		5.2.2. Estuaries
		5.2.3. Sea and oceans

3.3. Building Land Cover Diversity Index: Classes Issue

The scale at which data was collected was not the only parameter that may have influenced the index calculation. The typology of land cover could also have an effect on the index. In order to shed light on this issue and in particular to evaluate the reliability and accuracy of CORINE LC data we carried out a comparative analysis with the IACS data for Perugia Province in the centre of Italy.

3.3.1. Treatment of Italian IACS data

The Perugia IACS dataset was received as a shapefile coverage. The Perugia CORINE land cover grid (100 m resolution) dataset was extracted out of the GISCO databases. Both datasets were projected to the Transverse Mercator projection system. The IACS dataset was then converted from a shapefile to a grid coverage, using the extracted CORINE LC grid to set the analysis extent and grid resolution. This ensured that the gridded datasets were compatible. Spatial and cross tabulation analysis between both datasets was done in Arc View. Map query was a useful function for generating digital maps of data subsets, to illustrate the spatial coverage of particular classes and the extent to which similar classes from both datasets intersected.

Perugia province Map 8.

3.3.2. Localisation of the IACS arcels and breakdown of IACS and CORINE

Perugia province is 6335 km² and the IACS dataset covers an area of 1 810 km². Agriculture land covered by IACS therefore represents about 29 % of Perugia Province area. The breakdown of the IACS data in terms of percentage area is as follows: 76.5 % arable land, 11.2 % pastures, 6.8 % afforestation, 3.3 % permanent crops, 2.0 % buildings and storage facilities and 0.2 % inland water.

CORINE LC provides a total overview of land use in Perugia (Table 3), which indicates at level 1 that agricultural land use covers more than half of Perugia (51.6 %) and that the second most important category is Forest and semi-natural areas (43.3 %). There is a discrepancy therefore between the percentage of agricultural land identified by IACS (29 %) and by CORINE LC (52 %).

CORINE breakdown at level 1 Table 3.

	% of total	area (km²)
Artificial surfaces	2.9	181
Agriculture	51.6	3269
Forest and semi-natural areas	43.3	2745
Wetlands	0.1	6
Water bodies	2.0	128

The Agricultural category of CORINE LC broken down to three levels (Table 4) indicates that *non-irrigated arable land* (31 %) is the most important class, followed by *complex cultivation patterns* (7.0 %) and *land principally occupied by agriculture, with significant areas of natural vegetation* (6.6 %), and olive groves (3.6 %). Under the Forest and semi-natural areas category of CORINE LC, *broad-leaved forest* is by far the most important class (33.4 %), followed by *natural grassland* (3.7 %) and *transitional woodland shrub* (3.5 %).

Table 4. **CORINE land cover agricultural areas nomenclature and % of total area**

Level 1	Level 2		Level 3		% of total
2. Agricultural areas	2.1.	Arable land	2.1.1.	Non-irrigated arable land	30.9
			2.1.2.	Permanently irrigated land	0.0
			2.1.3.	Rice fields	0.0
	2.2.	Permanent crops	2.2.1.	Vineyards	0.6
			2.2.2.	Fruit trees and berry plantations	0.0
			2.2.3.	Olive groves	3.6
	2.3.	Pastures	2.3.1.	Pastures	2.5
	2.4.	Heterogeneous agricultural areas	2.4.1.	Annual crops associated with	0.3
				permanent crops	7.0
			2.4.2.	Complex cultivation patterns	6.6
			2.4.3.	Land principally occupied by agriculture, with significant areas of natural vegetation	0.0
			2.4.4.	Agro-forestry areas	

3.3.3. Agricultural areas in IACS and CORINE

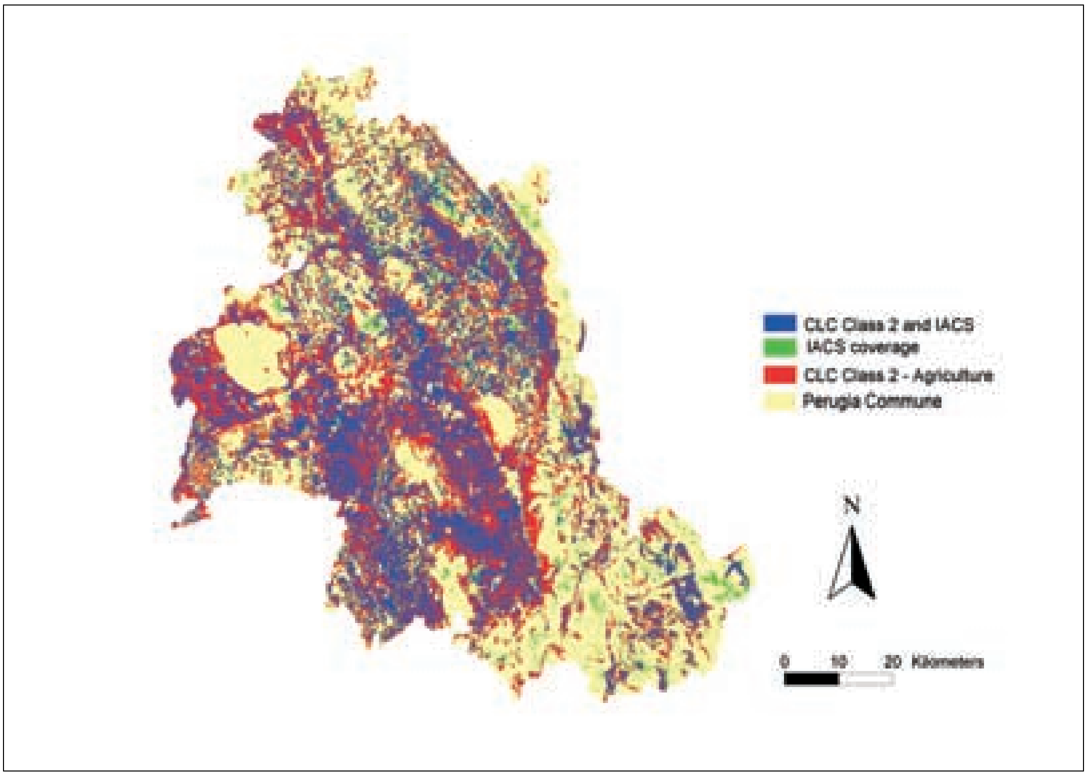
3.3.3.1. CORINE LC Agriculture and IACS

Map 9 shows the overlay of CORINE LC Agriculture and the entire IACS cover:

- In blue: CORINE LC *Agriculture* grids intersected by IACS grids (1451.6 km²)
- In red: CORINE LC *Agriculture* grids not intersected by IACS grids (1818 km²)
- In green: IACS grids not intersected by CORINE LC *Agriculture* grids (358 km²)

As indicated in section 3.2, the CORINE LC *Agricultural area* (3269 km²) is almost double the IACS area (1810 km²), and Map 9 indicates the overlay of the two datasets. The red shading, showing the CORINE LC *Agriculture* grids not intersected by IACS grids, are generally found at the margins of the blue shading, which represents the CORINE LC *Agriculture* grids intersected by IACS grids. The green shading, being IACS grids not intersected by CORINE LC *Agriculture* grids, are in close proximity to agricultural areas in the North West, but in the South East these are found in association with CORINE LC *Forests and semi-natural areas.*

Map 9. **Agricultural areas in IACS and CORINE**

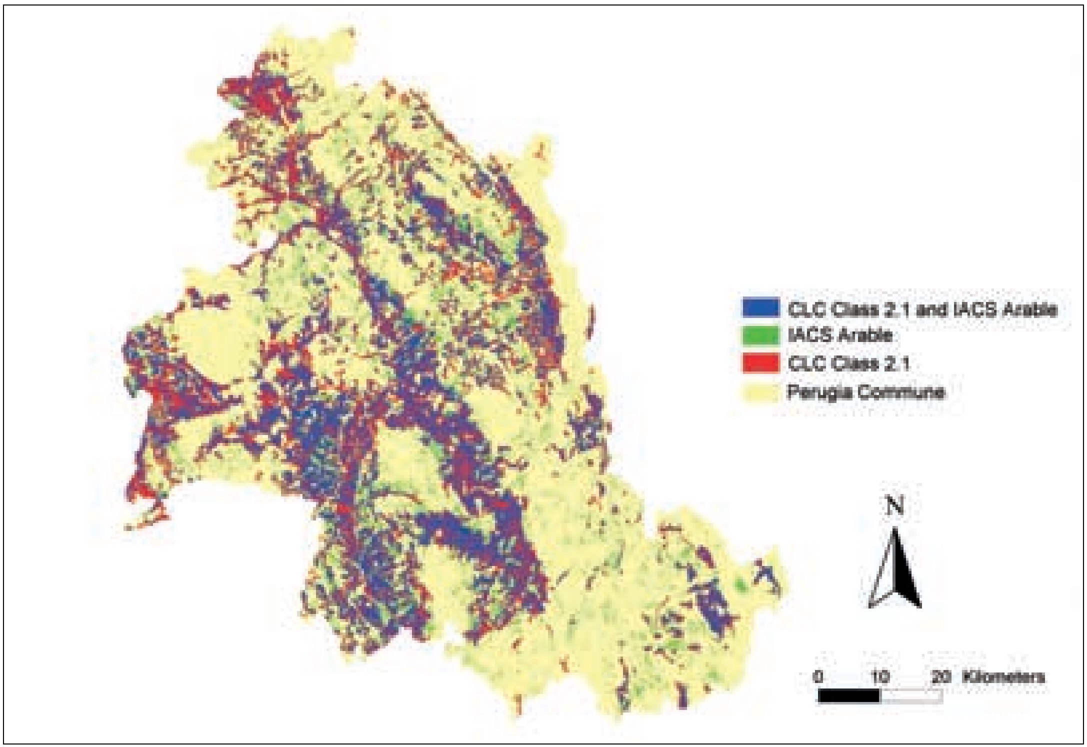


3.3.3.2. Arable land

Map 10 shows the overlay of CORINE LC *Arable Land* and IACS arable land classes:

- In blue: CORINE LC *Arable Land* grids intersected by IACS arable land grids (948 km²)
- In red: CORINE LC *Arable Land* grids not intersected by IACS arable land grids (1013 km²)
- In green: IACS *arable land* grids not intersected by CORINE LC *Arable Land* grids (472 km²)

Arable land in IACS and CORINE Map 10.



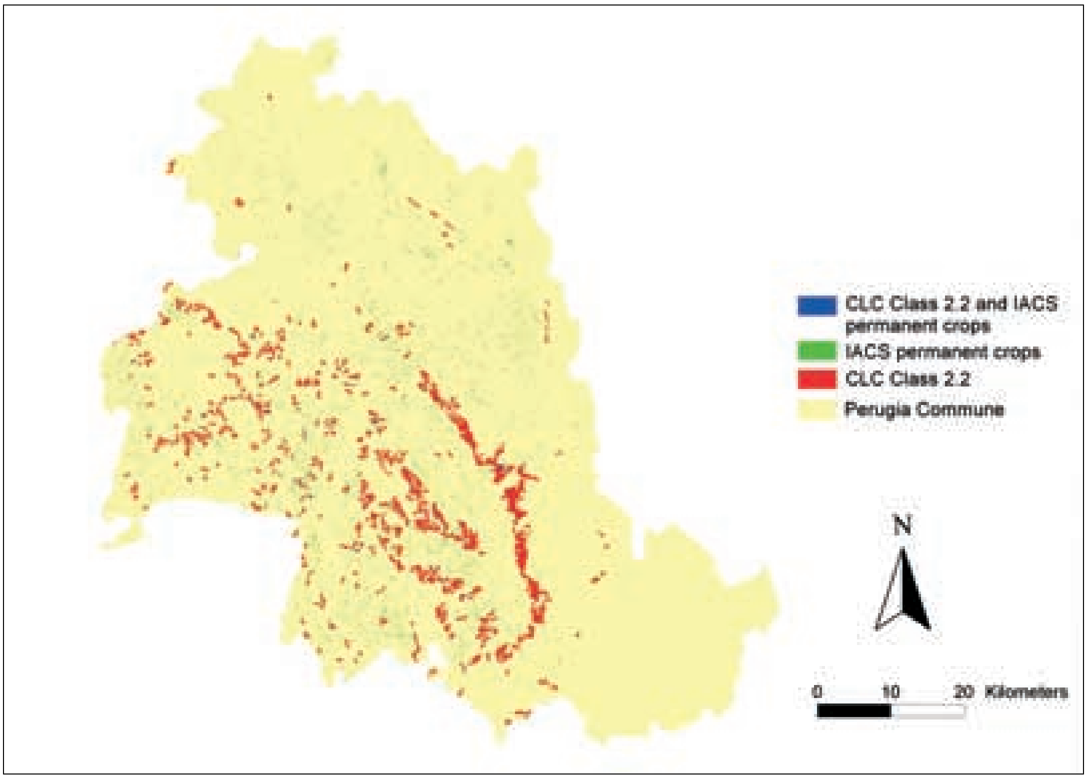
3.3.3.3. Permanent crops

Map 11 shows the overlay of CORINE LC *Permanent crops* and IACS permanent crops classes:

- In blue: CORINE LC *Permanent crops* grids intersected by IACS permanent crops grids (7 km²)
- In red: CORINE LC *Permanent crops* grids not intersected by IACS permanent crops grids (263 km²)
- In green: IACS permanent crops grids not intersected by CORINE LC *Permanent crops* grids (52 km²)

There is a very low agreement between the permanent crops grids of CORINE LC and IACS. An explanation for such a small area of intersection is can be the small size of the permanent crop plots in the IACS procedure comparison with the minimum size of the CORINE plots and also the low proportion of permanent crop holdings getting an area payment. In CORINE the permanent crops are situated in central and western areas, whereas in the IACS there is a substantial number of grids in the north eastern areas.

Map 11. **Permanent crops areas in IACS and CORINE**



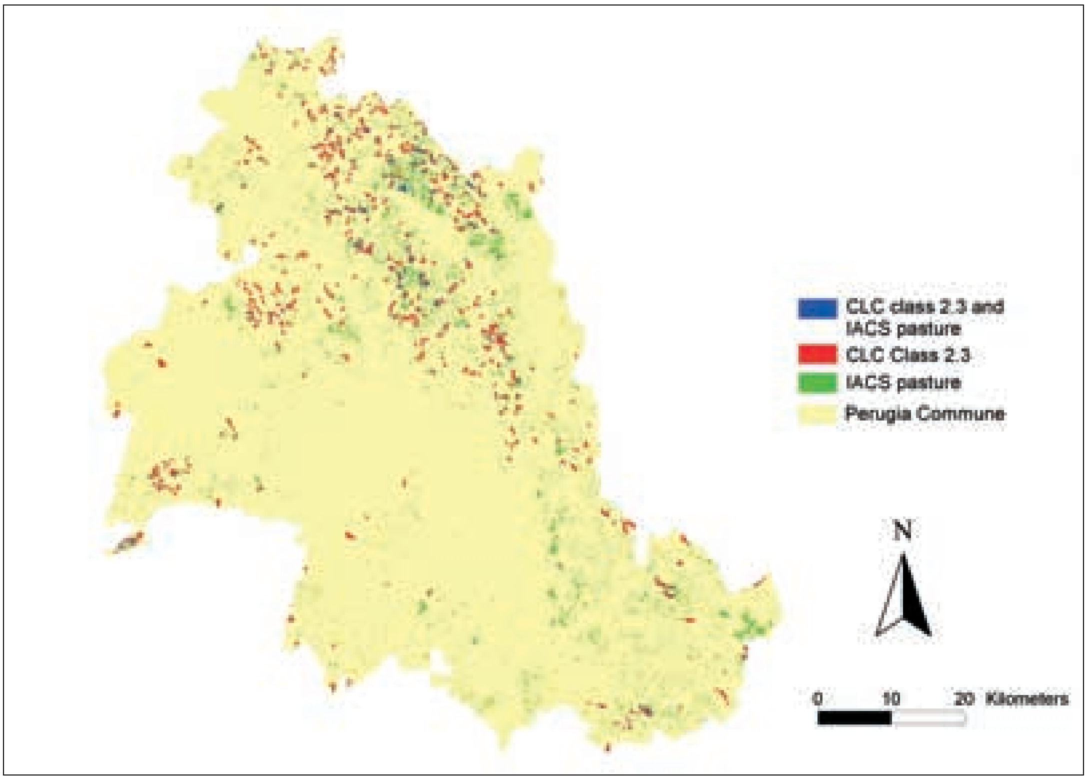
3.3.3.4. Pastures

Map 12 shows the overlay of CORINE LC *Pastures* and IACS pasture classes:

- In blue: CORINE LC *Pastures* grids intersected by IACS pasture grids (18 km^2)
- In red: CORINE LC *Pastures* grids not intersected by IACS pasture grids (141 km^2)
- In green: IACS pasture grids not intersected by CORINE LC *Pastures* grids (184 km^2)

Although there is a low number of corresponding pastures class grids, the pasture areas for both datasets are found mainly in the northern part of Perugia.

Map 12. **Pastures areas in IACS and CORINE**



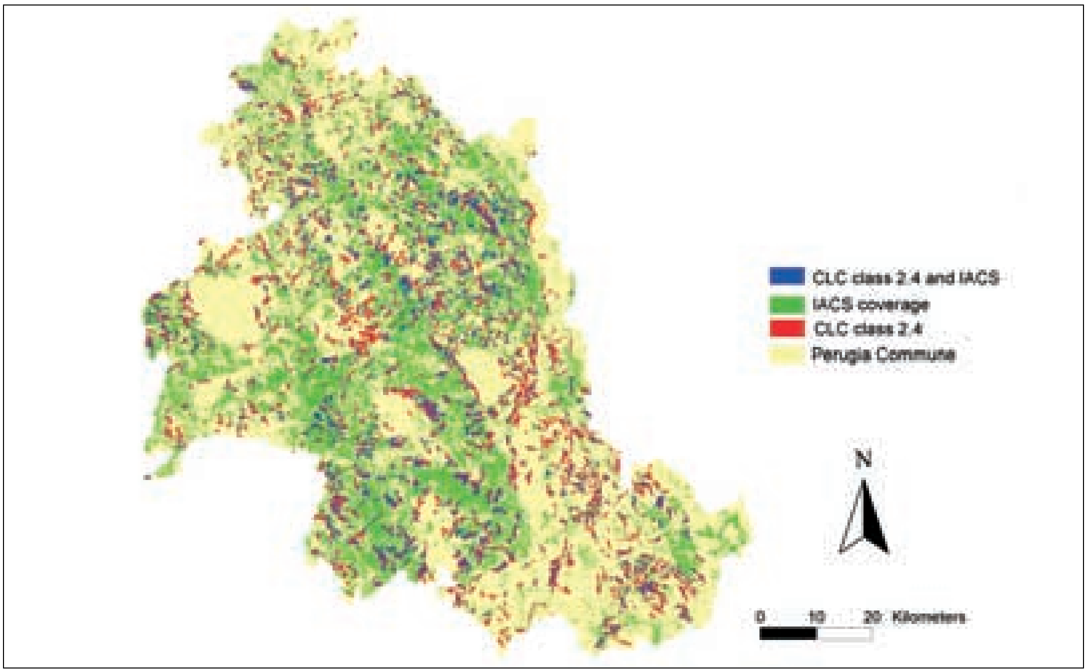
3.3.3.5. Heterogeneous agricultural areas

Map 13 shows the overlay of CORINE LC *Heterogeneous agricultural areas* and the total IACS coverage:

- In blue: CORINE LC *Heterogeneous agricultural areas* intersected by IACS grids (301 km²)
- In red: CORINE LC *Heterogeneous agricultural area* grids not intersected by IACS grids (579 km²)
- In green: IACS grids not intersected by CORINE LC *Heterogeneous agricultural area* grids (1509 km²)

The correspondence of CORINE LC *Heterogeneous agricultural areas* and IACS classes mainly occurs on the fringes of the IACS coverage, which fits the supposition that heterogeneity is more likely to occur on the margins of agricultural areas.

Heterogeneous agricultural areas in IACS and CORINE Map 13.



More indepth analysis was carried out of these heterogeneous agricultural areas, in particular areas mentioned as complex cultivation patterns (class 2.4.2) and *land principally occupied by agriculture, with significant areas of natural vegetation* (class 2.4.3).

Map 14 shows the overlay of CORINE LC *complex cultivation patterns* (class 2.4.2) and the total IACS coverage.

- In blue: CORINE LC *complex cultivation patterns* (class 2.4.2) intersected by IACS grids (168 km²)
- In red: CORINE LC *complex cultivation patterns* (class 2.4.2) grids not intersected by IACS grids (276 km²)
- In green: IACS grids not intersected by CORINE LC *complex cultivation patterns* (class 2.4.2) grids (1 642 km²)

The *complex cultivation patterns* are even more closely associated with the fringe areas of the IACS coverage than for the *Heterogeneous agricultural areas* (Map 13).

Map 14. **Complex cultivation patterns areas in IACS and CORINE**

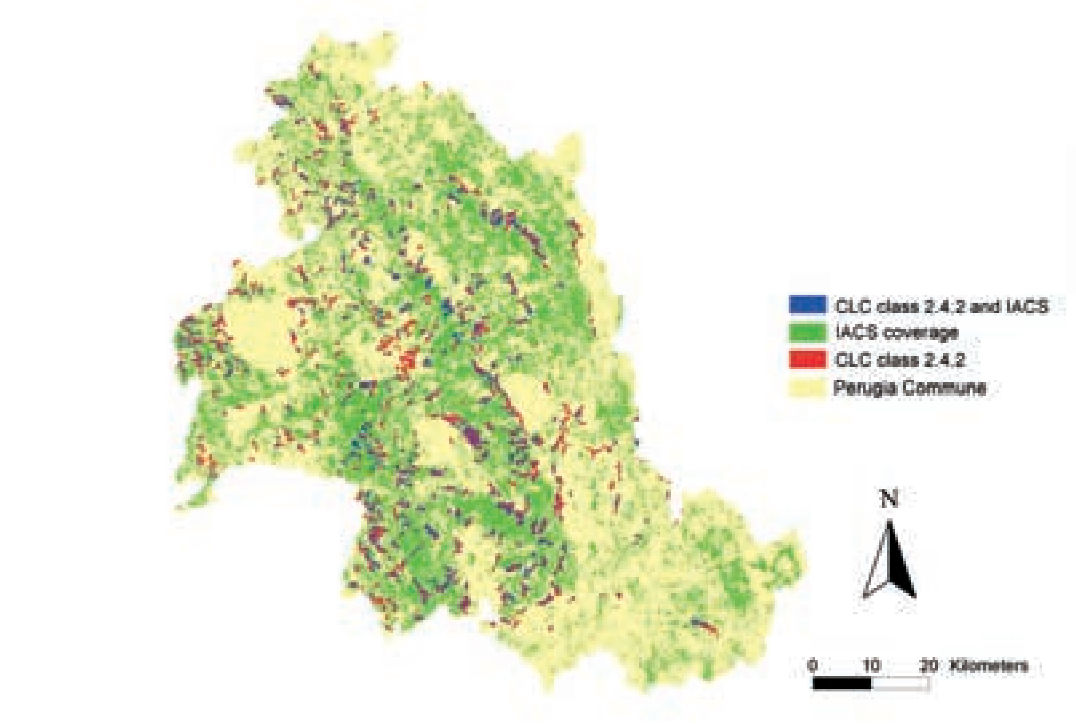


Table 5 compares the breakdown of IACS classes for the whole Perugia Province and for the area registered in CORINE as *complex cultivation patterns* (CORINE class 2.4.2).

Table 5. **Breakdown of IACS classes in total Perugia and in CORINE complex cultivation patterns**

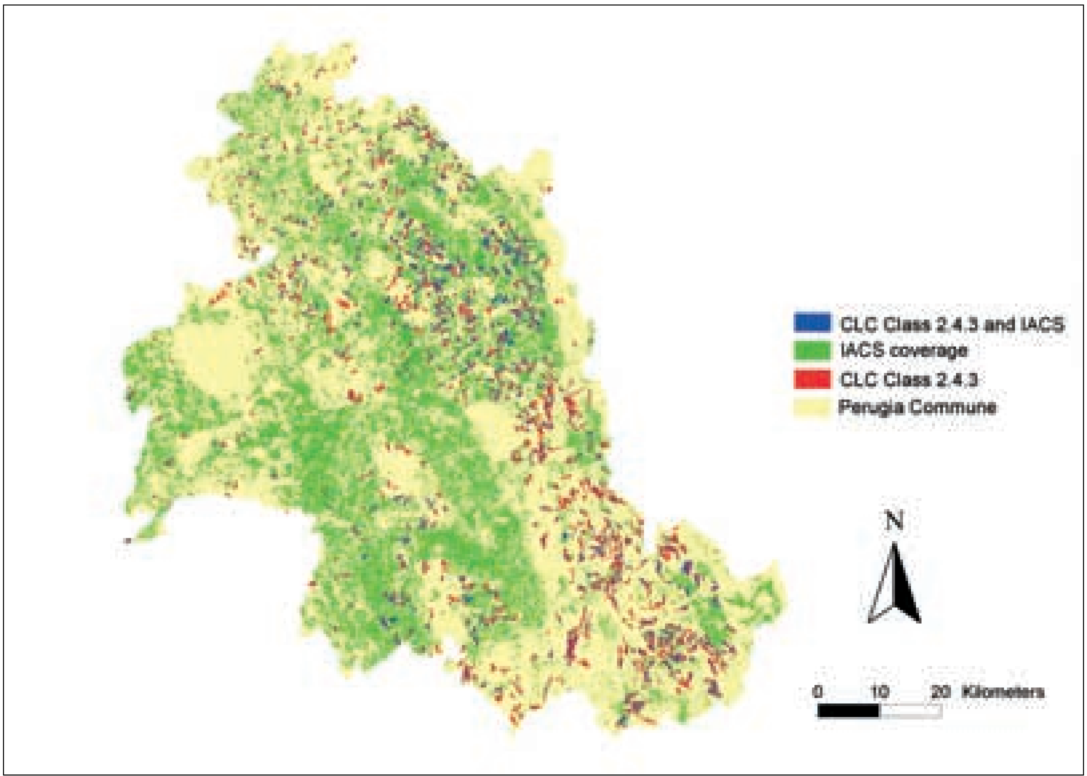
IACS classes	TOTAL	CORINE 2.4.2
Cereals (excl. maize & durum wheat)	30.1	33.6
Sunflower	14.2	18.0
Fodder crop	13.7	11.9
Pastures	11.2	3.3
Afforestation	6.8	4.4
Maize	5.2	3.7
Set-aside food	4.1	5.3
Durum wheat	2.9	2.9
Specialised orchards	2.5	5.3
Others	9.3	11.5

The main differences occurs for pastures with only 3.3 % for CORINE 2.4.2 class, whereas this value for the whole of Perugia is 11.2 %, and for specialised orchards, with 5.3 % for CORINE 2.4.2 class, and only 2.5 % for the whole of Perugia. Other minor differences occur between sunflower, afforestation and maize.

Map 15 shows the overlay of CORINE LC for *land principally occupied by agriculture, with significant areas of natural vegetation* (class 2.4.3) and the total IACS coverage.

- In blue: CORINE LC land principally occupied by agriculture, with significant areas of natural vegetation (class 2.4.3) intersected by IACS grids (128 km^2)
- In red: CORINE LC land principally occupied by agriculture, with significant areas of natural vegetation (class 2.4.3) grids not intersected by IACS grids (291 km^2)
- In green: IACS grids not intersected by CORINE LC land principally occupied by agriculture, with significant areas of natural vegetation (class 2.4.3) grids (1682 km^2)

Land principally occupied by agriculture with significant areas of natural vegetation in IACS and CORINE Map 15.



Closer inspection of the cross tabulation analysis reveals that 67 % of the intersection between CORINE 2.4.3 and the total IACS coverage is arable (o.w. 24 % cereals, 23 % fodder crops and 7 % sunflower), 20 % pastures and 10 % afforestation.

3.3.3.6. Summary

Table 6 summarises the result of the comparison of IACS and CORINE for Perugia. The differences between the two data sets as revealed by the cross tabulation analysis, overlay analysis and the resulting figures are due probably to a large degree from the different land use identification procedures adopted by the CORINE and IACS methodologies. CORINE land use interpreters draw polygons of more than 25 ha and attribute a single code that most represents the land cover classes within it. Whereas the IACS methodology is more at a parcel scale in the order of 100 square metres rather than hectares. In addition, IACS only registers holdings for which direct subsidy payments are made. Thus not all agricultural areas are recorded in IACS. For these reasons, it is normal that more agricultural areas are registered in CORINE LC than in IACS.

Share of IACS overlaying CORINE Table 6.

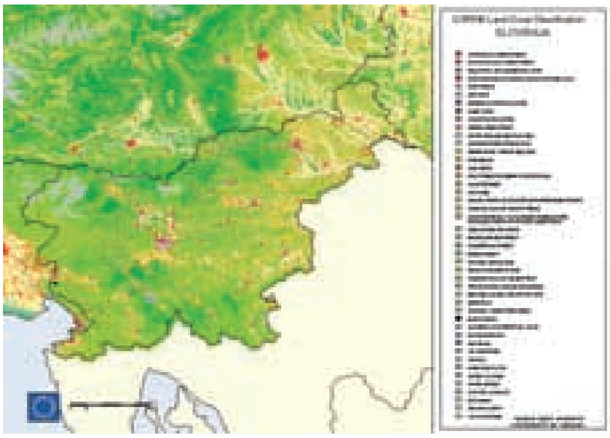
CORINE code	CORINE label	% of IACS covering CORINE	% of intersection (*)
2	Agricultural areas	44 %	44 %
2.1	Arable land	53 %	48 %
2.2	Permanent crops	18 %	2.2 %
2.3	Pastures	36 %	12 %
2.4	Heteregenous ag. areas	34 %	mainly arable
2.4.2	Complex cult. patterns	36 %	mainly arable
2.4.3	Land princ. occupied by agric.	33 %	mainly arable

(*) same classification in CORINE and IACS

3.4. Land Cover Diversity: The Structure Issue

After having tackled the issues of data and the one of classes typology we would like to test how far CORINE reproduces correctly land use structure by a direct comparison with the land use census, in our example the one carried out in Slovenia.

Map 16. **CORINE land cover — Slovenia**



3.4.1. Methodology

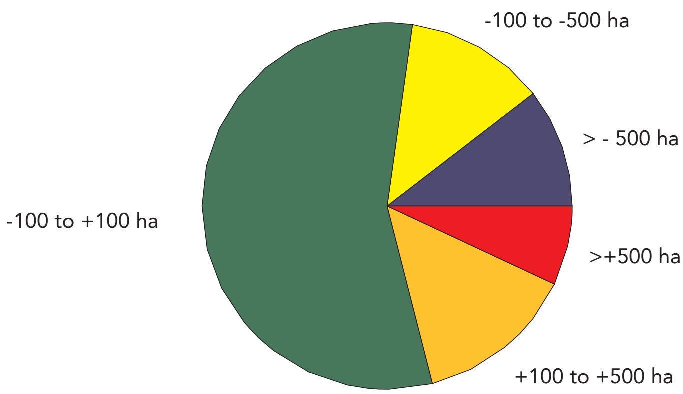
Land use classes utilised in this analysis are arable land, permanent crops, permanent grassland, wooded areas, inland waters and other uses. In order to carry out this comparison at communal level, it has been necessary to remove the CORINE data at regional level 4'[19]. The regional level 4' does not exist in the GISCO database, for that reason we needed to construct it by aggregating at a more precise scale. The result of this aggregation did not always fit with the limits of the communes. Thus, we first checked if the total area of a level 4' region was the same in the land use database and in our aggregation. If the result was divergent, the commune was not used to carry out the comparison.

3.4.2. Results

3.4.2.1. Total difference in regional 'level 4' area

For the whole of the Slovenian territory, the difference in total area calculated by both methods was negligible (3 hectares). However, for 10 of 62 communes the difference was higher than 500 hectares and thus these communes were not used for the comparison. Graph 3 shows the breakdown of the differences in regional level 4' total area. For more than 60 % of the communes (in green) the difference was lower than 100 hectares.

Graph 3. **Breakdown of the differences in regional 'level 4' area**



(19) See article 'Methodology' for explanation.

3.4.2.2. Differences in agricultural and non agricultural area

Table 6 shows the differences for the different land use classes for the 52 communes for which the difference in total area was less than 500 hectares. For these communes, the total difference in area was negligible, (ie.9 hectares).

DIFFERENCE in Land classes in ha (Land use — CORINE) Table 6.

	Average	SD	in % of the average LANDUSE class
Arable land	1 970	1 316	56 %
Permanent crops	563	521	63 %
Permanent grassland	7 075	6 300	75 %
Inland waters	-80	162	-266 %
Wooded areas	-2 362	4 338	-12 %
Other uses	-7 157	4 539	-286 %
Total	9		

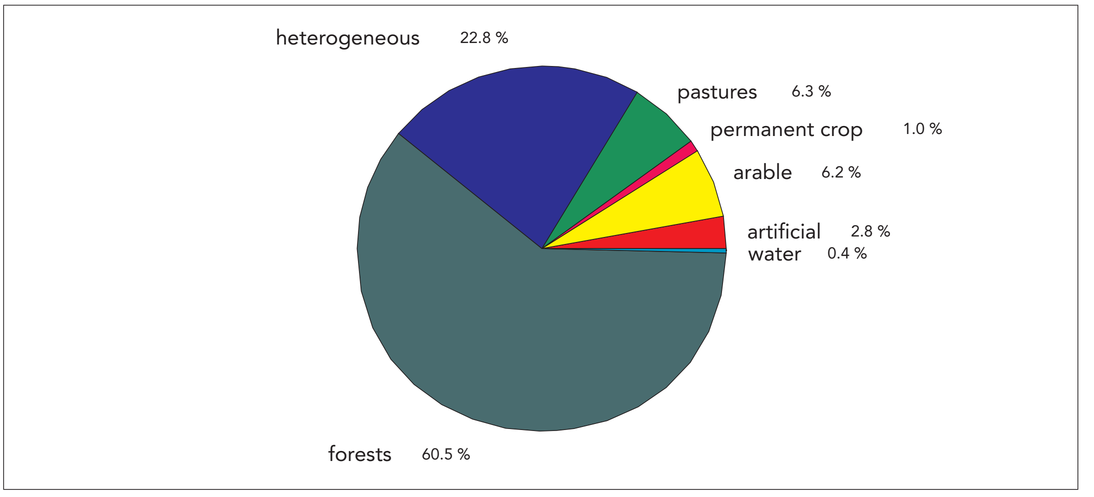
However, there were large discrepancies in the other classes with the exception of inland waters (- 80 ha on average). In all communes, CORINE underestimated agricultural areas and overestimated other uses. For wooded areas, it was less clear, 44 communes were overestimated by CORINE whereas 8 were underestimated. On average, permanent grassland was highly underestimated by CORINE by more than 7 000 hectares, whereas the other uses were overestimated by the same amount. CORINE underestimated arable land and permanent crops by, respectively, nearly 2 000 hectares and more than 500 hectares, wooded areas were overestimated by more than 2000 hectares.

More interesting was it to compare the communal average differences with the communal average land use. In this case, CORINE underestimated by more than 50 % the different agricultural classes, whereas it overestimated by more than 250 % other uses and inland waters. For the latter, the absolute difference was negligible. Only for wooded areas, the difference in percentage was small (overestimation of 12 %), this was certainly linked to the fact that more than 50 % of the whole Slovenian territory is wooded.

The results showed clearly the limits of the CORINE interpretation and the minimum polygon size of 25 hectares for a country like Slovenia where the area of agricultural holding was very low. In Slovenia, on average per holding, arable land was 2.2 hectares, permanent grassland was 3.3 hectares, orchards were 0.3 hectare and vineyard was 0.5 hectare. This certainly explained why in the CORINE database the class heterogeneous agricultural areas (code 2.4) was very important. On communal average, it represented 7 500 hectares whereas the 3 other agricultural classes together (2.1, 2.2 and 2.3) only 4 500 hectares.

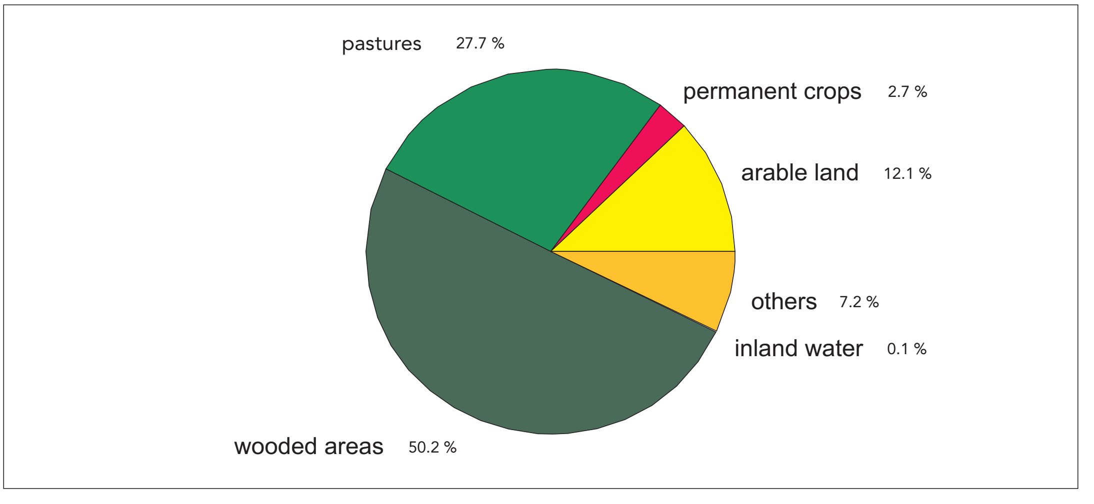
If the 4 agricultural CORINE classes were regrouped to level 1 of the CORINE classification, the average share of communal agricultural area was quite close to the land use census, 36.3 % instead of 42.5 %. Graph 4 illustrates the average communal breakdown of the main CORINE land cover classes.

Graph 4. **Breakdown of the main CORINE land cover classes**



Graph 5 illustrates the same average communal breakdown of the main land use classes.

Graph 5. **Breakdown of the main land use classes**



3.5.Conclusions

1. In the first part of the article, a simple agricultural land cover diversity index based on data coming from the Integrated Administration and Control System inventory was presented. The different land use classes are calculated in a square cell of 3 x 3 km² and then summed up at NUTS 3 level although other administrative or geographical levels would be possible. The index is the highest in agricultural areas where IACS coverage is the largest.

2. This index has been compared with the Shannon index calculated for the same 3 x 3 km² cells and a strong correlation ($r^2 = 0.81$) between both indices has been underlined. Both indices have also been compared with a land cover diversity index based on the CORINE inventory. No correlation between both indices has been observed: the latter cover the full range — not only agricultural — of land use. Moreover, in this study case the CORINE classes, permanent irrigated land, rice fields, vineyards and agro-forestry are not relevant.

3. The scale of CORINE and IACS is quite different: a minimum polygon size of 25 hectares for CORINE and 100 square metres for IACS. The IACS database is richer ([20]) and more detailed than CORINE but limited to agricultural land of holdings receiving direct payments. Therefore, the question is that of

overlapping and articulating' different layers of information.

4. In the second part of this paper, we compared for one region agricultural land classes given by administrative data (IACS system) and the land cover inventory (CORINE). More than 80 % of administrative data fits with the CORINE inventory. However, the CORINE inventory gives an estimate of agricultural area which is double the area derived by IACS. What is the explanation? On the one hand, IACS probably underestimates agriculture area by including only areas of holdings receiving direct payment and on the other hand, CORINE LC is based on the photo-interpretation of 25 hectares polygons, which include non-agriculture areas.

5. For the same region, we also analysed more indepth the CORINE heterogeneous agricultural areas (class 2.4). IACS overlay for this class is a little bit lower than for agricultural areas (one third). Comparison of the breakdown of IACS land uses in total Perugia and CORINE complex cultivation patterns shows only small differences except for pastures and specialised orchards. It means that this class is a real agricultural class and not a bin for photo-interpreters unable to interpret.

6. Finally, we confronted the land use census with the CORINE inventory. At communal level for Slovenia, large discrepancies were found between the land use census and the CORINE LC inventory. In all communes, CORINE LC underestimated agricultural areas. However, the average agricultural parcel area in Slovenia is very small (2.2 hectares for arable land) and fragmented. This has to be compared again with the minimum 25 hectares polygon size of CORINE. This is also the reason why the CORINE heterogeneous agricultural area class is very important.

7. The authors would like to thank the Belgian Ministry of Agriculture (Centre de Recherche Agronomique) and the Italian Control Agency (AGEA) for having provide some of their IACS data as well as the University of Ljubljana for the data of the land use census.

(20) The IACS database contains information of different years and would thus also allow carrying out temporal analyses.

References

Agenzia Regionale per la Protezione Ambientale della Toscana (ARPAT), *Rapporto sullo stato dell'ambiente in Toscana,* Rapporto 2000.

Association internationale ruralité, environment, dévelopement, *Evolution Agricole et paysage rurale,* ed. Patrice Collignon, 1992.

Boggia A., Pennachi F., *Sviluppo agricolo sostenibile del bacino del lago Trasimeno,* ISBN 88-87652-02-3, Perugia, 1999.

Brossard T., Wieber J.-C., *Le paysage: trois définitions, un mode d'analyse et de cartographie*, Espace géographique, 1, 1984, pp. 5-12.

Burel F., Baudry J., *Ecologie du paysage,* Tec & Doc, Paris, 1999.

Dion, R., Essai sur la formation du paysage rural français, Arrault, Tours, 1934.

European Commission, *CORINE land cover. Technical guide*, EUR 12585 EN, OPOCE, Luxembourg, 1994.

European Environment Agency, *Spatial and ecological assessment of the TEN: Demonstration of indicators and GIS methods*, ISBN 92-9167-112-6, Copenhagen, 1998.

Hoyiat M., Simon L. et alii, *Paysages: évolution et dynamique*, Programme SHS du CNRS, ENS de Fontenay-Saint-Cloud, 1996.

Klopatek J. M., Gardner R. H., Ed. *Landscape ecological analysis*, Springer, New York, 1999.

Larcena D., Puech D., *La dynamique des paysages: de l'approche spatiale par télédétection à la mise en place d'indicateurs comptables,* Revue de l'économie méridionale, Montpellier, 1998, N°183.

Lassus B., *The landscape approach*, Philadelphia, AAUP, 1998.

Noirfalise, A., *Paysages: L'europe de la diversité,* CE Commission, 1988.

Perigord M., *Le paysage en France*, PUF, Paris, 1996.

Ray Haines-Young, David R. Gree and Stephen H. Cousins, *Landscape ecology and GIS*, 1993-92.

Rougerie G., Beroutchachvili N., *Géosystèmes et paysages, bilan et méthodes,* Armand Colin, Paris, 1991.

Sanderson J., Harris L.D., *Landscape Ecology A Top Down Approach,* Lewis Publishers, Boca Raton, Florida 1999.

Schama S., *Le paysage et la mémoire*, Seuil, Paris 1999.

Sereni E., *Storia del paesaggio agrario italiano*, Laterza, Bari, 1961.

Turner M. G., Garner R. H. (eds) *Quantitative methods in landscape ecology. The analysis and interpretation of landscape heterogeneity*, Springer, New York, 1990.

Voelm, F. & Glitzner, I., Institute for *Assessment of barrier effects on Red deer due to motorways in Austria,* Wildlife Biology and Game Management, 1998.

Walz, U., *Indicators of landscape structure,* Germany, Institute of Ecological and Spatial Development, 1998.

Westmacott & Worthington, *Agricultural landscapes: a third look,* UK Countryside Commission, 1997.

Wicherek S., Paysages agraires et Environnement, *Principes écologiques de gestion en Europe et au Canada*, CNRS Editions, Paris, 1999.

Willems E., Vandevoort C., Willekens A., Buffaria B., *Landscape and land cover diversity index,* in From land cover to landscape diversity in the European Union, Report of the European Commission, 2000.

4. Land cover in the context of the Natura 2000 network

Els De Roeck*, Danny Vandenbroucke*, Angelo Salsi, Steve Peedell*****
(*K.U.Leuven/Ground for GIS, **DG ENV — LIFE Unit, ***JRC)

4.1. General background of the Natura 2000 programme

Since the early 1970s, nature conservation has been an important part of the European Union's environmental policy. The rapid and sometimes alarming decline in species populations and the deterioration of natural habitats on the territory of the Union called for urgent measures at policy level. In 1979 the European Union implemented the Birds Directive on the conservation of wild birds (21), and in the early 1990s, the Habitats Directive on the conservation of natural habitats and of wild fauna and flora (22) was adopted. All EU Member States were henceforward legally bound to designate a part of their territory as protected areas, and to establish for those areas the appropriate measures to guarantee a favorable conservation status of the populations of animal and plant species, and of the natural habitats of Community importance. The totality of these protected areas shall, once approved, form the so-called Natura 2000 network: a coherent network of sites across Europe, in which nature conservation is top priority. It is expected that approximately 12 000 to 15 000 sites will cover between 10 and 15 % of the European territory of the Union.

The sites are of two types: Special Protection Areas (SPAs) designated for the protection of wild birds and Sites of Community Importance (SCIs) proposed for the protection of wild fauna and flora. The data about the sites are compiled by the national competent authorities and submitted to the Directorate General for Environment (DG ENV) of the European Commission. A project has been set up by DG Environment, called *'GIS for Natura 2000'*, to establish an operational database and to develop a system to exploit the data and make them of use for the work carried out by European Commission officials and other users. The construction of the geographic data layer of Natura 2000 is done through a collaborative effort between DG Environment of the European Commission, the Space Applications Institute of the Joint Research Center and the GISCO service of Eurostat. The project started in 1998 and will last for at least three years.

From the experience of the first year of the project, it has become clear that not only technical problems have to be tackled. The accuracy and alacrity with which the Member States submit the information to DG Environment also considerably influence the degree of progress that can be made. This process was speeded up by providing the Member States with adequate information regarding the objectives of the current activities.

It should be noted here that the GISCO database already contains a first dataset referring to Natura 2000. It contains the limits of the so-called bio-geographical regions (Mediterranean, Alpine, Continental, Boreal, Macaronesian, Atlantic and the new regions foreseen for candidate countries). The analysis of the coherence of the network is made region by region.

4.2. Use of the NATURA2000 geo-layer within a GIS environment

Alongside the development of an operational geographic data layer, a user-friendly GIS application is under development. Its functionality is in the first place determined through a user needs assessment. An application for in-house use will allow full (read-only) access to the data and useful background information, whereas a second envisaged application will aim at diffusing the essential information elements available to the general public over the World Wide Web. In addition, appropriate technology will be used and customized for the maintenance of the database, in a completely separate application.

(21) Council Directive 79/40/EEC of 2 April 1979 O.J. L 103 of 25/4/79.
(22) Council Directive 92/43/EEC of 21 May 1992 O.J. L 206 of 22/7/92.

4.2.1. Overview of 'use-cases' of the Natura 2000 data within DG ENV

In the first instance, the officials dealing with the establishment and the follow up of the Natura 2000 network, constitute the main users of any system that will be developed around the (geographic) database. Their tasks are related to the evaluation, assessment and reporting activities concerning the areas proposed or designated for Natura 2000. The user needs assessment carried out in this main group of users revealed the existence of six main types of use of the database ([23]). In the next paragraphs we describe three of the most common use-cases: handling complaints, evaluating LIFE-nature projects and providing opinions on co-financed projects.

Handling complaints

The most important use-case is the handling of complaints related to infringements of EU nature conservation legislation. Currently, DG Environment is dealing with more than 1200 formal complaints. Of these, 50 to 60 % are directly or indirectly related to nature conservation problems. To this one should add all the written and oral questions tabled by members of the European Parliament as well as Petitions. Some examples of data pertinent for handling a complaint are specified below.

Evaluating LIFE-nature projects

For the evaluation a LIFE-Nature ([24]) project, the key task lies in the verification of the type of the site(s) involved (under which directive the site has been proposed), and of the contours of the site(s). Other elements that might need crosschecking include the distribution of cited species across Europe, the overlap with existing LIFE-Nature projects, the eligibility for community funding, etc. Visualization of related data on the screen is the most important functionality to be incorporated in the GIS application.

Providing opinions on co-financed projects

In some cases, DG Environment is requested to provide an opinion regarding proposed projects under one or another Commission funded programme ([25]), since they may have a considerable — often negative — impact on the natural environment. For this task, the visualization of the area affected by the proposed project and its relative position to one or more Natura 2000 sites has to be verified. This can serve as a basis of an impact study. The display of additional data about the area in question might also assist DG Environment to identify areas that should have been designated as Natura 2000 sites, but have not (yet) been proposed by the Member State concerned.

4.2.2. Some exam les of com laints

An element of major importance for handling a complaint obviously is the presence or absence of a Natura 2000 site and/or the existence of a LIFE project in the area concerned. In case there is a Natura 2000 site and/or a LIFE project in place at the location under examination, the desk officer needs to retrieve all available information about them to evaluate the relevancy of the complaint. Furthermore, the availability of information about sites and projects that are operational in the surrounding area may be of key importance to correctly handle the complaint in question. The ability to display this information, together with the visualization of nearby locations for which a (similar) complaint has been filed, will contribute to a better understanding of the overall situation in the area in question.

Often, it is not enough to only know the location of Natura 2000 sites and LIFE projects. Additional background information sometimes proves to be essential to formulate a correct and just answer to the complaint that has been filed. The listing below gives an impression of what a desk officer might need to look up:

- in which administrative region is the location of interest situated;
- which (major) roads run through the area of interest;
- how does the altitude and the slope vary;
- where are cities or villages located and what is their population;
- how is the land cover or the land use distribution in the area;
- where are polluting industrial nucleus' situated;
- does the location of interest lie in an area eligible for community funding;
- etc.

To answer these and similar questions, the concurrent display of data layers is needed. One of the major problems to overcome is

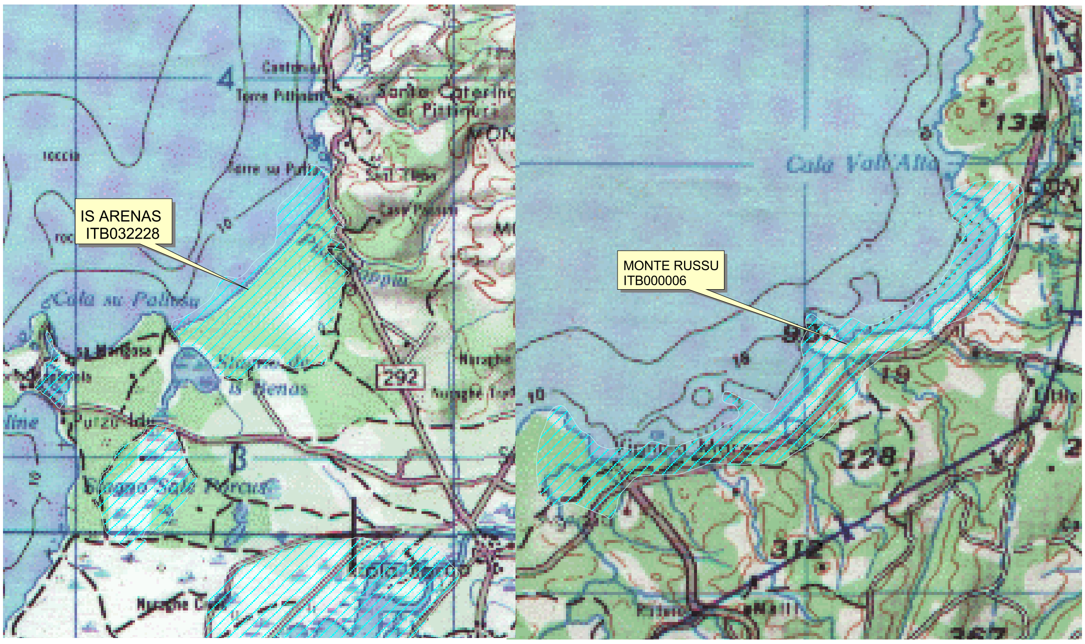
(23) The User Needs Assessment was carried out by the Joint Research Centre.
(24) Regulation (EC) 1655/2000 Official Journal (O.J. L 192 of 28/7/2000).
(25) Typical examples are the Structural Funds, Cohesion Fund, etc.

the non-availability of high quality data at an appropriate scale.

We start with a simple example (not directly related to a specific complaint) of two sites in Sardegna to illustrate this: Monte Russu (code ITB000006), situated in the north of Sardegna and Is Arenas (code ITB032228) in the south-west of the same region. Besides the boundaries of the Natura 2000 sites at scale 1/100 000, a scanned topographic map at scale 1/250 000 is available for this region. For having general background information on the two sites, this might be enough. In general, the sites and topographic data match quite well. Nevertheless, when discussion comes up about particular locations, and more specific questions must be answered, this becomes less obvious. For example, it is not obvious if the provincial road n° 200 passes through or marks the limit of the site Monte. It seems from Figure 1 that the road intersects the site, but we can not be really sure. The road could be at the border of the site, or just inside or outside the site, or even — as can be seen on the map — partially inside and partially outside the site. For Is Arenas, the local road n° 292 touches the site almost perfectly'.

This simple test case gives a first feeling of the typical problems caused by the lack of contextual data at an appropriate scale (1/100 000 or better) and of high quality.

Figure 1. **Extract from the topographic map of Sardegna and the sites of Monte Russu and Is Arenas**

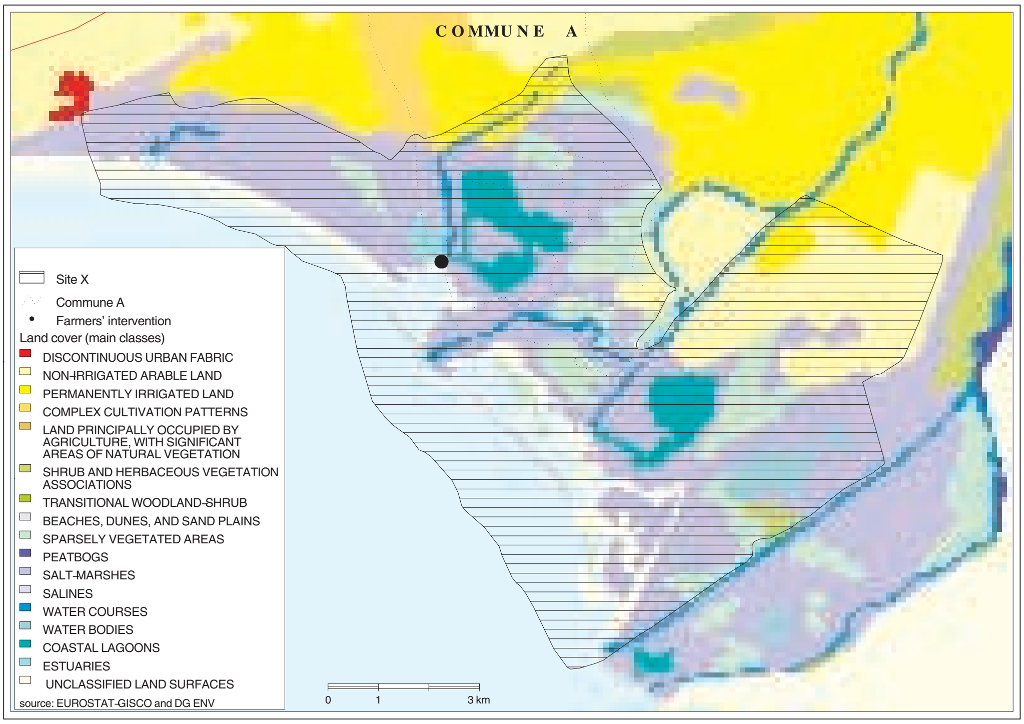


In the next paragraphs we want to illustrate the way complaints are addressed to the Commission and how the Natura 2000 database, topographic and land cover data can be used to understand and answer them. The major data requirements for most of the complaints are twofold. On the one hand some topographic data and a gazetteer are needed to be able to understand and situate the sites in question. This means that the names of populated places and municipalities, rivers, roads, mountain peaks, etc., are of utmost importance together with their relative position. On the other hand information is needed regarding land cover (and even land use) to have information regarding the main activities and natural phenomena. The most important issue here is that the data needs to be up-to-date. We will clarify this with three examples.

Example 1: nature conservation and agriculture in a wetland

In this example agricultural practices were considered as conflicting with ecological systems. In the complaint reference was made to activities that could lead to the drainage of the lagoons in part of this area. These activities were apparently realised to avoid the salinisation of agricultural land. To situate the area Figure 2 was prepared. The CORINE land cover, raster version 100m was used to indicate the land cover classes such as: agricultural land, lagoons, wetlands and rivers, and salinised land. Although this only gives a temporary snapshot (1990), it gives a clear picture to understand the problem. The agricultural areas are very close to- and even within the site. The figure also makes clear that the activities blamed for the potential drainage of the area took place inside the site, in the northern part of the lagoon (see the spot on the map). This last element can be derived from the relative position of the commune A. From the descriptive database, information can be extracted regarding the nesting populations, the habitats, etc.

The wetland case and land cover data from CORINE Figure 2.

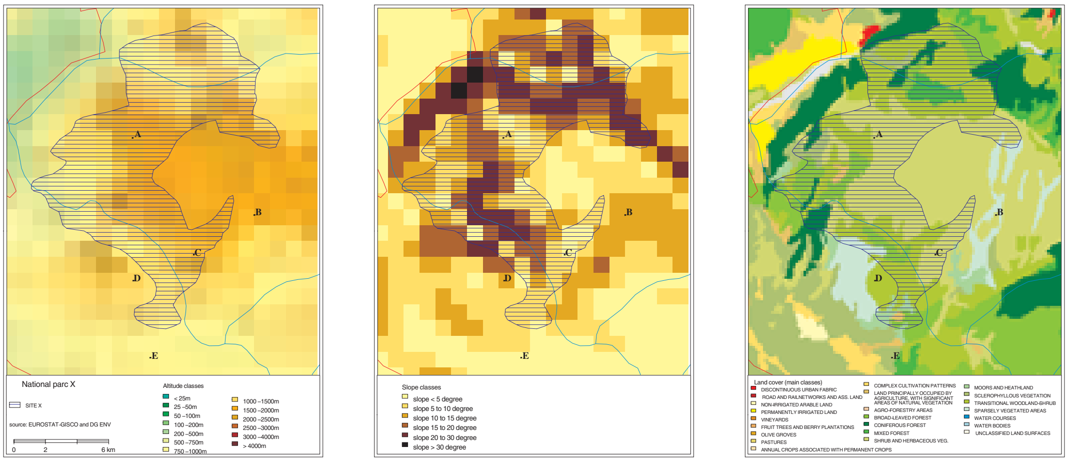


Example 2: nature conservation and tourism in a National ParK

The complaint refers to the development of the area as a tourist centre in the heart of a mountain area famous for its bird species and natural habitats. To be able to develop tourism in the region, the local authorities wanted to build new road infrastructures besides the already existing ones. We want to give some extracts from this complaint to illustrate the way it was formulated towards the Commission. (..). the touristic development which threatens to degrade finally the mountain region of the area X. More precisely it concerns a road construction aiming at connecting the municipality of A and the municipality of B cutting into pieces the mountain chain and a landscape of an extraordinary beauty that until now remained intact.' (..).'Already today, a road network is in place in the area.'

As can be learned from these extracts, the complaint refers specifically to spatial data, including topography (altitude and slope), road infrastructure, place names, etc. Figures 3, 4 and 5 give some data from the GISCO database used to situate the problem. It is clear from these figures that the information is useful to situate the problem, but not detailed enough to verify the correctness of the elements as stated in the complaint.

Figure 3., 4. and 5. **Extracts from the GISCO database: altitude, slope and land cover**



Desk officers need to be able to map this information in a more flexible way: mountain and altitude data, local roads — and if possible roads under construction and planned roads -, vegetation and landscape data, but also tourist sites (in the form of POIs ([26])), administrative data such as regions and municipalities. The data need to be visualised and queried, while it should be possible to edit/add some data. The latter could be done by integrating data coming from the Member State or by interpreting documents (indication of a new local road). Also, from this example it becomes clear that data at level 1/250 000 to 1/100 000 should be collected, at least for the areas concerned.

Example 3: nature conservation and economical development, the case of the A20 (Germany)

In the beginning of the 90s a discussion and later an investigation started concerning the construction of a new Trans European Network road, the A20 in Mecklenburg-Vorpommern, Germany. One of the main objectives of the road construction was to support socio-economic development of this Land, one of the Objective 1 areas in Europe. By adding this new road, the German local authorities aimed at opening up the main cities of the region for economic development. The proposed A20 was originally foreseen to pass by the main cities of the coast (Wismar, Rostock, Stralsund, Greifswald amongst other) and link Western Germany on the one hand and Poland on the other hand.

A complaint was addressed to the Commission stating that this development — although it was recognised to be useful as such — would interfere with the preservation of at least 7 Natura 2000 sites. It was admitted by the organisation that formulated the complaint that the construction of the road would not be possible without influencing these seven sites. Therefore, the discussion was concentrated on the sites where damage would have a major impact on some of the most precious habitats. We describe here the essentials of the complaint to better understand how complaints arrive at the Commission.

The information provided pointed out that the Recknitz and Trebel valleys were not only SPAs under Directive 79/409/EC, but also hosted a range of important habitats under Directive 92/43/EC, including priority habitats (44A1-44A4, for both valleys together at least 1449 ha). Together with the Peene valley, they constitute the largest fen complex in central Europe and are an important corridor between habitat systems on the Baltic coast and further east. For the Trebel, a proposed crossing at Tribsees following the line of an existing highway close to the town of Tribsees was held to be the least negative. This route would have crossed the Trebel at one of the narrowest points of the valley. A more northerly route crossing the Trebel near Bad Sülze would go straight through a biotope for the priority bird species Aquila pomarina. For the Recknitz, a proposed crossing south of Tessin was held to be more destructive than for the north of Tessin. The northern route (more or less parallel to the B105 along the coast) was considered to have a lower impact on the Trebel and Recknitz ecosystems than any of the central routes. This route appeared,

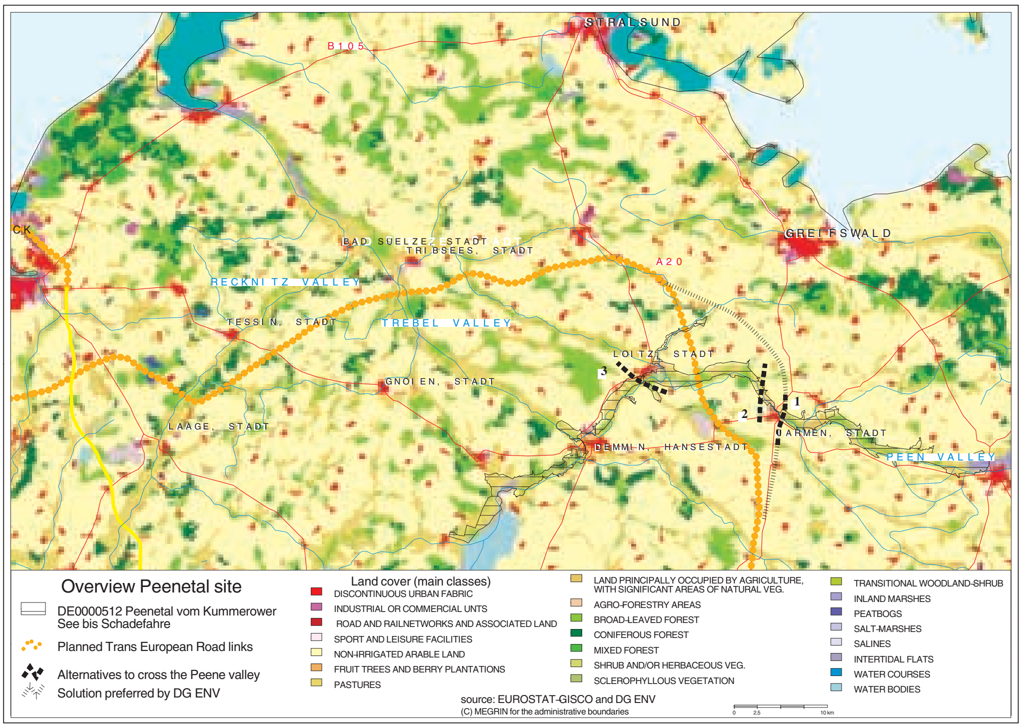
(26) POI or Point of Interest

though, to be less desirable from the point of view of traffic management. Finally there was also a southern route. This route would have followed a trajectory Laage-Gnoien-Demmin and appeared to have the great advantage that all three ecosystems (Recknitz, Trebel and Peene) would be crossed simultaneously at one point instead of three different points as in the northern and central routes.

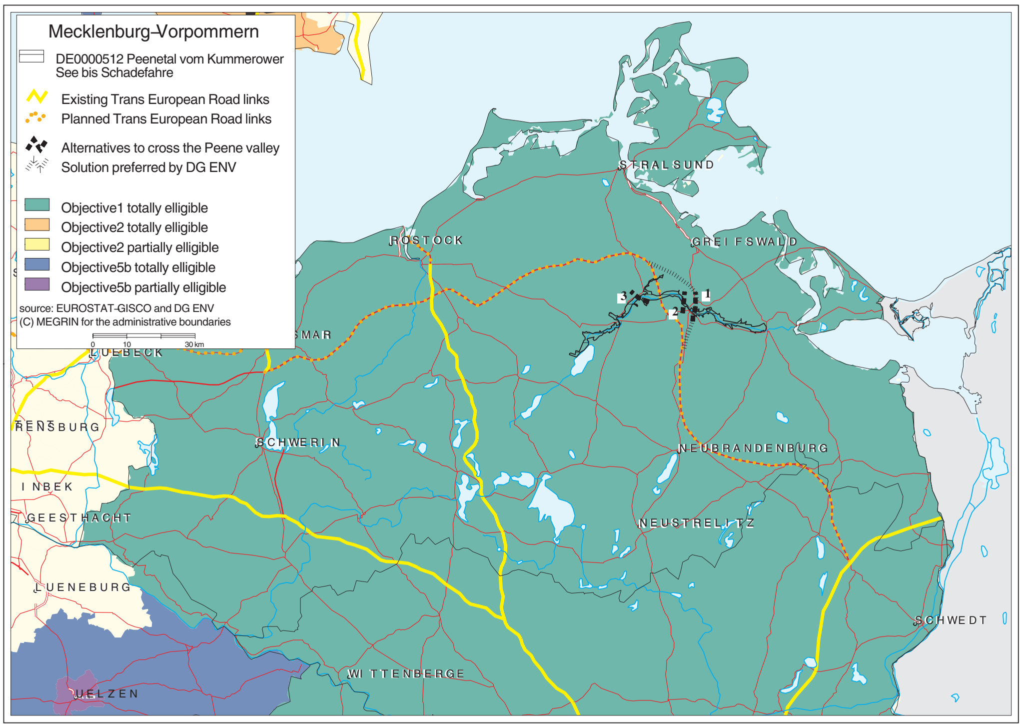
As can be seen from the description above the main issue during the discussion was the place for the crossing of the different sites, the characteristics of the sites at these sections from the ecological point of view and the overall situation within the region.

Finally, the discussion focused on three alternatives when the decision had to be taken for the Peene valley: crossing the valley (1) east of Jarmen, (2) west of Jarmen or (3) west of Loitz. A fourth alternative south of Jarmen was no longer taken into account.

Map of the A20 in Mecklenburg Vorpommern Figures 6.



Figures 7. **Map of the A20 in Mecklenburg Vorpommern**



During the discussions several socio-economic arguments came to the forefront. For example certain cross-sections could not be taken into account at all due to various elements. The dropping of the southern route appeared to be justified by the fact that it was not suitable in terms of traffic management (too far south to be of use for the coast, in particular Stralsund). It would also lead to a greater impact on the environment because of the greater distance vehicles would have to cover to reach e.g. Stralsund. If this route was to serve as a motorway to reduce traffic on the coastal highway B105, then indeed the southern route and even the central routes appeared to be less suitable. However, it was to serve as a fast link between Hamburg and the east (Stettin), then the southern route was actually the shortest in terms of mileage. As for Stralsund itself, the southern route appeared to be the least suitable because it would require a long feeder route' from e.g. Demmin to Stralsund. Figure 6 gives an idea of the planned road by 1995. Note that the data coming from the GISCO database does not yet reflect the discussions on the alternatives and the final option chosen.

Information was also provided concerning a simulation made in which traffic from Hamburg was routed over the central route to Stralsund and over the southern route plus a feeder route, again to Stralsund. Assuming an average speed of 120 km/h, the difference in time to reach Stralsund by the longer southern route instead of the central route was calculated as only 134 seconds — the total difference in kilometers appeared not that great for traffic from Rostock or further west.

All this information was analysed by the Commission during its evaluation of this important case.

In 1996, a final opinion was delivered by the Commission based on article 6.4 [27] of the Habitats Directive. This article states that if, in spite of a negative assessment of the implication for a site hosting priority habitats type and/or species a project must be carried out for imperative reasons of overriding public interest, an opinion of the Commission is required. The preferred solution was that where the crossing of the river Peene would take place east of Jarmen (precisely 300 m east of the existing crossing by the national road B96 and close to a commercial distribution centre — alternative 1 in figures 6 and 7). The above described example of the A20 in Germany revealed that

(27) Commission Opinion of 18 December 1995 O.J. L 006 of 9/1/96.

for DG Environment staff to be able to situate the problem, data was needed at different levels. At level 1, data describes the overall situation of the region, the socio-economic parameters, the general road infrastructure, but also calculated information (e.g. unlocking parameters), etc. At the second level more detailed information of the populated places, the land cover, the road construction is needed. Information for both levels can be provided by the Commission as far as data is available at scale levels 1/1 000 000 and 1/100 000. The third level needs even more detail and must be prepared by the Member States (indicating more detailed information on the species, the landscape, etc.). It is also clear that during the discussion it must be possible to include in a flexible way new alternatives (data) from the Member States, together with existing data from the Commission and that the final results should be integrated in the database of the Commission (GISCO).

4.2.3. Other users

As described before, the geographic data layer of the Natura 2000 network shall eventually be integrated in the GISCO geographic reference database. As such, it will be available for other EU institutions and the general public through the normal GISCO dissemination procedures. Since the Natura 2000 sites are of major importance to the entire European Community, the non-confidential data will be made available to the general public through an Internet application. This will also contribute to the transparency of EU policies to the wider public.

4.3. Natura 2000 sites and CORINE land cover data, the case of Greece

Besides the above use cases, many potential applications of the geographic data layer of the Natura 2000 network exist. The most interesting ones probably lie in the spatial analysis of the data layer and its confrontation with other, relevant geographic data layers. Obviously, several problems arise when combining data that originate from different sources.

Confrontation of data implies that their respective scale must be comparable. This will probably be the major problem for analyses that have to be carried out for the entire European territory. Most pan-European datasets are not available in a harmonized form at a scale that is comparable with the scale of 1/100 000 of the Natura 2000 data layer. Generalisation of the most detailed data layer can in some cases be a useful technique to overcome part of the problem of scale. However, data loss must be taken into account and the sense or non-sense of spatial analysis of the Natura 2000 data at too small a scale has to be evaluated. Indeed, as can be seen from the examples above, for a considerable number of questions asked regarding the Natura 2000 sites, data at an even larger scale than 1/100 000 are required.

Other factors to be taken into account are the comparability of the data (nomenclatures applied, time of monitoring, methodology followed, elements of human interpretation), the availability of complete projection and reference information, reliability of the source, official value of the data, etc. For data that is incorporated in the GISCO reference database, however, these issues have been tackled before incorporation in the database, and should therefore not pose any problems.

Disaggregation of certain data might be useful and/or necessary, but as for generalization techniques, care has to be taken when manipulating data to suit the purpose of the analysis (reference is made to work presented by other authors in this publication).

4.3.1. Land cover classes in the Natura 2000 sites of Greece

To know what types of land cover occur in a site that is designated for nature conservation and what their relative importance is, can give insight into a number of features or can possibly highlight cases that merit close follow up. One can thus verify, for instance, to what extent human activities (urban, industrial and agricultural activities) interfere or interact with nature conservation areas. It is also possible to determine whether a specific land cover type of major importance for a certain (group of) species is duly represented in the proposed nature conservation network. Investigation regarding the diversity of land cover classes within the area covered by the Natura 2000 sites can be appropriate in a given region or for a given species.

As an example, an assessment of the repartition of land cover types in the territory covered by the Natura 2000 sites has been done for Greece. It must be stressed here that the analysis has been carried out on non-validated/old data (version of 1999).

The data used
For the current analysis, two main (geographical) datasets have been used ([28]). The first one comprised the boundaries and site codes of the Sites of Community Importance and of the Special Protections Areas in ARC/INFO vector format. It concerns a first version of a preliminary database that was obtained by DG Environment from the Greek authorities in 1999. The second dataset used was the CORINE land cover (CLC) dataset for Greece ([29]), in ARC/INFO raster format with a resolution of 100m. The first dataset was re-projected from the Transverse Mercator into the Lambert-Azimuthal projection, using the appropriate projection parameters. However, it was discovered that a considerable shift exists between the NATURA2000 sites boundaries and some features that were clearly distinguishable in the CORINE land cover data. On-screen measurements of the difference in coordinates showed that the shift between both layers was systematic. To ascertain better overlap between both datasets, the values for False Easting and False Northing were set to zero. In order to facilitate the spatial analysis, the coverages containing the site boundaries were converted into raster format, using the same resolution and origin as the CLC grid.

Carrying out the analysis
In a first step, the CLC and the site boundary grids were superimposed, transferring the land cover information onto the Natura 2000 sites. From this, an overview was derived of the repartition of the area of each site over the different land cover classes that occur in the site. From there, an overall picture was obtained on the importance of the various types of land cover within the Natura 2000 sites, by calculating total, average, minimum and maximum area and frequency of occurrence of each land cover type. The results were aggregated at CLC levels 1 and 2 to assess the importance of major land cover types. This information was then evaluated in collaboration with the DG Environment Desk Officer for Greece, in terms of relevance for nature conservation issues.

When the results are interpreted, due attention was given to issues for which the results obtained for SCIs and SPAs showed large differences.

Results
Based on the information in the preliminary spatial database, about 20.7 % of the Greek territory is covered by SCIs. This figure includes both terrestrial and marine sites in relation to the terrestrial territory of the country. About 4.6 % of the territory is covered by SPAs. It is expected that some additional SPAs will be defined in the near future because of insufficient SPA classification. This element must also be borne in mind when reading the results of the analysis.

(28) At the time the article was written, Greece submitted a series of new datasets with updated and more accurate data collected within the framework of the 'mapping habitat type' project. Therefore, the results of the analysis carried out in the following paragraphs have to be treated with caution.
(29) The images used for the Greek CORINE Land Cover inventory date from 1987, whereas the Natura 2000 database was established in the late nineties. This discrepancy in time period could possibly partially influence the figures presented in this document. No absolute measures of the degree of influence are available at the time of writing.

Repartition of the Greek national territory, the area under sites of community importance and of the special protection areas over the major CORINE land cover types

Table 1.

Source: (*) EEA, NATLAN; (**) own calculations

Major CLC type	National total (*)		Sites of Community Importance (**)		Special Protection Areas (**)	
	Area (ha)	%	Area (ha)	%	Area (ha)	%
Artificial surfaces	16 7615	1,4	6 487	0,3	1 624	0,3
Agricultural areas	4 642 629	39,2	285 725	11,0	67 401	12,1
Forests and semi-natural areas	6 903 945	58,2	1 352 171	52,2	341 898	61,6
Humid zones	38 306	0,3	33 401	1,3	17 821	3,2
Water bodies	100 152	0,8	622 249	24,0	105 046	18,9
Unclassified			288 249	11,1	21 203	3,8
Total	11 852 647	100	2 588 117 [1]	100	554 993	100

(1) The total area under SCI and under SPA does not entirely correspond to the officially reported respective areas. This is partially due to the above described discrepancy between preliminary spatial and descriptive databases, and partially to thedifference that usually appears between officially reported area of a feature and the area that s calculated in a GIS.

Land cover repartition in Natura 2000 sites

Table 1 shows the repartition over the major CLC types of the national territory as well as SCIs and SPAs. The first does not include the areas for which no data is available in the CORINE land cover inventory; this explains the difference of the national total in the table and the official area of Greece (118. 526km² versus 131 626km²). At national level, a considerable part of the area is covered by forests/semi-natural areas (58,2 %) and agricultural areas (39,2 %). Artificial surfaces (such as urban and all types of industrial areas, infrastructure, etc.), humid zones and water bodies represent only a minor occupation of the land covering respectively 1.4 %, 0.3 % and 0.8 % of the national territory. The relative importance of semi-natural areas and forests (52,2 %) is far less in the Greek SCIs than at the national level. Logically, the SCIs cover much smaller artificial surfaces (0.3 %) and far more humid areas (25,3 %, with 20 % of SCI under marine area). The fairly large part of the area of SCIs covered by agricultural activity (11 %) reflects the fact that extensive agricultural practices can indeed generate an environment favorable for biodiversity. As for the SCIs, only a negligible portion of the SPAs falls within artificial surfaces (0.3 %), and the area of SPAs covered by agricultural activities is more or less the same (12,1 %). However, the relative importance of semi-natural areas and forests has slightly increased (61,6 %) and about a quarter of the area under SPAs status is made up by humid areas (22,1 %), with the marine environment (sea and ocean) representing 11 % of the area under SPA. The very nature of bird protection leads to the inclusion in the protected sites of a large wetland surface.

Diversity in land cover classes within the Natura 2000 sites

The frequency analysis carried out on the data resulting from the spatial overlay revealed that the diversity of land cover types within the Natura 2000 sites is very high. Generally, the larger the site, the more different CORINE land cover types occur, with a few exceptions (mainly for the marine sites).

On average, eight different CORINE land cover types are present in the SCIs. About one quarter of the sites (48) present two to three land cover types. Almost half of the sites (96) show between four and ten different CLC classes on their surface, and about one third (72) between eleven and nineteen different land cover types. In four sites, between twenty and twenty-four different land cover types occur. It concerns fairly large SCIs, but definitely not the largest sites (the average area of those sites is +/- 35 000 ha, whereas the average site area is just under 10 000 ha, with a maximum site area of just over 250 000 ha).

For the SPAs, the diversity is higher. The average number of different CORINE land cover types occurring on the SPAs is fourteen. One third of the SPAs (14) show between four and ten different CLC classes on their surface, and twenty-one SPAs are covered by between eleven and eighteen different land cover types. One SPA has more than twenty land cover types. The average size of SPAs and SCIs is similar: 10 673 ha versus 11253 ha.

The scale of the CORINE land cover inventory is, however, quite small in relation to the complexity of the vegetation pattern that occurs in Greece. As such, the above figures — even though they already reveal a positive situation — are probably an underestimation of the diversity in land cover and vegetation types that are present on the sites.

Importance of specific types of land cover which are of major importance for nature conservation

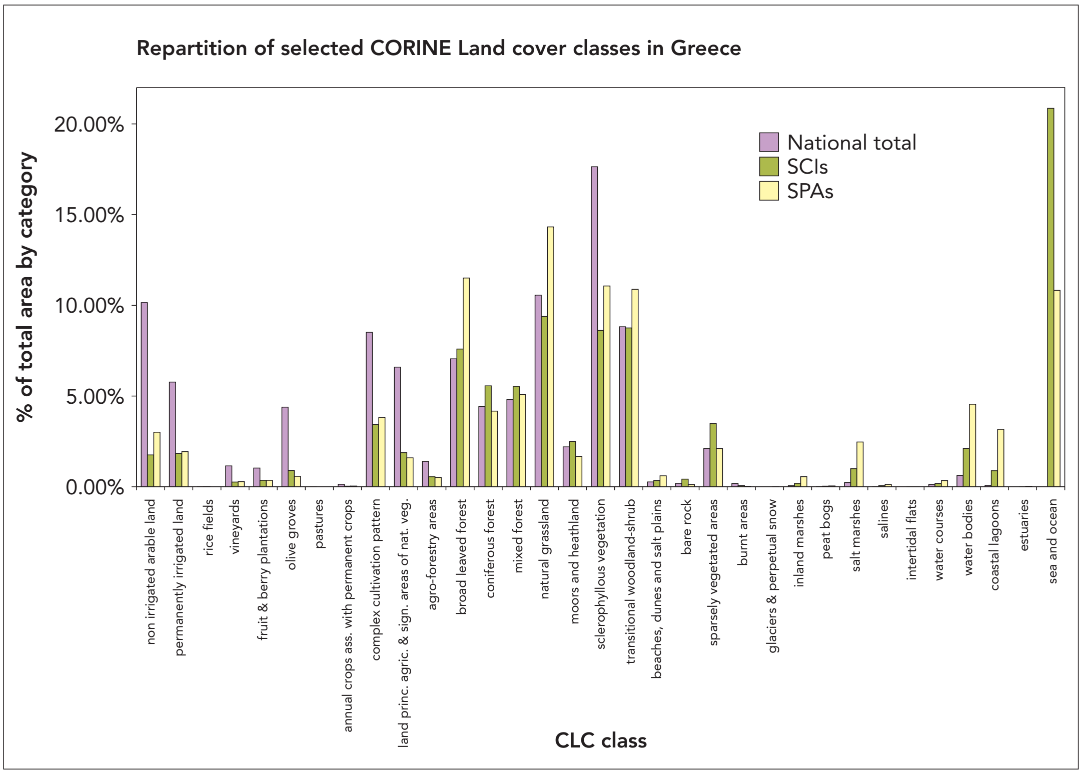
The land cover types that may be considered as relevant for nature conservation purposes are fairly well represented within the area covered by the SCIs: about 1 438 183 ha or 56 % of the SCI area is covered by those land cover types (30). For the SPAs, the situation is even better with 417 243 ha or 75 % of the recorded SPA area being covered by these land cover types.

Occurrence level of land cover types

Looking at the frequency of occurrence of the various land cover types, complex cultivation patterns and agricultural areas, with significant areas of natural vegetation, were found on half of the sites. This confirms the described co-existence of extensive agriculture and nature conservation areas. Furthermore, the natural grassland, the sclerophyllous vegetation, plus the transitional woodland-shrub and sea and ocean were found to be the most frequently present land cover types within the sites. No anomalies with respect to the occurrence of any land cover type were detected. The discontinuous urban fabric category has been identified in one third of both SCIs and SPAs. This is due to the large size of the sites and the way they were originally delimited. The area covered is, however, insignificant (0.2 % for SCIs and 0.26 % for SPAs).

Figure 8. **CORINE land cover classes in Greece**

Source: EEA/GISCO/ own calculationsnt



4.3.2. Recorded land use in the Natura 2000 sites against CORINE land cover data

The database containing the descriptive data reflecting the characteristics of the Natura 2000 sites, is constructed using the information the Member States reported to DG Environment. This data is transferred by the Commission to the ETC/NC in Paris, which puts them together in an operational database. This database describes the general characteristics of the sites such as name, code, area, position, administrative data as to which organism is responsible for the sites, etc. Part of the information refers to the

(30) Important land cover types for nature conservation in Greece: non irrigated arable land; complex cultivation pattern; broad-leaved forest; coniferous forest; mixed forest; natural grassland; sclerophyllous vegetation; transitional woodland-shrub; beaches, dunes and salt plains; bare rock; salt marshes; water bodies; water courses; coastal lagoons; estuaries.

degree in which the sites cover marine areas and to the habitat groups that occur in the Natura 2000 sites. It is this type of information that has been compared with the CORINE land cover database.

To this purpose, the detailed data contained in the descriptive database were aggregated and totaled for the SCIs, on the one hand, and the SPAs on the other. As such, a general picture was generated of the repartition of the total area under the respective types of sites, over land and marine areas, and over various types of general habitats. This data could then be compared with the results of the analysis based on the CLC inventory.

Land area versus marine area in the Natura 2000 sites

The information about the repartition between land area and marine area in the descriptive database is of a fairly basic nature. For each Natura 2000 site the percentage of the area that is not covered by any NUTS ([31]) region is recorded. Based upon this data, the total marine area in the Natura 2000 sites was calculated and compared with the CLC class Sea and Ocean' (code 5.2.3). As can be seen from Table 2, the distribution patterns of both data sources correspond well.

The repartition of land and marine area in descriptive database and in CLC inventory Table 2.

	SCI		SPA	
	ha	%	ha	%
Descriptive database (RegCod table)				
Land area	2 010 549	78 %	487 706	88 %
Marine area	554 825	22 %	65 252	12 %
Total area	2 565 374		552 958	
CORINE land cover inventory				
Land area *	2 048 451	79 %	494 905	89 %
Marine area **	539 666	21 %	60 088	11 %
Total area	2 588 117		554 993	

* All CLC classes excluding class 5.2.3 'Sea and Ocean', but inclusive not-classified territory.
** CLC class 5.2.3 'Sea and Ocean'

Source: ETC/NC and own calculations

Part of the area that has not been classified in the CLC inventory covers coastal areas of Greece, where a fairly large number of the Natura 2000 sites are situated (about one quarter of the SCIs, with the not classified area accounting for 11 % of the SCI area, and about one ninth of the SPAs, with the not classified area accounting for 4 % of the SPA area). As such, a considerable part of these sites that overlap with not-classified areas, would be covered by Sea and Ocean' should the CLC inventory have been complete. Clarifying and straightening out the differences in area for the individual sites in both data sources (viz. the descriptive database and the geographic datasets) is one of the elements of the work carried out in the framework of the construction of the geographic data layer for the Natura 2000 sites.

General habitats and CLC in Natura2000 sites

The descriptive database contains for each site the repartition of its area over 23 different habitat groups. These figures are the result of estimates made by the Greek authorities, based on a variety of methods for area estimation, including measurement on maps, field inventorying, photo-interpretation, that have been applied on a sample basis. The total area under each of those habitat types, in respectively SCIs and SPAs, is given in Table 3. The most important classes are shaded in grey. Four classes account for more than 65 % (SCI) and almost 52 % (SPA) of the area: marine areas and sea inlets (N01); heath, scrub, maquis and Garrigue, Phygrana (N08); broad-leaved deciduous woodland (N16) and coniferous woodland (N17).

(31) NUTS = Nomenclature des Unités Territoriales Statistiques: a hierachical nomenclature for administrative regions, which is used by Eurostat for collection of statistical data at EU level.

Table 3. **Main habitat classes for the SCI and SPA sites**

Source: ETC/NC and own calculations

CODE	Description	SCI		SPA		
		Area (ha)	Area %	Area (ha)	Area %	CLCcdl1
N01	Marine areas, sea inlets	469 878	18,2 %	18 167	3,3 %	5
N02	Tidal rivers, estuaries, mud flats, sand flats, Lagoons (including saltwork basins)	33 878	1,3 %	21 724	3,9 %	4
N03	Salt marshes, salt pastures, salt steppes	30 154	1,2 %	19 906	3,6 %	4
N04	Coastal sand dunes, sand beaches, machair	20 490	0,8 %	4 994	0,9 %	3
N05	Shingle, sea cliffs, islets	17 975	0,7 %	2 608	0,5 %	4
N06	Inland water bodies (standing water, running water)	89 868	3,5 %	32 927	6,0 %	5
N07	Bogs, marshes, water fringed vegetation, fens	38 066	1,5 %	14 869	2,7 %	4
N08	Heath, scrub, maquis and garrigue, phygrana	556 649	21,5 %	115 293	20,9 %	3
N09	Dry grassland, steppes	69 900	2,7 %	22 253	4,0 %	3
N10	Humid grassland, mesophile grassland	20 459	0,8 %	5 926	1,1 %	3
N11	Alpine and sub-alpine grassland	134 696	5,2 %	16 113	2,9 %	3
N12	Extensive cereal cultures (including rotation cultures with regular fallowing)	86 919	3,4 %	6 320	1,1 %	2
N13	Ricefields	2 694	0,1 %	1 428	0,3 %	2
N14	Improved grassland	1 279	0,0 %			2
N15	Other arable land	6 6864	2,6 %	37 836	6,9 %	2
N16	Broad leaved deciduous woodland	333 461	12,9 %	99 703	18,1 %	3
N17	Coniferous woodland	332 733	12,9 %	52 927	9,6 %	3
N18	Evergreen woodland	55 771	2,2 %	28 512	5,2 %	3
N19	Mixed woodland	34 114	1,3 %	6 702	1,2 %	3
N20	Artificial forest monoculture (e.g. plantations of poplar or exotic trees)	7 121	0,3 %	1 182	0,2 %	3
N21	Non-forest areas cultivated with woody plants (including orchards, groves, vineyards, dehesas)	41 405	1,6 %	2 107	0,4 %	2
N22	Inland rocks, screes, sands, permanent snow and ice	61 829	2,4 %	12 288	2,2 %	3
N23	Other land (including Towns, Villages, Roads, Waste places, Mines, Industrial sites)	43 759	1,7 %	4 831	0,9 %	1
Total		**2 583 842**	100 %	**550 338**	**100 %**	

The mentioned habitat classes in Table 3 can be regrouped according to the main CORINE land cover classes: artificial surfaces (1), agricultural areas (2), forests and semi-natural areas (3), humid zones (4) and water bodies (5). The areas of SCI and SPA sites found in the five principle classes (level 1) of CLC are provided in Table 4. Between brackets we have indicated the relative area that was calculated from the CORINE land cover data (see also table 1). We can conclude from this table that the habitat classes match quite well with the CLC classes.

Habitat types recalculated to the 5 main CORINE land cover classes — Table 4.

Source: ETC/NC and own calculations

CLC class	SCI		SPA	
	Area (ha)	Area (%)	Area (ha)	Area (%)
artificial surfaces	43 759	1,7 (0,3)	4 831	0,9 (0,3)
agricultural areas	199 161	7,7 (11,0)	47 691	8,7 (12,1)
forests and semi-natural areas	1 627 222	63,0 (52,2)	365 892	66,5 (61,6)
humid zones	120 074	4,6 (1,3)	59 106	10,7 (3,2)
water bodies	593 625	23,0 (24,0)	72 818	13,2 (18,9)
Total	**2 583 841**	**100,0**	**550 338**	**100,0**

- Artificial surfaces seem to be more important in the Natura 2000 database than in the CLC database.
- The agricultural areas seem to be overestimated in the CLC database.
- Exactly the opposite is the case for the forests and semi-natural areas: these are more important in the Natura 2000 habitat database.
- Humid zones are far more important in the Natura 2000 database than in the CLC database.
- Water bodies seem to be slightly more important in the CORINE land cover database than in the Natura 2000 habitat database.

All these facts can probably be explained by the methodology used. The CORINE land cover database is less detailed' because units are always larger than 25 ha. Therefore, classes that tend to cover smaller zones (e.g. artificial areas) are underestimated while on the contrary, classes that tend to cover larger zones (e.g. agricultural areas) are probably overestimated. This does not mean that the data from the Natura 2000 descriptive database is by default more reliable, neither does it give more detailed information. In fact, the habitat data in the descriptive database give only a rough estimation, but the estimation is always relative to the size of the sites, and these are often small, even smaller than the smallest unit of the CLC database (25 ha).

4.4. Conclusions

✓ The first deliverables of the project GIS for Natura 2000 show how important this initiative may be as a tool to support the set-up of the network. The control of quality and completeness of the geographical information delivered by Member States would be impossible without a GIS application.

✓ Once the network is established the need for GIS tools will increase. This is clearly demonstrated by the increasing number of complaints related to Natura 2000. The level of detail required for their analysis highlights the basic limiting factor of this project: the lack of contextual data at an appropriate scale and the pan-European coverage.

✓ Today, the GISCO database only contains useful information to overview the complaint, but a very limited number of layers at the scale required for analysis/ problem solving (1/100 000 or better) and none of these refer to basic topographic features, such as water bodies, roads, etc. This is obviously a problem not only for Natura 2000, but for many other initiatives currently being developed by Community institutions. The lack of a legal Community framework and the conservative policy of the national cartographic agencies are the major causes for this gap. This situation may only be solved through a coordinated effort (a good example for this is the COGI[32] inter-service group set up by Eurostat).

✓ The CORINE land cover database is certainly one of the most interesting data layers available from the GISCO database at a scale comparable to that of Natura 2000. The cross comparison between CLC and Natura 2000 for Greece shows how important CLC data may be for this project.

✓ Unfortunately, the level of detail of CLC only allows for performing some statistical analysis. This is mostly due to two factors: the broad definition of land cover types and the insufficient geographical resolution.

✓ To overrun these limits more detailed data could be collected within Natura 2000 sites during the CLC update. In particular, land cover types need to be disaggregated to the

(32) Commissions Geographic Information Inter service Group

same level of definition as Annex I in the Habitats Directive.

- ✓ Another solution could be found in the construction of a database of scanned topographic maps for at least the Natura 2000 area concerned, or by using data extracted from images from the IMAGE 2000 project of the Commission or from similar initiatives.
- ✓ It should be noted that this type of data is frequently available at national or local level for many sites. Frequently this information has been gathered thanks to the Community financial support (e.g. LIFE). This means that in these cases, data only need to be integrated and harmonised, as they have already been collected.
- ✓ If this effort is realised, CLC may become the reference data layer for any application developed by the project GIS for Natura 2000.

Acknowledgements

Fotios Papoulias, DG Environment, Nature and Biodiversity Unit (B.2) for reading and commenting on the paper.

Oliver Schall, DG Environment, Nature and Biodiversity Unit (B.2) for reading and commenting on the paper.

Anne Burrill DG Environment, Territorial Unit (B.3) for reading and commenting on the paper.

Amandine de le Court, Ground for GIS, K.U.Leuven R&D for performing some calculations and preparing the maps.

Paul Campling, Ground for GIS, K.U.Leuven R&D for proof reading the paper.

5. Reassignment of the Farm Structure Survey's data

Maxime Kayadjanian*, Claude Vidal** (* LANDSIS, ** Eurostat/F1)

5.1. Introduction

The reform of the Common Agricultural Policy (CAP) aims to promote environmentally friendly farming practices with a view to sustainable development and the conservation of both the quality and diversity of rural areas. This reform must also maintain high productivity levels to enable farmers to remain competitive. The assessment of agricultural policies and of their impact on the countryside and landscapes proves increasingly necessary.

First indispensable stage in this evaluation is the study of the spatial units that constitute the underlying structure of these territories. A major part of the statistical data is available for administrative units (NUTS regions), with no direct way to assign them to units more relevant from the geographical point of view (drainage basins, landscape units, etc.).

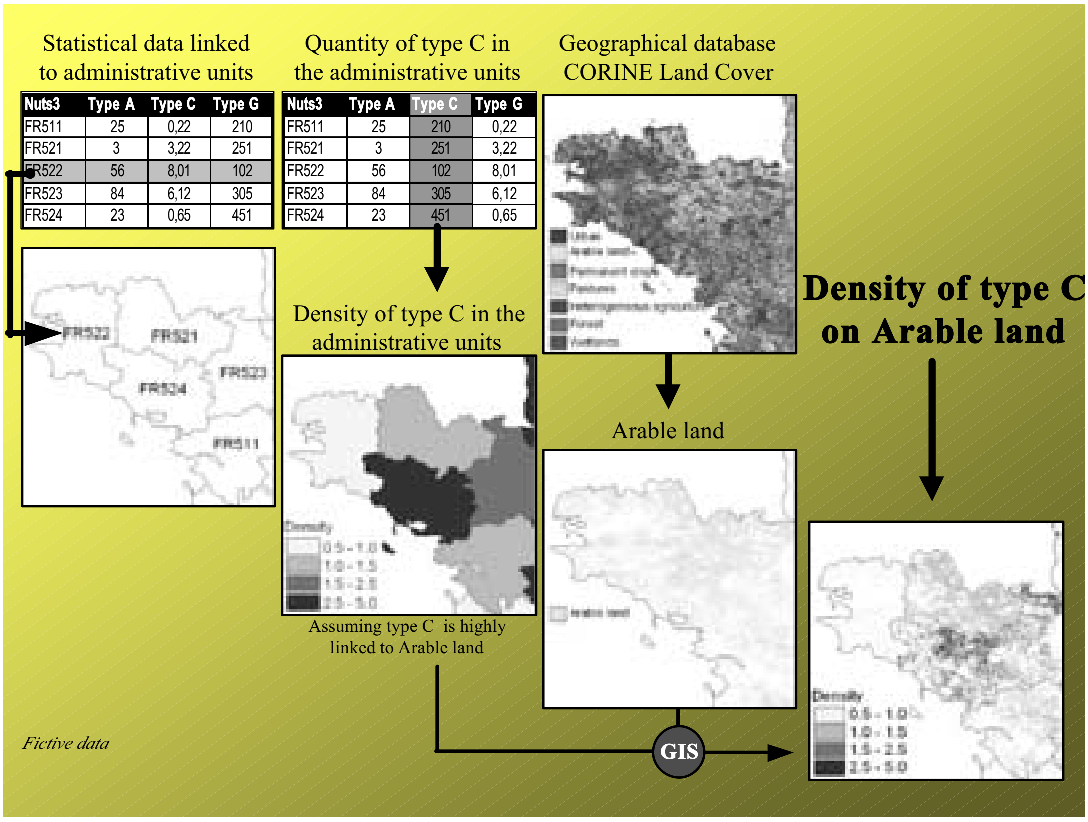
If Eurostat can have tools allowing such an assignment it would open new prospects in terms of data analysis in the fields of agriculture and environment. It would make it possible to develop relevant agro-environmental indicator calculations.

Among the various databases on agriculture available at Eurostat, the farm structure surveys (named Structure surveys in the remaining part of this article) — and in particular the basic surveys which take place every ten years — have a special place. These are to date the only surveys to report at the same time on the structure of the holdings, their managers, the land use, livestock, and the labour force. They provide inter alia data concerning the various crops as well as the number and types of agricultural machinery. Results are given for relatively fine administrative units at the European scale-survey districts (geographical limits based on the NUTS regions). This geographical level is not, however, appropriate to carry out certain environmental studies. Structure survey data require for this objective a comparison with other sources of information. Parameters such as topography, pedology and climatology of the different types of agricultural land are fundamental if one wants, for example, to evaluate the risks of erosion or of pollution of watercourses by pesticides. Knowing agricultural areas by type of crop within survey districts is insufficient. It is necessary to localise this information more precisely. This will allow the reallocation of data into suitable areas, such as drainage basins, while limiting the loss of information ([33]).

The geographical database CORINE land cover (CLC) can be used as a means for the spatial disaggregation of Structure survey data into a more accurate geographical level (Figure 1), and is a first step towards a satisfactory spatial analysis.

(33) See paper of C. Vidal, J. Gallego & M. Kayadjanian, infra.

Figure 1. **General principle of spatial disaggregation of statistical data on CORINE land cover [1]**



(1) Inspired by the HYDROSOL project of the French Institute of the Environment ("IFEN - HYDROSOL)

Structure surveys and CLC commonly describe land cover and land use. Definition of an interface between their nomenclature is a precondition for this spatial disaggregation.

5.2. Characteristics of the Farm Structure Surveys

Structure surveys consist of a census organised every ten years to which are added intermediate surveys by sample survey every two or three years. The data collected by the statistical Member State services are forwarded to Eurostat which checks and stores them in the Community database. The first survey, carried out in 1966/67, arose from the need to have harmonised information at the Community level. Since then, regulatory texts have defined the methodological framework and the contents of the Structure survey questionnaires.

5.2.1. Sco e of the survey and contents
Holdings surveyed are those where agricultural area (AA) is higher than 1 ha or an equivalent level of economic activity. This threshold can be higher for certain countries insofar as, in agreement with the regulatory texts, the holdings that are not taken into account by the survey do not add up more than 1 % of the total agricultural economic activity. Thus, the part of Community agricultural activity not covered by the censures is very low.

The Structure survey comprises the following headings:

- **Land use** (Table 1). Data is collected in ha and concerns the principal following classes: arable land (D), kitchen gardens (E), meadows and pasture (F), permanent crops (G), others: wood, roads, buildings, etc. (H). The holding AA is the sum of D+E+F+G areas, and the total area of the holding is the sum of AA+H area. Lastly, in the category I irrigated arable areas, area under glass, comprising successive crops as well as the annual crops associated with the permanent crops are entered. This category repeats areas entered in the previous classes;
- **Information on the holding** (legal status, type of tenure, presence of accounts, etc.) and the holder (manager, education level, employee, other gainful activities, etc.);
- **Livestock**: equidae, cattle, sheep, goats, pigs, poultry and others;

- **Agricultural machinery**: tractors, reaping-machines, milking facilities, etc.;
- **Holding labour force**;
- **Ty ology information**: The standard gross margin (SGM) which allows the classification of agricultural holdings by Farm type also gives the economic size of an agricultural holding.

Member States locate the data they collect on these variables to the holding's survey district (e.g. commune), based on where the headquarter of the holding is situated but this detailed information is not available for reasons of statistical confidentiality.

Simplified classification of land use in the Structure surveys 1989/90 1 Table 1.

Code	Principal headings	Secondary headings
D	Cereals	D/01 common wheat and spelt
		D/02 durum wheat
		D/03 rye
		D/04 barley
		D/05 oats
		D/06 grain maize
		D/07 rice
		D/08 Other cereals
	Dried vegetables	D/09/a as pure crops for fodder
		D/09/b others
	Root crops	D/10 potatoes
		D/11 sugar beets
		D/12 fodder roots and brassicas
	Industrial plants	D/13/a tobacco
		D/13/b hops
		D/13/c cotton
		D/13/d other oilseed crop or fibre plants and other industrial plants
	Fresh vegetables, melons, strawberries	D/14 fresh vegetables, melons, strawberries outdoor
		D/15 fresh vegetables, melons, strawberries under greenhouse
	Flowers and ornamental plants	D/16 Flowers and ornamental plants outdoor
		D/17 Flowers and ornamental plants under greenhouse
	Forage plants	D/18/a temporary grass
		D/18/b other forage plants
	Other crops of arable land	D/19 arable land seeds and seedlings
		D/20 other arable land crops
		D/21 fallow land
E	Kitchen gardens	E kitchen gardens
F	Permanent pastures and meadows	F permanent pastures and meadows
G	Permanent crops	G/01 fruit and berry plantations
		G/02 citrus plantations
		G/03 olive plantations
		G/04 vineyards
		G/05 nurseries
		G/06 other permanent crops
		G/07 permanent crops under glass
H	Unutilized agricultural land	H/01 and H/03 unused AA
		H/02 woodland
I	Successive secondary crops	I/01 successive secondary crops
		I/02 mushrooms
		I/03 irrigable and irrigated area
		I/04 ground area covered by greenhouses in use
		I/05 combined crops

(1) Regulation 89/651/EEC (OJ N° L391 of 30.12.89).

5.2.2. Date of the surveys and Geogra hical coverage

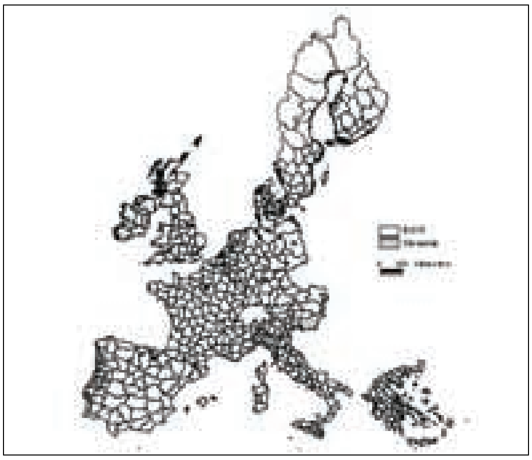
Table 2. **Type and date of surveys since 1966/67**

Type of survey	Year of survey	EU-	Duration
Sample survey	1966/67	6	
Agricultural census	**1970/71**	**6**	
Sample survey	1975	9	march-75/march-76
Sample survey	1977	9	
Agricultural census	**1979/80**	**9+EL**	**apr-79/juin-80**
Sample survey	1983	10	oct-82/janv-84
Sample survey	1985	10+E+P	dec-84/march-86
Sample survey	1987	12	dec-86/march-88
Agricultural census	**1989/90**	**12**	**dec-88/march-91**
Sample survey	1993	12	dec-92/march-94
Sample survey	1995	15	dec-94/march-96
Sample survey	1997	15	dec-96/march-98

EU-6 = B, D, F, I, L, NL;
EU-9 = EU-6 + DK, IRL, UK;
EU-10 = EU-9 + EL;
EU-12 = EU10 + E, P;
EU-15 = EU-12 + O, S, SW

Some CEECs such as Slovenia have already organised their agricultural surveys in accordance with the specifications from the Structure survey within the framework of their future accession to the European Union and the integration of the acquis communautaire .

5.2.3. S atial observation unit
Structure survey data is provided by some Member States at very detailed geographical level. But for reasons of confidentiality, they are generally useable only at district level (except for Ireland where data is available at communal level. Within the framework of this work, the Irish data has been aggregated into units of homogeneous size similar to the other countries in the survey).



Map 1. District limits (EU15) and national limits of Slovenia

Districts are based on administrative levels of Europe (NUTS 2 + 3 regions), see Map 1.

The data on Slovenia, available at national level, was also used.

5.3. Relocalisation of the Structure survey's data

5.3.1. Objective
Structure survey nomenclature (Table 1) distinguishes between detailed agricultural land use classes. Even though the area of each class is known for every district the location within the district is not. CLC allows land cover/land use spatial units (or polygons) of at least 25 ha to be located. Its nomenclature, however, is much less detailed (34). It comprises 11 classes for agricultural land, against 60 for the Structure survey nomenclature.

To start with CLC's spatial units that correspond to arable land as indicated by Structure surveys must be located. (Figure 1) details the different steps which aim to relocate the surfaces of the Structure survey. The first stage of this process is to match CLC classes and the Structure nomenclature. This preliminary work will define for each district

(34) cf. description of the nomenclature annexed in the paper of C. Vidal, J. Gallego & M. Kayadjanian, infra.

the share of each Structure class within the CLC polygons. Each CLC polygon has, therefore, to be allotted a probability of containing a given type of Structure land use. For example, for a certain district, cereals from Structure might be distributed as 90 % on the CLC's arable lands and as 10 % on CLC's heterogeneous agricultural areas contained in this district.

5.3.2. Pros ects

Once this initial stage has been carried out, it becomes possible to redistribute quantitative data other than land use from the Structure surveys (for example livestock) by defining distribution rules using covariables. It will also be possible to relocate variables coming from other sources of information (for example the average quantity of inputs used for a given type of culture). At this stage, the database CLC could be combined with other layers of geographical information (climate, topography, pedology, socio-economic data) in order to refine the location rules to be applied to the variables that one wishes to spatialise. These various operations should provide (35) a well stocked database able to deal with the problems of sustainable agriculture at an appropriate geographical level.

For example, to deal with the nitrogen pollution potential, it will be possible to reaggregate the data in this base according to drainage basins. These drainage basins will of course have to be of an appropriate size in order that the Structure survey data is representative.

5.4. Comparison of nomenclatures

Comparing the two nomenclatures requires:

- the choice of aggregation level of the classes for which a correspondence has to be set up;
- the validation of this choice by comparison of the respective surfaces area of the related classes.

Building links between classes of two nomenclatures requires as far as possible equivalence of coverage and information level.

5.4.1. Data used

The Structure survey of 1989/90 was used. This survey corresponds to a census in the twelve Member States of the European Community at the time. Its completion date represents the average date of the satellite images sources from CLC dating as it does from 1985 to 1995. The comparison data concerns the Community of Twelve, except the United Kingdom for which CLC data was not available.

In order to extend the analysis to future Member States, data on Slovenia was also incorporated. Data used was for 1997 and was available at the national level. Slovenia is currently the only future Member State to provide results of agricultural surveys which conform with Structure.

In order to be able to compare agricultural areas entered in both databases, CLC surfaces were consolidated by district (at the national level for Slovenia) by applying this zoning system to the CLC base.

5.4.2. Highest aggregated level of the nomenclatures

An aggregate cross check to determine if both databases furnish comparable overall information, the values of the classes were calculated at their highest aggregated level. Thus the agricultural areas (i.e. class 2) for CLC and the agricultural area (AA, classes D, E, F, G) for the Structure survey were compared. Map 2 represents the relative differences (36) between AA and the CLC's agricultural areas. The surfaces coloured in maroon are those where CLC's agricultural areas underestimate Structure's AA. Those in blue are areas in which CLC's agricultural areas are higher than total AA of the holdings in Structure.

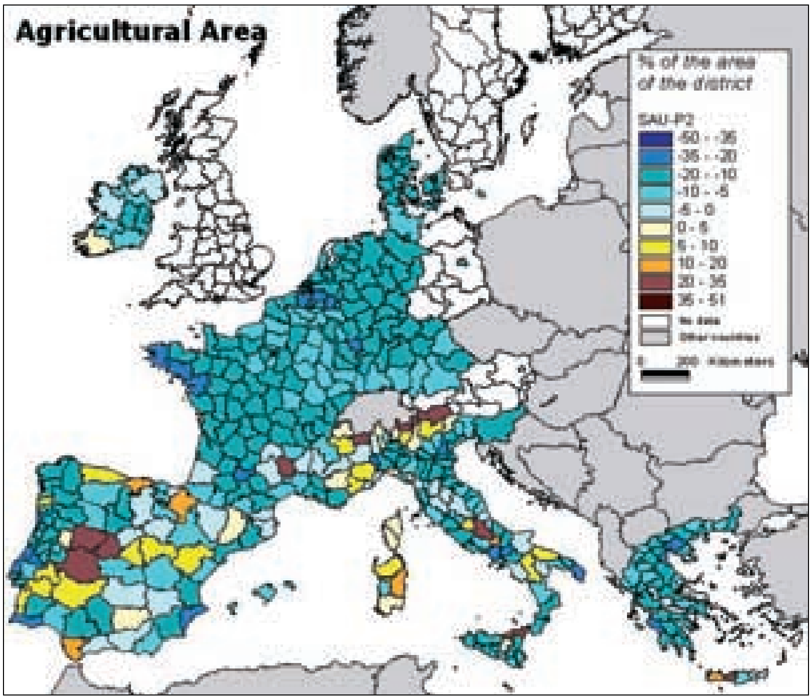
(35) ibid.
(36) Let us note that expressing the value of the differences in percentage terms of the area of the district gives the advantage to the differences occupying important surfaces. The differences weaker in terms of area but important relatively, show up less. Nevertheless, this type of presentation makes it possible to highlight the most significant problems.

Table 3. **Percentage of districts where agricultural areas are larger in CLC than in the Structure survey**

Country	%
Belgium	100
Denmark	100
Germany	100
Greece	93
Spain	73
France	92
Ireland	88
Italy	77
Luxembourg	100
The Netherlands	100
Portugal	82
EU-11	**92**

5.4.2.1. Large differences
Overall, CORINE land cover gives more agricultural area than the Structure survey. This relationship is found all over Europe (92 % of the districts, i.e. 85 % of the area of the 11 countries of the European Union concerned). These differences are marked in certain districts and can reach almost the third of the area of a district (southern Brittany, north of Belgium, south of Portugal).

Map 2. **Comparison between AA and agricultural areas**



The differences are expressed in percentage of the area of the district. Maroon indicates positive differences — Structure > CLC -, the blue indicates negative differences — CLC > Structure -. This cartographic principle is included in the comparison maps that follow.

Where agricultural areas reported in the Structure survey are higher than those of CLC (mainly Spain, south of France and Italy) the differences can also be quite important.

5.4.2.2. Some explanations and problems
Because of differences of cover European farms versus the whole European surface area and of data collection methods (census versus photo-interpretation), three factors should be considered to explain the differences between agricultural areas in the two sources of information:

- semantics: the definition of CLC classes is such that non agricultural land can be entered as agricultural. For example the item 243 (Land principally occupied by agriculture, with significant areas of natural vegetation') contains by definition a part of non-agricultural land.
  Conversely, other non-agricultural classes can take into account areas entered as agricultural in the Structure survey. For example the item CLC 321 (natural grassland') can include surfaces classified in Structure to permanent meadows and pastures' (item F).
  Finally, some surfaces used by agriculture are not entered in the Structure survey. It is the case of the communal meadows in Ireland and in Scotland and of certain mountain pastures. They are taken as agricultural (i.e. meadow) in CLC.

- geometrical: the minimum size of 25 ha of CLC's mapping units means that a number of non agricultural polygons are classified as agricultural while they are only partially agricultural (urban fringes, semi-natural spaces).
  Conversely, polygons classified as non agricultural territories in CLC can include a part of agricultural land.
  This problem will be found at a more detailed level within the nomenclature for the agricultural classes. Thus, a part of meadows or of permanent crops can be included in polygons of arable lands and conversely.

- temporal: the acquisition period of the data was spread out over 2 years for the Structure survey (1989 to 1990), and over 10 years for CLC (1985 to 1995). This temporal difference explains certain differences. Additionally, the reform of the Common Agricultural Policy (introduction of agri-environmental measures in 1992, in particular) has also induced change within certain rural areas.

Linking both nomenclatures has to be repeated with the classes of a more detailed level to identify those where there exists strong differences. This second stage furnishes various explanations to the questions thrown up by the analysis of the differences.

5.4.3. Linking nomenclatures at a finer level
Level 3 of CLC's nomenclature, being the most detailed one, has 11 agricultural classes related to those of the Structure survey. Some CLC classes can directly be connected to those of the Structure survey (i.e. rice plantations, vineyards, etc.). Others can be connected only to aggregated classes of the Structure survey (i.e. arable lands). Finally, others have no explicit link, such as the item 242 of CLC, Complex cultivation patterns' (Table 4).

Matching the CLC and Structure nomenclatures — Table 4.

CORINE land cover		Structure
Level 2	*Level 3*	
21 Arable land	211 non-irrigated arable land	D arable lands (except irrigated arable lands)
	212 permanently irrigated land	I/03/b Irrigated arable land
	213 rice fields0	D/07 rice
22 Permanent crops	221 vineyards	G/04 vines
	222 fruit trees and berry plantations	G/01 plantations of fruit trees and soft fruit
	223 olive grooves	G/03 olive groves
23 Pastures	231 pastures	F permanent meadows and pastures
24 Heterogeneous agricultural areas	241 annual crops associated with permanent crops	I/05/b combined crops: permanent crops — annual crops
	242 complex cultivation patterns	?
	243 land principally occupied by agriculture with significant areas of natural vegetation	?
	244 agro-forestry areas	I/05/a combined crops: agricultural crops — forestry species

One finds here a sharper expression of the problems relating to the differences in the definition of the classes. Part of the land put by CLC in heterogeneous classes, in particular items 242 and 243, are either completely or partially neglected in the comparison of arable land or meadows. These classes should have been taken into account, but there is no basis for allocating them.

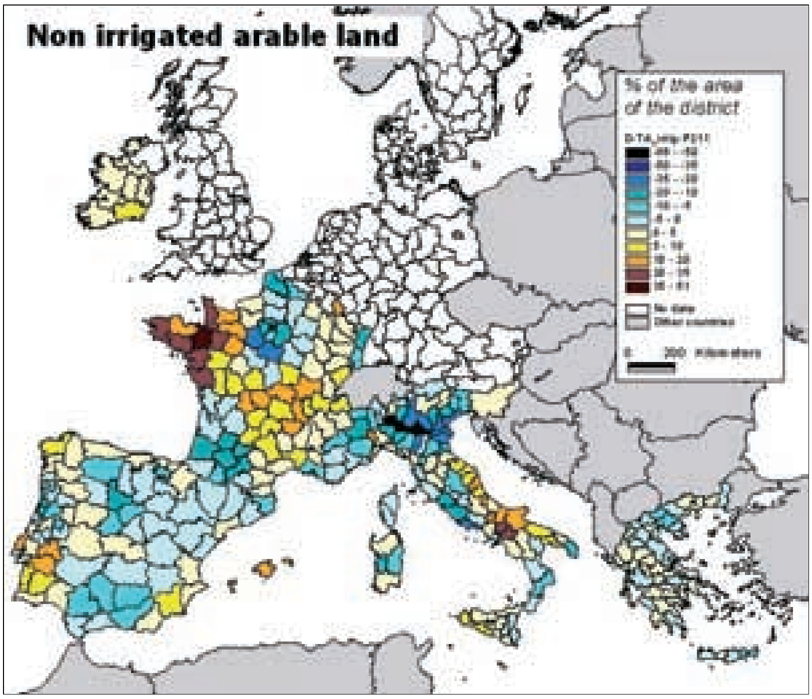
5.4.3.1 Comparison of arable lands
Item 211 of CLC, 'non-irrigated arable land', does not have any direct equivalent in Structure. This comparison was carried out with arable lands (item D) from which the irrigated arable land category was subtracted (Map 3). The area of *'irrigated arable lands'* (Structure) was calculated from *'cultivated area irrigated at least once a year'* (I/03/b) minus 'fruit and berry plantations' (I/03/b/8), 'citrus fruits irrigated at least once during the year'' (I/03/b/9) and 'vines irrigated at least once during the year' (I/03/b/10).

'Irrigable areas' (I/03/a) are not taken into account because they do not correspond to the definition of the item 212 of CLC: 'crops irrigated permanently or periodically, using a permanent infrastructure *(irrigation channels, drainage network). Most of these crops could not be cultivated without an artificial water supply. Does not include sporadically irrigated land'*.

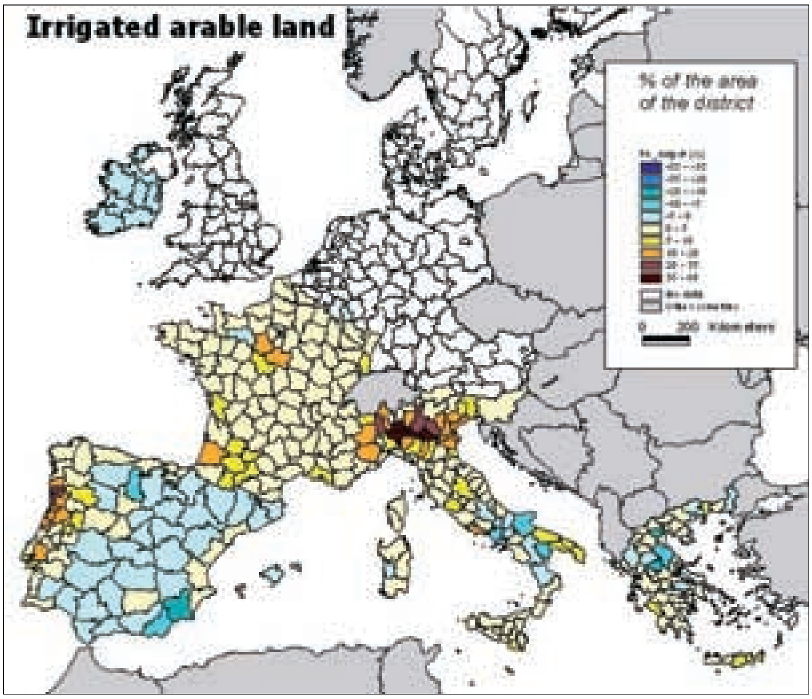
Results show strong shortfall in arable lands not irrigated for the Structure survey in the plain of the Po. These important differences are also found in the Parisian basin and in Aquitaine. The positive differences concern mainly the west of France and a part of Italy.

The comparison of irrigated arable lands (Map 4) constitutes, for numerous districts, the negative image of the previous one. Brittany, the south of the Paris basin, Piedmont, the plain of the Po, the south of Italy, Aquitaine and the centre of Portugal showed less irrigated land in CLC than in Structure. Evidently, CLC has a systematic

Map 3. **Comparison of non irrigated arable lands**



Map 4. **Comparison of irrigated arable lands**



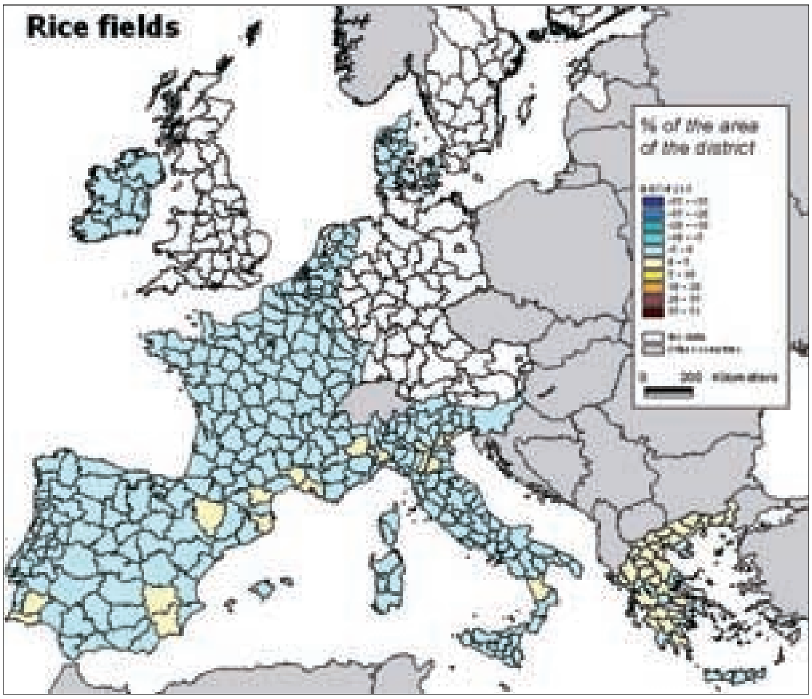
tendency to underestimate this item. A partial explanation for this lies in the definition of the item 212. It includes in its definition that 'sprinkler irrigation does not have to be taken into account, only irrigation by immersion techniques or flows should be considered'. The land classified as irrigated in CLC is that for which the infrastructure is visible on satellite images, i.e. the surfaces equipped with heavy infrastructures. This can explain, for example, why the Beauce is under-represented.

Another part of the explanation lies in the date of the source images of CLC. The comparison of the rice plantations reveals only weak differences (Map 5). However, surfaces are generally not very extensive (less than 1 % of the surface of the district), except for Piedmont, Camargue, the south of Spain and Portugal. In the northern half of Europe, the aberrant negative differences (presence of rice fields in CLC) involve errors of photo-interpretation or of coding.

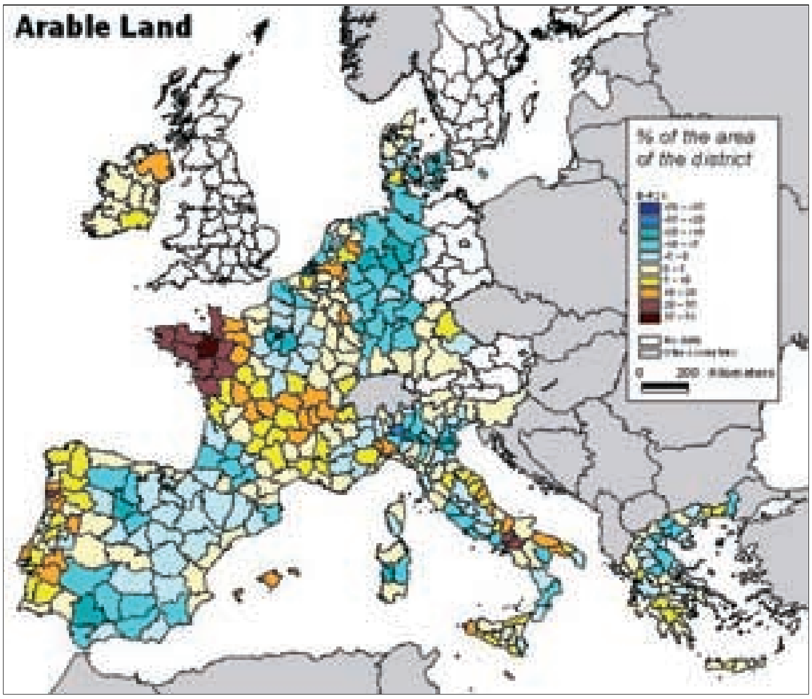
Comparisons on the one hand between irrigated arable lands and on the other hand between non-irrigated arable lands are not very conclusive. The criteria are how to distinguish between the, vary between the two datasets. These two classes have therefore been aggregated. Rice fields have been included. Map 6 gives the comparison of the arable lands on a more aggregated level, defined by item 21 of CLC and by class D of the Structure survey.

The results clean up a part of the differences observed on Map 3 and Map 4. Positive as well as negative differences are reduced for the plain of the Po and the Parisian basin. The comparison remains less than satisfactory for Brittany, Southern Italy and Portugal. On the one hand a large part of the arable area falls in the mixed CLC classes and, in particular, for the areas with small agricultural plots. On the other hand the CLC class 'arable lands' includes by definition a part of the permanent crops and induces biases in the comparison. The definition of the item 211 includes objects classified differently in the Structure survey: *'cereals, legumes, fodder crops, root crops and fallow land. Includes flowers and trees (nurseries) and vegetables, whether open field or under plastic or glass (includes market gardening). Includes aromatic, medicinal and culinary plants. Does not include permanent pasture'.*

Comparison of rice fields — Map 5.



Comparison of arable lands — Map 6.



5.4.3.2 Comparison of permanent crops
Comparisons here give satisfactory results. Differences vary in general between –5 % and +5 %. Some isolated districts show more marked differences: south of Portugal (Map 7), the mouth of the Po (Map 8) and the southern end of Italy (Map 9).

As with rice fields, the negative differences observed in three-quarters of the north of Europe for olive groves are due to errors ascribable to CLC. For this reason and the fact that G/02 (citrus plantations) and G/06 items (others) have no direct correspondence in CLC nomenclature, it was necessary to compare permanent crops with a more aggregated level, namely item 22 of CLC and the heading G of Structure. Class E (kitchen garden), moreover, was included in this aggregate, because it is small low and does not have a direct correspondence in the CLC nomenclature (included in 242). Differences (Map 10) are mostly included in the interval [-5 %; +5 %]. About thirty districts show larger positive or negative differences up to at most 15 % to 20 %.

G/05 (nurseries) and G/07 headings (permanent crops under glass) of Structure could have been aggregated to the heading 'arable lands'. Indeed, CLC count them together. Their negligible surface within the districts (less than 0.1 % on average) did not make this transfer essential.

5.4.3.3 Comparison of the meadows
For pastures (Map 11), only 16 districts record negative differences (between –17 % and –0.2 %). Differences are strongly positive in the Cevennes, in Sardinia, Italian Alps, the Abruzzi and in Extremadura (from +30 % to +50 %). Pastures areas are especially lacking within CLC in the south of Europe. The photo-interpretation has difficulty in distinguishing them from other surfaces, such as semi-natural areas.

5.4.3.4 Comparison of agro-forestry land
For agro-forestry land (Map 12) and annual crops grown in association with permanent crops (Map 13) comparisons give suitable results. Agro-forestry spaces in Extremadura are over-estimated by CLC. As for pastures, CLC has difficulty with the transition areas between agricultural and natural spaces. For Spain, Map 12 constitutes the negative of Map 11.

5.4.3.5. Comparison of annual crops grow in association with permanent crops
Heterogeneous agricultural area types constitute a characteristic of the mixed CLC classes (item 24). Their comparison with Structure classes is problematic because Structure clearly distinguishes the various types of culture. Only comparison between the item 241 of CLC (annual crops associated with permanent crops') with the I/05/b station of the Structure survey (combined crops: annual crops — permanent crops') can be carried out (Map 13), but even then only for certain countries. This question was in fact optional in the Structure survey in 1989/90.

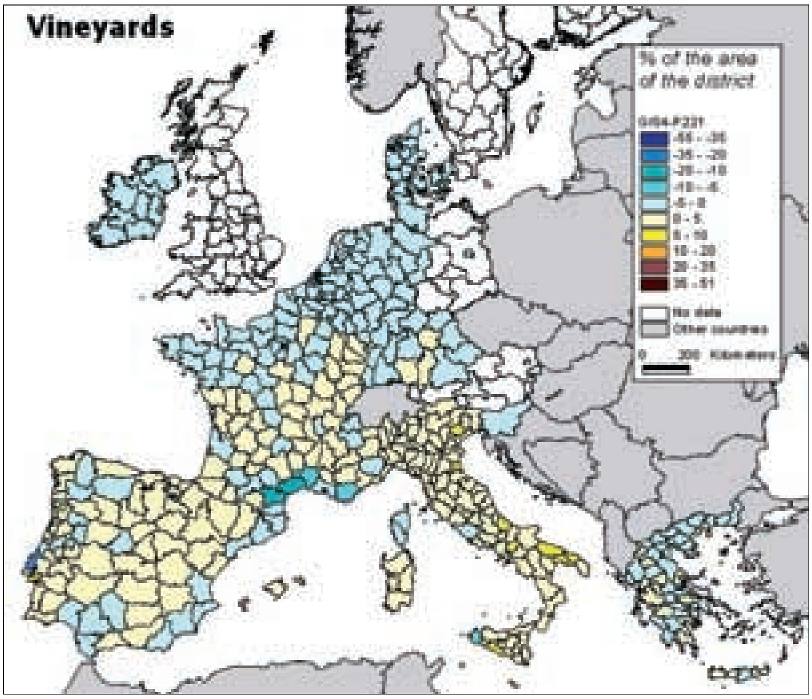
5.4.4. Pro osals for im rovement
5.4.4.1. Arable land aggregate
Photo-interpretation of the satellite images at the time of the development of CLC does not distinguish permanent pasture from temporary pasture. In the Structure survey definition, the difference is that a temporary pasture has a lifetime lower than 5 years. Class D/18 (forage plants), under heading arable lands' of Structure was therefore aggregated with class F (permanent meadows and pastures').

Differences (Map 14) show clear improvement of the results (cf. Map 6) in particular in Brittany, in Portugal and in Italy on the Adriatic coast.

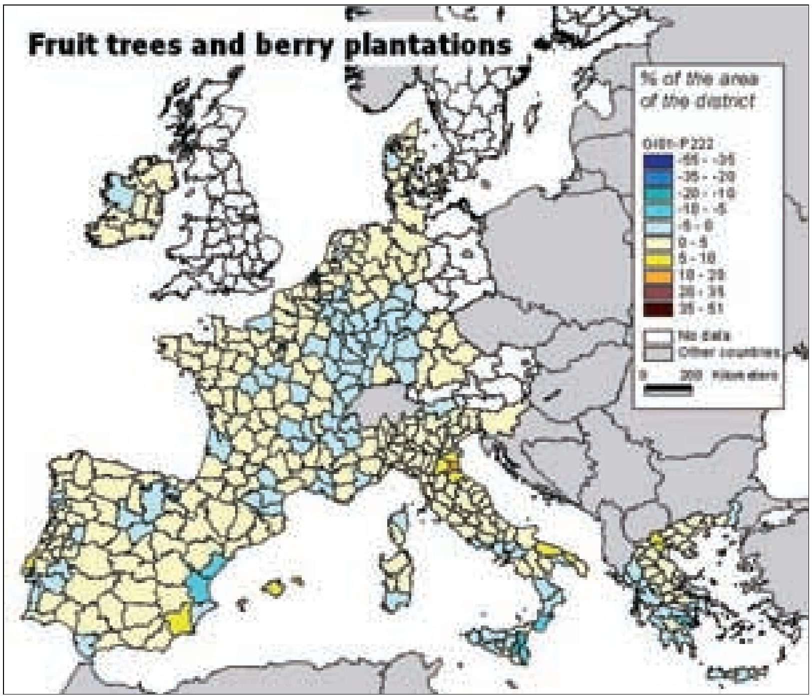
The positive differences concern mainly Portugal and the *massif central* region in France but in rather small proportions (+10 % to +15 %) They reach more than 21 % only for two districts in the south of Italy in Campania.

The negative gaps widen in the plain of the Po, Northern Germany and Denmark where they vary between –40 % and –10 %. These are significant differences and can be interpreted that too many fodder plants were transferred towards permanent pastures. It is advisable therefore to refine this reassignment subsequently.

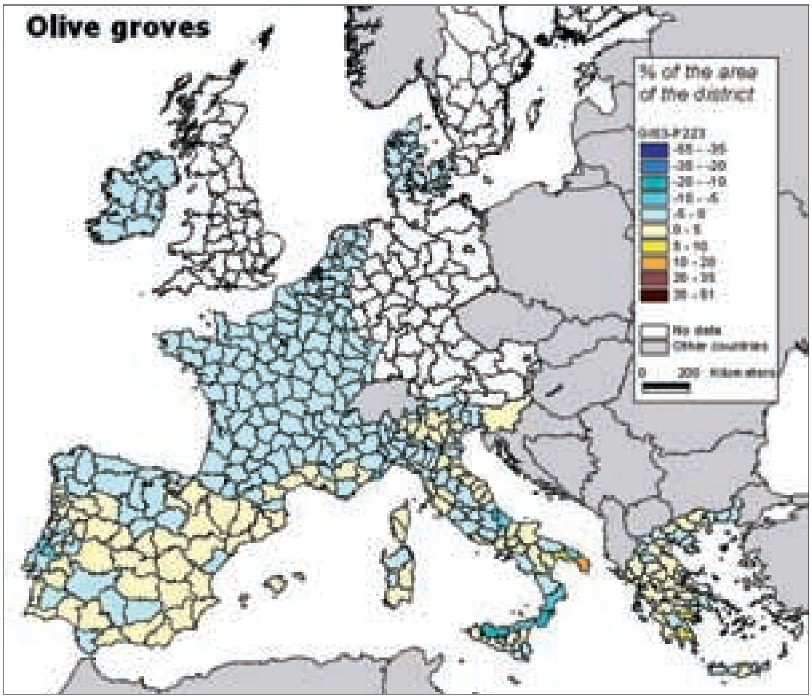
Map 7.	Comparison of vineyards



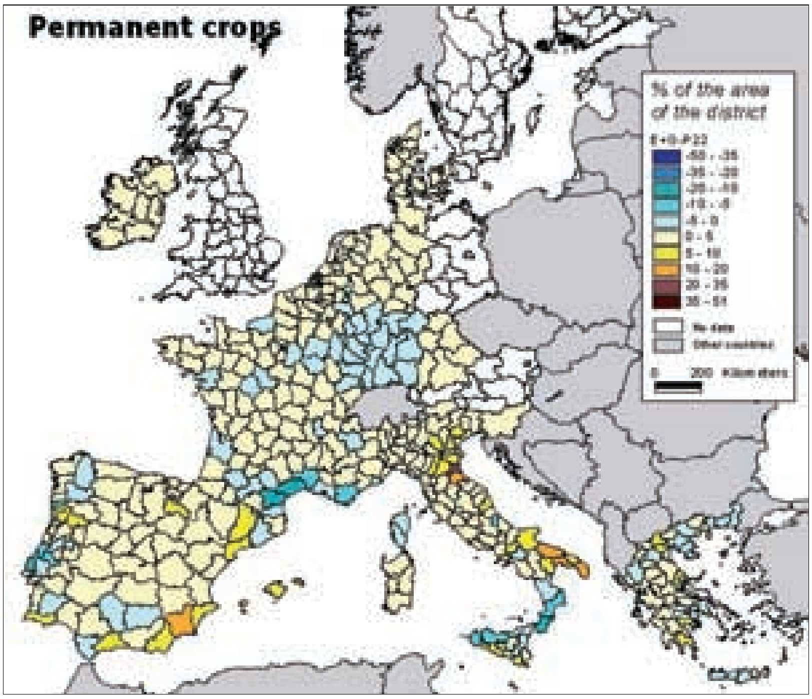
Map 8.	Comparison of the fruit trees and soft fruit



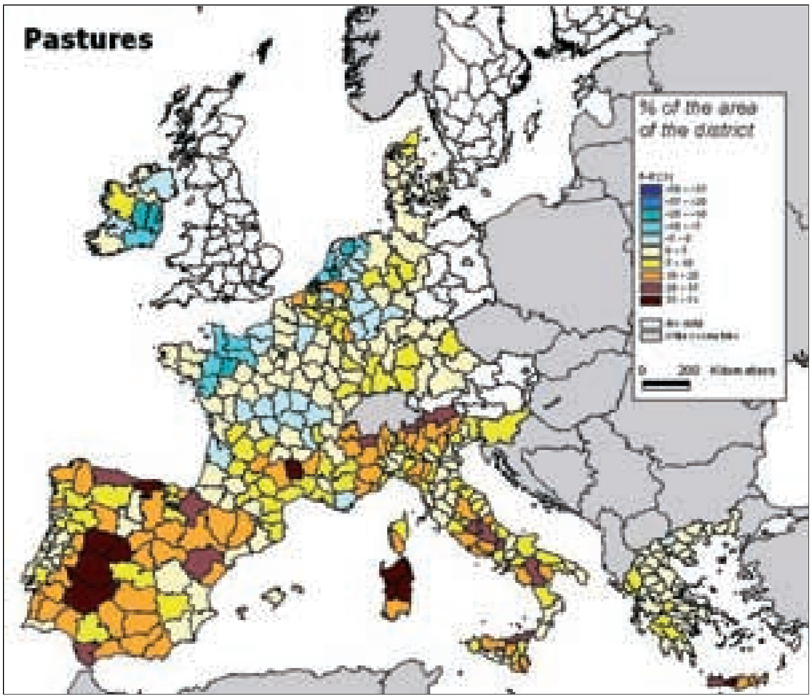
Comparison of the olive groves Map 9.



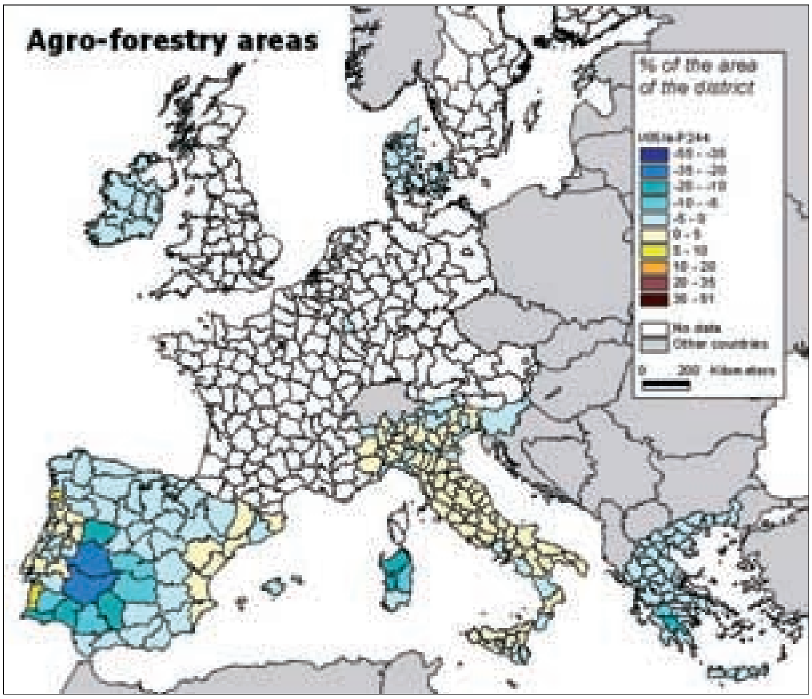
Comparison of the permanent crops Map 10.



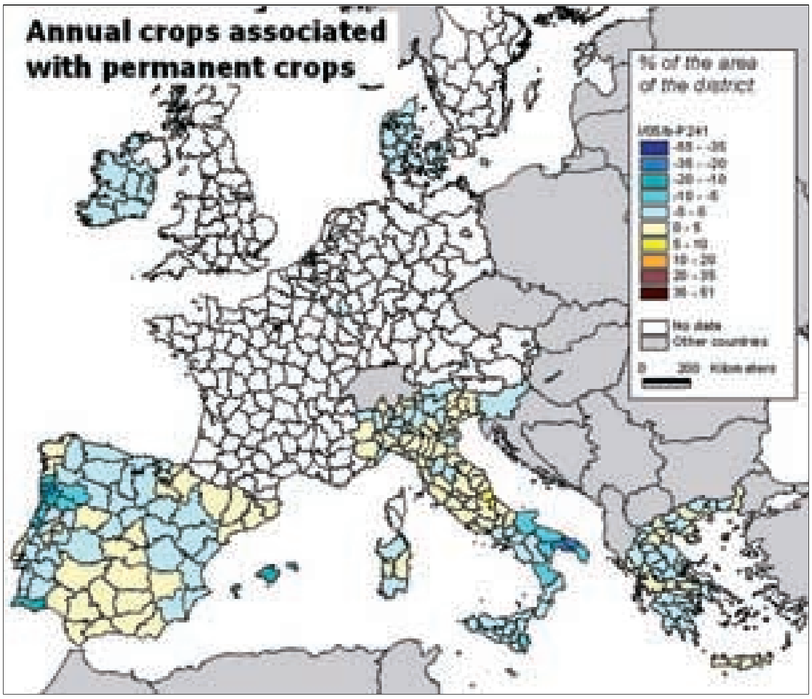
Map 11.	Comparison of pastures



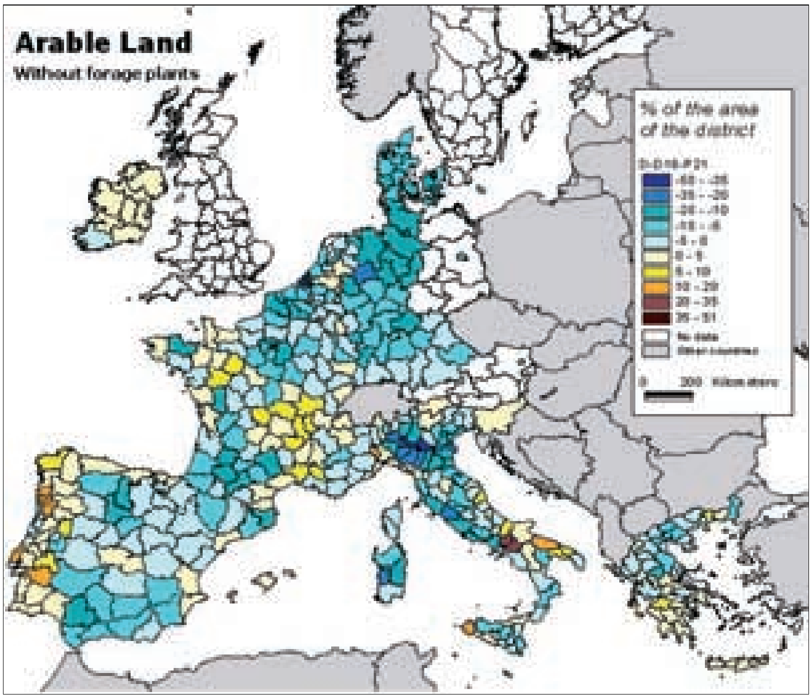
Map 12.	Comparison of agro-forestery



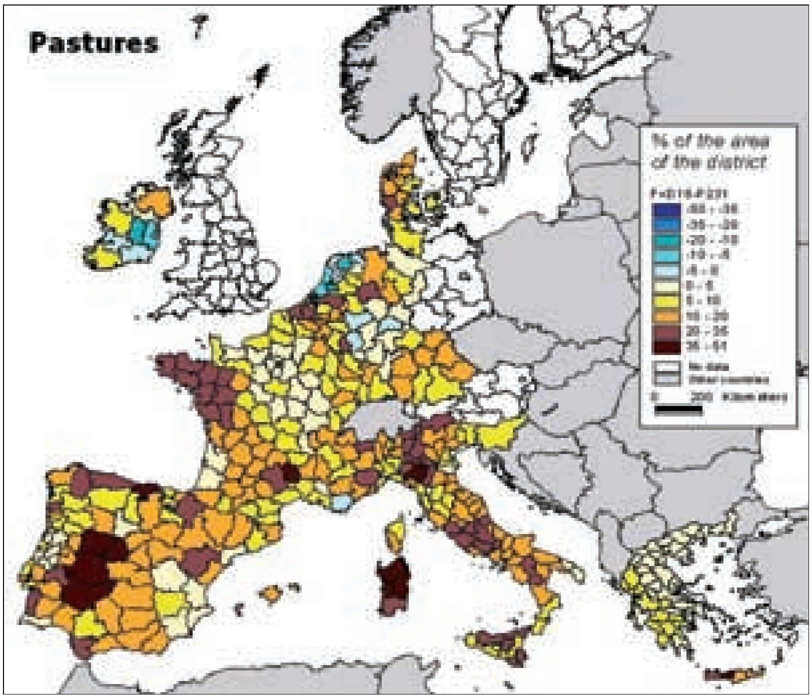
Comparison of annual crops associated to the permanent crops — Map 13.



Comparison of the arable lands not taking the D/18 station into account — Map 14.



Map 15. **Comparison of the meadows taking into account the D/18 station**



5.4.4.2. Pastures

Transfer of class D/18 from arable lands' (D) towards meadows and pastures' (F) reveals, on the other hand, the problem of pastures. The comparison of this aggregate with the pastures' of CLC accentuates the positive differences in Brittany, in Flanders and in Denmark (Map 15). However, distribution of CLC class 24 *heterogeneous agricultural areas*' (Map 16) explains partly these positive differences. The annual crops (meadows, arable lands) can occupy up to 75 % of the surface of these areas. It is probable that in regions where plots are very fragmented such as Brittany areas of pasture reach some 75 %. It is for this reason, that as a first approximation a CLC aggregate including pastures' (item 231) and heterogeneous agricultural areas' (item 24) was constituted and was named pasture 1' Comparison with aggregate pasture' of Structure (items F and D/18) are shown in Map 18. The surfaces counted as permanent meadows and pastures' in Structure are not easily identifiable by photo-interpretation in the semi-natural type scrub and/or herbaceous vegetation associations' (item 32 of CLC) and sparsely vegetated areas' (item 333). They tend to be classified as natural by CLC. This classification of areas can explain in part the differences on pastures observed in the Mediterranean area and in the Alps.

It is the reason for which, in a second approach, a CLC aggregate including meadows' (item 231), scrub and/or herbaceous vegetation associations' (item 32) and sparsely vegetated areas' (item 333) was created and named pastures 2'. The comparison with the aggregate pastures' of Structure (items F and D/18) is shown in Map 19.

These various regroupings show that the reallocation of heterogeneous classes to pastures explains a part of the difference between Structure survey and CLC. In certain cases, the allocation of a part of these classes (according to rules to be defined) would be sufficient. In other cases, this transfer can appear insufficient. Classes like semi-natural spaces have then also to be partially incorporated.

Stressing delocalised farming systems (transhumance, tele-control of major crops) can also serve as a covariable to a reassignment of these types of surface.

Links of aggregated nomenclatures Table 5.

Aggregate	CORINE land cover	Structure
Arable lands	21	D-d/18
Permanent crops	22	G+E
Pastures (1)	23+24	F + d/18
Pastures (2)	23+31+332	F+d/18

5.5. Possible future work

5.5.1. Integration of time factor

The comparison of nomenclatures and the analysis of the area differences between both databases can be refined by the integration of the factor time. The integration of the chrono-geographical coordinates of the satellite images sources of CLC should enable:

1) Identifying districts for which CLC's image interpretation is indeed 1990 ± 1 year;
2) using the Structure data intermediate surveys (1985, 1987, 1993 or 1995) that correspond closely to the date of the satellite images. This adjustment to one of the two bases will mitigate the effect of time.

5.5.2. Integration of s atial units

Some districts include urban, semi-urban and rural areas. The comparison between Structure survey data and CLC appears particularly delicate here. Because of fragmentation the threshold of 25 ha of CLC leads to some agriculture land appearing as urban and vice versa. The use of smaller geographical units would make the process of splitting up heterogeneous polygons more precise and more reliable. The use of neighbourhood analysis techniques can cover part of these interpretation defects.

Structure survey data is collected at the NUTS level 5. For reasons of statistical confidentiality, they are sent to Eurostat at district level. By taking into account data at the NUTS level 5, more homogeneous geographical units regarding land cover and land use could be dealt with. Ireland, for which data at the NUTS level 5 is available, will serve as a test area for more precise reassignment methods.

5.5.3. Cases where CORINE land cover underestimates AA

In some districts, CLC agricultural surfaces are lower than Structure survey AA. In this case, AA is taken as value of reference and the later stages consist in upgrading certain heterogeneous CLC polygons (at least partially) in arable lands', permanent crops', or meadows'.

The previous examples show that, by adding to agricultural surfaces certain heterogeneous CLC classes, it is possible to fill the gaps with AA, in particular for arable lands and permanent crops. The pasture case is more complex and, in certain areas, it will be necessary to reallocate a part of the Structure survey meadows to the semi-natural areas'. The latter stages consist of identifying relevant covariables to define reassignment rules. Some approaches such as neighbourhood analysis of polygons can be considered. Other tools exist at Commission level: digital terrain models, soil maps, climatological maps, agricultural plot maps of IACS (37).

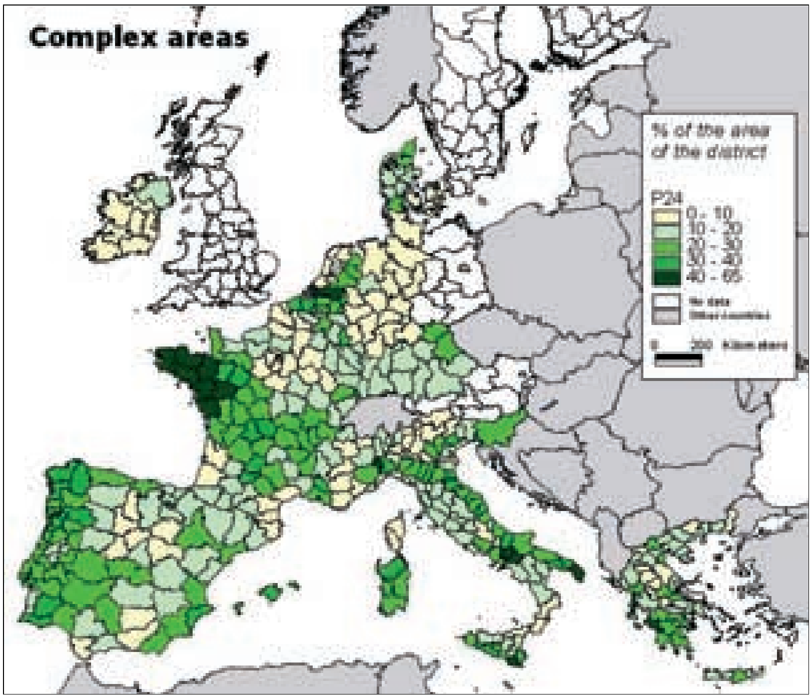
5.5.4. Cases where AA is lower than agricultural areas in CORINE land cover

In some districts, AA is lower than the sum of the surfaces of polygons classified as agricultural areas' in CLC. This situation is more delicate because it contains two explanatory elements but each of a different nature.

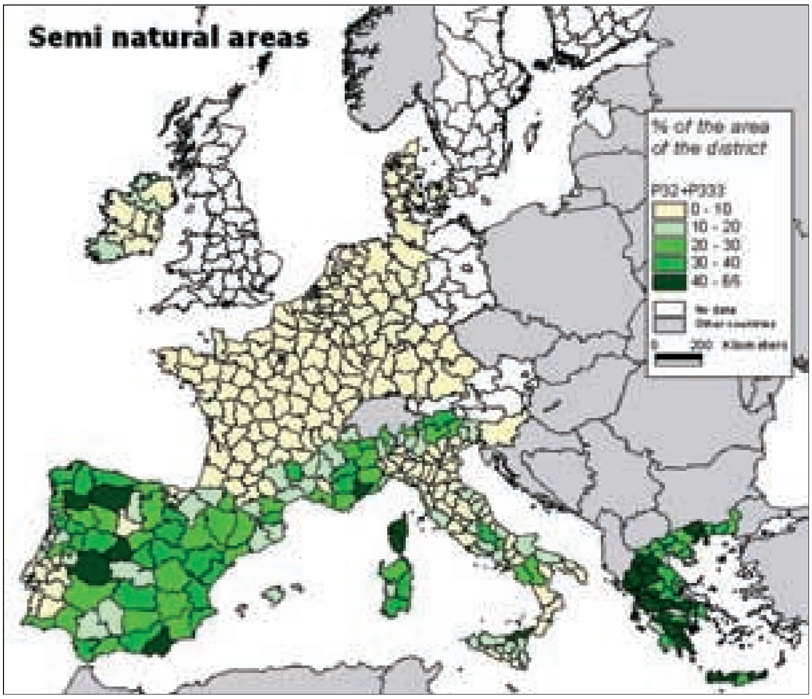
On the one hand, because of the threshold of 25 ha of CLC's polygons, non-agricultural areas are included in polygons classified as agricultural area by photo-interpretation. This can happen for semi-urban areas, regions with fragmented plots, of thickets or of small wooded enclaves, etc. On the other land AA does not reflect the total sum of agricultural areas. The smallest holdings are not surveyed. Even if this ones concerns small areas, it can be of importance locally. In certain, well defined cases, areas are not surveyed because they do not belong to a single specific owner. This is the case in particular of mountain pastures and of communal meadows.

(37) See paper of E. Willems & al., infra

Map 16. **Heterogeneous areas (item 24)**

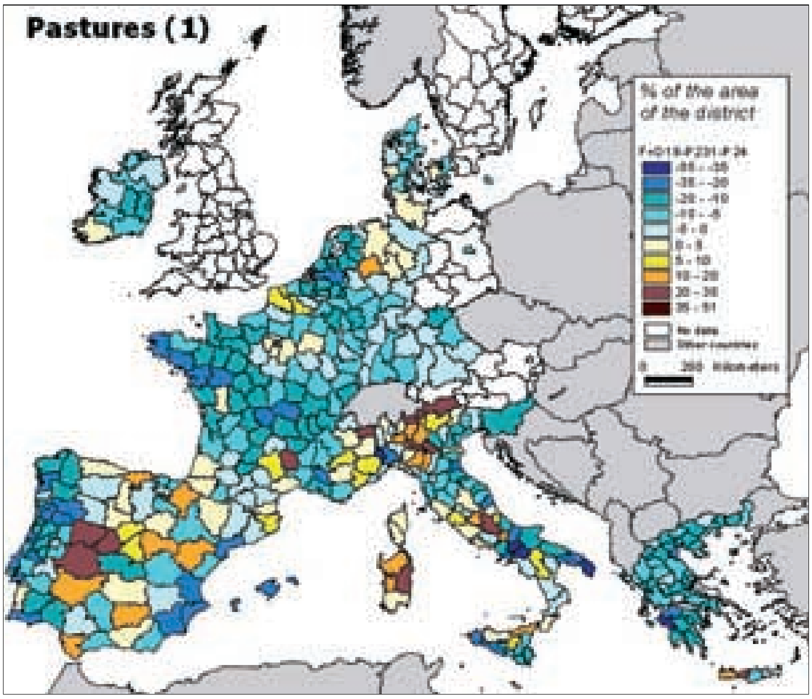


Map 17. **Semi-natural areas (items 32+333)**

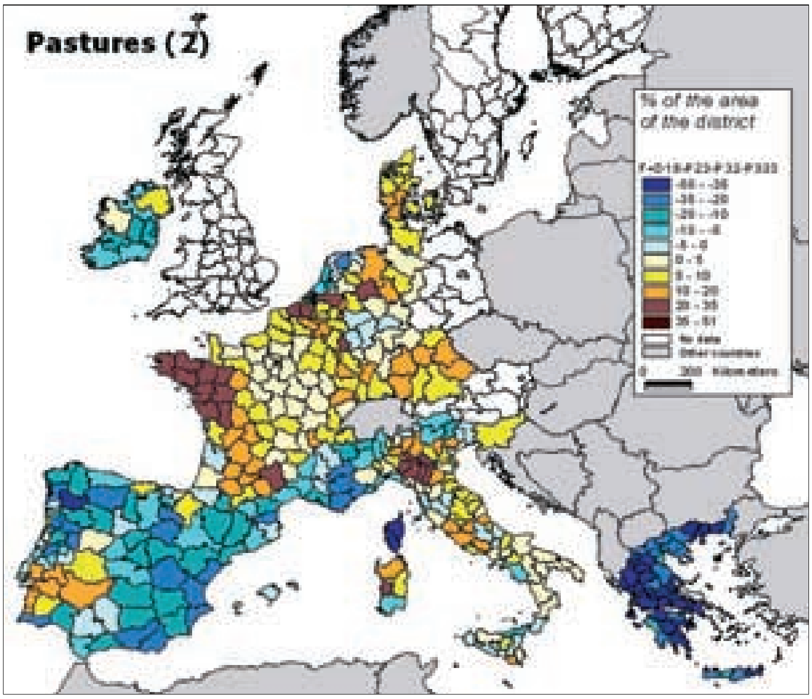


Finally, another element of explanation is the headquarters of the holding. By definition, in the Structure survey, all the land of a holding is assigned to the commune' where the headquarter is situated. This rule, gives the ever increasing size of holdings, can create differences in the calculation of AA by district. For all these reasons, the reassignment of AA to CLC polygons will require the detail of the geographical context and an analysis of the surrounding situation to be taken into account. The introduction of probabilities of heterogeneity could improve the precision of the process.

Map 15 – Map 16 Map 18.



Map 15 – Map 17 Map 19.



6. Using CORINE land cover to map population density

Javier Gallego, Steve Peedell (JRC)

6.1. Introduction

Population density data are available to the European Commission (EC) at the level of the commune (NUTS5). The size of communes is very heterogeneous across the EU. Hence this level of spatial resolution may be insufficient in many cases for planning or modelling purposes or to assess the impact of EU policies. In some countries, as in France, where most communes have a rather small area (approx. 15 km^2 in the average), the resolution may be sufficient, but it is clearly insufficient in other countries where the communes tend to be larger.

CORINE land cover (CLC) gives useful geo-referenced information for disaggregation. This geographic database provides information that is spatially more detailed than the commune limits. A certain commune may contain for example one part of dense urban nucleus, agricultural land with some sparse population, and natural vegetation areas with very little or no population. The objective is to disaggregate population data, imputing different densities to different land cover categories. One possible approach to tackle this problem might be based on the EM algorithm (Dempster, 1977, Ambroise and Govaert, 1998), but the underlying parametric model may be debatable. Here we test a more empirical method.

6.2. Available data

The latest population data available to the Commission at commune level correspond to 1991, therefore the commune boundaries of the SABE database, version 1991 have been used (http://www.megrin.org/SABE/Sabe.html). The CLC data has been used as raster data (1 ha pixel).

For the computations hereafter, in the communes in which CORINE land cover is partially missing, the average population density of the commune is assigned to the areas with missing data and the general rule is appled to the rest. CLC is not available for Sweden and Finland, and the SABE commune boundaries are not available for Scotland. These areas have been excluded from the study (the column UK corresponds to England, Wales and Northern Ireland). The French département Seine St Denis' has been also excluded because of code errors in the version of CLC used.

6.2.1. Communes without urban area in CORINE land cover

A large number of communes appear as not having any urban area (class 1.1) according to CORINE land cover.

Table 6. **Communes with no urban area in CORINE land cover**

	<1000 inhab		1000-5000 inhab		>5000 inhab		Total	
	N communes	population (*1000)	N communes	population (*1000)	N communes	population (*1000)	N communes	population (*1000)
AT	488	301	772	1 353	5	40	1 265	1 695
DE	1 860	656	185	268	0	0	2 045	924
ES	1 269	247	151	333	22	157	1 442	737
FR	12 860	2 699	80	108	2	11	12 942	2 819
GR	2 841	865	110	164	5	77	2 956	1 107
IE	1 973	767	109	143	0	0	2 082	910
IT	886	410	357	617	6	48	1 249	1 074
LU	12	6	1	1	0	0	13	7
NL	2	1	2	2	0	0	4	4
PT	1 585	742	624	1 107	10	63	2 219	1 912
UK	106	79	509	878	1	5	616	961
Total	23 882	6 773	2 900	4 975	51	403	26 833	12 151

There are several possible explanations for the absence of urban area in CORINE land cover for a commune:

- The area of the built area is below the CORINE land cover threshold (25 ha). This is the explanation for the great majority of cases.
- Geometric inconsistencies between CORINE land cover and SABE. For example all the urban area of a commune may be in a linear development along the coast, but appears to fall outside the commune because of location inaccuracy.
- Errors in the data (CORINE land cover or SABE). In a few cases, crossing population data with commune borders can help focus attention on inconsistencies between land cover and population density. Some of them might turn out to be errors in datasets. Anomalies have been detected for several communes of Seine St Denis in the outskirts of Paris. Although this area has been excluded, similar errors may still remain in the data.

The 12.1 million inhabitants in this category of commune represents less than 4 % of the 327 million inhabitants of the studied area, but it is a significant part of the rural population; hence this case must be considered specifically. The most populated communes without CLC urban classes correspond to areas with very scattered settlements and no major error could be detected in CLC.

6.3. Modified areal weighting with given coefficients

Population data can be disaggregated with the help of CORINE land cover assuming that the ratio between the population density of two land cover classes is the same for any commune. This a simplified version of modified areal weighting. We can initially assume that the coefficients are known.

We call
X_m: population in commune m.
S_{cm}: area of land cover type c in commune m.
Y_{cm}: density of population for land cover type c in commune m. Inside each commune Y_{cm}: assumed to be proportional to given coefficients U_c for each land cover type.

$$Y_{cm} = U_c W_m \quad (1)$$

W_m: adjustment factor to ensure that the total population in each commune matches the administrative data.

therefore $X_m = \sum_c S_{cm} Y_{cm} \quad (2)$

and it was assumed that

$$X_m = \sum_c S_{cm} U_c W_m \Rightarrow W_m = \frac{X_m}{\sum_c S_{cm} U_c} \quad (3)$$

Hence the densities were computed in a first approach as

$$Y_{cm} = U_c \frac{X_m}{\sum_c S_{cm} U_c} \quad (4)$$

This disaggregation has been carried out with an initial set of coefficients provided by EEA for an aggregated CORINE land cover nomenclature (Table 7). For the other classes we assume there is no population.

Table 7. **Grouped CORINE land cover classes and initial coefficients**

grouped class	Initial coefficient U_c	CORINE Class	Label
1	32	111	Continuous urban fabric
2	25	112	Discontinuous urban fabric
3	1	121	Industrial or commercial units
4	1	122, 123, 124	Road and rail networks, ports, airports
5	1	141, 142	Green urban areas, sport and leisure facilities
6	3	211	Non-irrigated arable land
7	3	212	Permanently irrigated land
8	1	213	Rice fields
9	5	22	Permanent crops
10	3	231	Pastures
11	5	241	Annual and permanent crops associated
12	5	242	Complex cultivation patterns
13	3	243	Agriculture, with natural vegetation
14	1	244	Agro-forestry areas
15	1	31, 324	Forest and woodland
16	1	32, 41	Other natural vegetation

6.4. Search for weighting coefficients

Assuming an approximately homogeneous behaviour of W_m, the expression

$$X_m = \sum_c S_{cm} U_c W_m \quad (5)$$

can be interpreted as

$$X_m = \sum_c S_{cm} U_c W + \varepsilon_m \quad (6)$$

where the residuals ε_m are small, i.e. we can write it as a simple linear regression. A simple regression gives completely unacceptable coefficients with several negative values and very high values for classes such as green urban or sport areas. This phenomenon confirms that the approach is too simplistic. In fact the weighting coefficients are not the same for all communes. Separating urban and rural areas is not sufficient to make coefficients homogeneous.

6.4.1. Disaggregation of regional data to assess the validity of weighting coefficients

The best way to assess the disaggregation of the commune populations is comparing the results with data at infra-commune level, but, at the current stage, such data is generally not available to the EC. One possible way to overcome this limitation would be:

- disaggregate regional data with CLC using a set of coefficients;
- reaggregate the attributed population on commune basis.;
- compare with the known population per commune and compute a disagreement indicator;
- modify the coefficients to reduce the disagreement.

X_r is the population in region r.
S_{cr} is the area of land cover type c in region r.
Y_{cr} is the density of population we attribute to land cover type c in region r.
W_r is an adjustment factor to ensure that the total population in each region coincides with the known total.

Thus,

$$X_r = \sum_c S_{cr} U_c W_r$$ => The densities

attributed are $$Y_{cr} = U_c \frac{X_r}{\sum_c S_{cr} U_c} \quad (7)$$

and the population attributed to each commune m in region r is

$$X_m^* = \sum_c S_{cm} Y_{cr} \quad (8)$$

Computing the ratio between the attributed population and the known population

$$\psi_m = \frac{X_m^*}{X_m} \quad (9)$$

or an aggregated difference between attributed and real population at regional or European level
.

$$\delta_r = \sum_{m \in r} \left| X_m^* - X_m \right| \quad \delta = \sum_{m} \left| X_m^* - X_m \right| \quad (10)$$

It can be checked that $\delta_r \leq 2X_r$.

The maximum value of the deviation would happen when all the population is attributed to communes with real population 0.

For each region we can compute

$$\rho_{cr} = corr\left(\psi_m, \frac{S_{cm}}{S_m}\right).$$

If the correlation $\rho_{cr} > 0$, this would mean that a too high population has been generally attributed to communes where the CORINE land cover class *c* has a high proportion. We can try to compensate this tendency by reducing the coefficient for this region and land cover:

$$U'_{cr} = U_c\left(1 - \frac{\rho_{cr} \times \delta_r}{2 \times X_r}\right)$$

The coefficient U'_{cr} raises when the correlation is negative.

The term $\frac{\delta_r}{2 \times X_r}$ is introduced to moderate the modification of the coefficient when the attribution is close to the real population.

The coefficient adjustment can be repeated in an iterative way until the difference indicator δ becomes stable. To avoid some extreme effects on the coefficients, limits have been introduced so that the ratio between the maximum and minimum density in a commune is constrained not to exceed 10 000.

6.4.2. A lication with 16 grou ed classes

The total deviation with the same coefficient to all CLC classes gives an indicator of 322×10^6 for a total population of 321×10^6 inhabitants. This corresponds to the inaccuracy of representing the average population density by NUTS2 compared with the representation by communes.

Deviation indicator with uniform disaggregation for NUTS2 regions Table 8.

	population	deviation no weight	ratio %
AT	7 796	6 398	82
BE	9 979	6 640	67
DE	59 940	49 036	82
DK	5 146	4 574	89
ES	37 182	45 991	124
FR	55 072	63 794	116
GR	9 692	11 846	122
IE	3 364	3 689	110
IT	56 705	54 779	97
LU	383	386	101
NL	14 950	12 349	83
PT	9 371	10 690	114
UK	51 468	52 583	102
total	**321 047**	**322 755**	**101**

With the coefficients in Table 7, the disagreement indicator goes down to around 241×10^6. The application of the algorithm described above improves the deviation, so that results become stable at slightly under 137×10^6.

Looking at the average values of the coefficients (Table 9) by region, we observe:

- the three arable land classes have similar average coefficients;
- the coefficients for permanent crops are close to the coefficients of two of the complex classes (2.4);
- agro-forestry presents similar coefficients to those of forest and semi-natural vegetation.

Figure 2. **Total deviation d of population attribution with 16 CORINE groups**

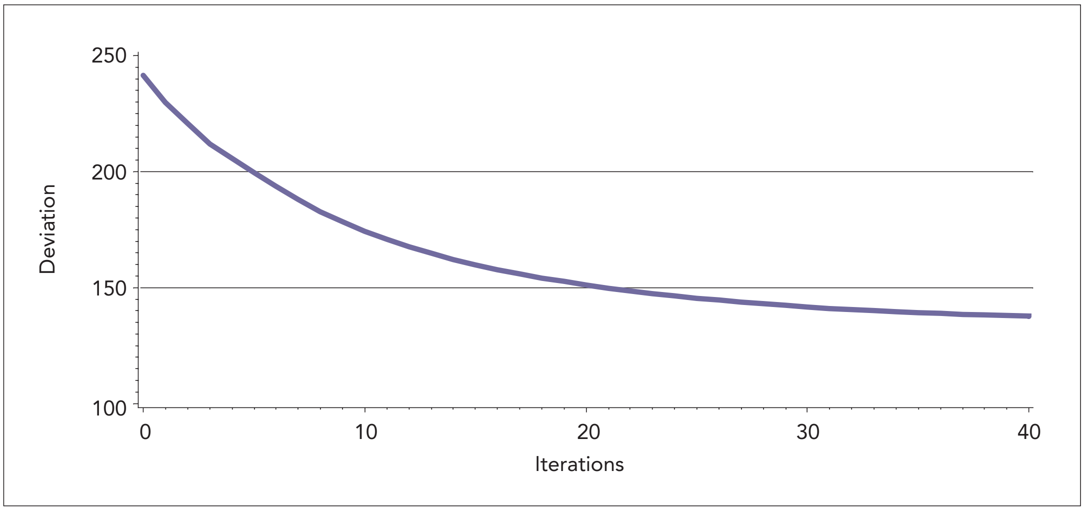


Table 9. **Average coefficients after 40 iterations**

Land Use	
Urban dense	198.41
Urban discontinuous	176.56
Industrial and commercial	9.56
Transport	3.71
Green urban	3.42
Arable non irrigated	2.98
Arable irrigated	3.36
Rice	2.90
Permanent crops	4.95
Pastures	2.99
Arable with permanent crops	5.40
Complex agricultural	5.92
Agricultural and natural	3.24
Agroforestry	0.86
Forest	0.94
Natural vegetation	0.63

6.4.3. Stratification and further grou ing of CORINE land cover classes.

The ratio between the density in different land cover classes is not the same in densely populated areas and in more rural areas. Therefore communes have been stratified in each region applying a very simple criterion:

1. Dense communes: population density higher than twice the average density in its NUTS2 region;
2. Less dense: population density lower than twice the average density in its NUTS2 region, but urban area reported in CORINE land cover;
3. No urban: No urban area reported in CORINE land cover.

A cluster analysis of the 16 CLC classes based on the table of coefficients by NUTS2 after 40 iterations also gives some indication of the CORINE land cover classes that have a similar behaviour for different region typologies. Taking into account the results of cluster analysis and the meaning of classes, the classes have been regrouped into 8 (Table 10) and the iterative algorithm has been rerun. The classes Industrial and commercial', transport' and green urban and sport facilities' have been aggregated with the class urban discontinuous' with fixed weights inspired in Table 9:

Urban 2 = Urban discontinuous + 0.05 * Industrial_commercial + 0.02*transport + 0.02*green urban

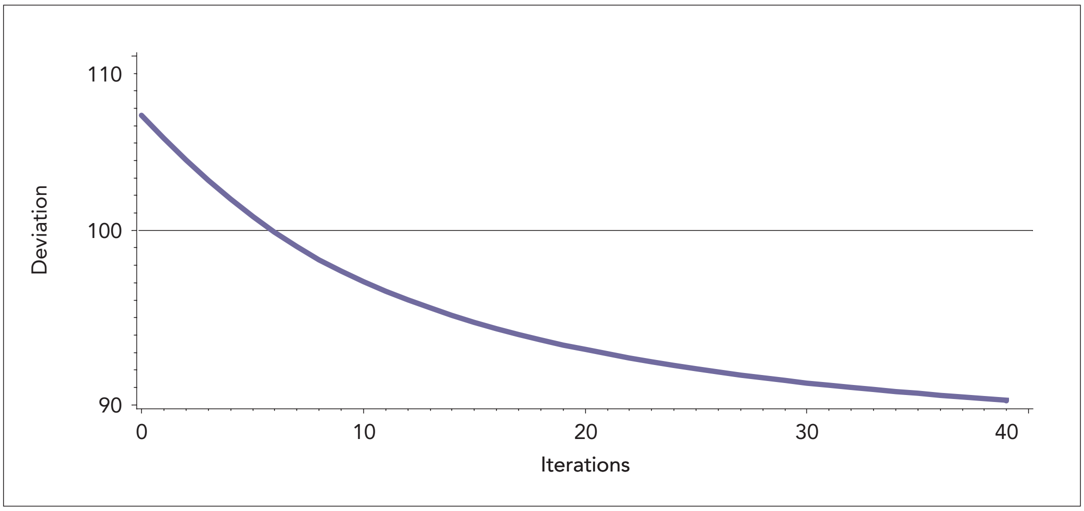
Therefore 6 classes have been kept for further analysis.

Applying the iteration algorithm with 3 strata and 6 CORINE land cover classes, a deviation indicator that approaches 90×10^6 is obtained.

New grouped classes Table 10.

grouped class	CORINE Class	Label
1	111	Continuous urban fabric
Urbdisc (2a)	112	Discontinuous urban fabric
ind (2b)	121	Industrial or commercial units
otha (2c)	122	Road and rail networks and associated land
otha (2c)	123	Port areas
otha (2c)	124	Airports
otha (2c)	141	Green urban areas
otha (2c)	142	Sport and leisure facilities
3	211	Non-irrigated arable land
3	212	Permanently irrigated land
3	213	Rice fields
4	221	Vineyards
4	222	Fruit trees and berry plantations
4	223	Olive growes
5	231	Pastures
4	241	Annual and permanent crops associated
4	242	Complex cultivation patterns
5	243	Agriculture, with natural vegetation
6	244	Agro-forestry areas
6	311	Broad leaved forest
6	312	Coniferous forest
6	313	Mixed forest
6	321	Natural grassland
6	322	Moors and heathland
6	323	Sclerophyllous vegetation
6	324	Transitional woodland-shrub

Total deviation of population attribution with 6 CORINE groups and 3 density strata per region Figure 3.



6.5. Suggested disaggregation

The algorithm gives weighting coefficients for each stratum in each NUTS2 region. For most regions the coefficients are similar, but some outliers appear. Clustering NUTS2 regions does not show a clear grouping linked with population settlement styles. Therefore a first disaggregation rule has been proposed based on a set of coefficients for each stratum that are the same for all NUTS2 regions.

The population density we attribute to land cover class *c* in commune *m* is computed as $Y_{cm} = U_c W_m$. The coefficients U_c keep their meaning if they are multiplied by any constant *K* and the coefficient W_m is divided by *K*. The values of U_c given in Table 11 correspond to a choice of *K* such that the median of W_m in each stratum is 1. They still cannot be interpreted as population density for land cover class *c*', but as median density for each land cover class in each stratum.

Table 11. **Disaggregation coefficients with 6 CLC classes and three strata of communes**

	Urban dense	Urban 2	Arable	Permanent crops and complex	Pastures	Forest & natural vegetation
Stratum 1	1 445.9	619.1	10.2	15.4	5.1	3.3
2	947.4	622.4	17.4	30.9	11.3	5.2
3			32.0	69.3	22.8	8.6

Table 12. **% of population in each CLC class with the suggested disaggregation**

	AT	BE	DE	DK	ES	FR	GR	IE	IT	LU	NL	PT	UK	All
Urban dense	11.8	5.7	2.5	10.1	54.1	13.2	33.2	8.4	23.8	14.8	85.3	18.3	16.1	21.9
Urban discontinuous	50.2	86.4	80.0	67.4	18.2	64.6	33.9	50.2	50.2	66.8	2.6	40.6	71.8	56.5
Indust. & commer.	0.2	0.6	0.6	0.3	0.6	0.8	0.3	0.4	0.7	0.7	0.4	0.4	0.3	0.6
Transport	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.1	0.1
Green urban	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.2	0.0	0.0	0.1	0.0	0.2	0.1
Arable non irrigated	8.6	2.4	8.1	16.5	6.2	7.1	4.4	4.1	7.9	2.5	3.1	6.1	4.8	6.7
Arable irrigated	0.0	0.0	0.0	0.0	2.2	0.0	2.7	0.0	0.5	0.0	0.0	0.1	0.0	0.4
Rice	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.3	0.0	0.1
Permanent crops	0.6	0.1	0.5	0.0	4.8	1.6	7.2	0.0	4.3	0.3	0.1	7.2	0.0	2.1
Pasture	6.6	0.7	2.2	0.2	0.3	2.5	0.0	28.9	0.3	1.4	3.7	0.0	5.1	2.4
Arable & perm crops	0.0	0.0	0.0	0.0	0.4	0.0	0.2	0.0	1.3	0.3	0.0	10.0	0.0	0.6
Complex agricultural	11.3	3.1	3.4	2.9	6.3	6.7	8.5	2.1	5.6	9.1	3.8	5.7	0.8	4.7
Agric. & natural veg.	0.5	0.4	0.4	1.4	1.3	0.8	2.3	3.0	1.6	1.4	0.3	3.7	0.1	0.9
Agroforestry	0.0	0.0	0.0	0.0	0.5	0.0	0.2	0.0	0.0	0.0	0.0	0.4	0.0	0.1
Forest	8.7	0.5	2.2	0.8	2.7	2.0	3.1	1.3	2.8	2.4	0.4	5.8	0.4	2.1
Natural vegetation	1.5	0.0	0.0	0.1	2.0	0.3	4.0	1.3	0.6	0.0	0.1	1.3	0.4	0.7

Table 12 gives the % of population that has been attributed to each CLC class in each country. The meaning of this table might need some clarification. For example 28.9' in the cell Ireland pasture does not mean that 28.9 % of the Irish population live in pasture fields, but that this amount of population has been attributed to areas coded as pasture' by CLC. This may correspond to a large number of small urban nuclei inside the CLC pasture' class.

6.5.1. Further coefficient tuning.
These coefficients can be seen as a starting point for a manual tuning procedure, that allows taking into account additional knowledge on the population settlement in specific areas. An Arc-view tool has been built with the following functionalities:

- Viewing the attributed population density for a subset of land cover classes;
- Interactive selection of a subset of communes based on a set of criteria that

includes a geographic window, average density or % of urban area;
- Attributing new coefficients to the selected set of communes.

6.6. Quality assessment of the disaggregation in Arezzo (Italy)

A first assessment has been made on the basis of the spatial behaviour of the deviation between the actual commune population and the population attributed by disaggregation of NUTS2 data. A link is observed with urban agglomerations: the population attributed to cities is generally smaller than their actual population. The opposite happens for rural areas, where the population attributed is generally larger than the actual one. This indicates that the stratification that has been made is not sufficient. The distance to major urban agglomerations seems to be another element to be considered.

A different quality assessment has been made by comparing the results of disaggregating commune data with data at sub-communal level (census sections) for a test site in Arezzo, including 27 communes subdivided into 1 656 census sections (see chapter 8 in this volume for a description of the site). These data are available as a geo-referenced layer in a GIS environment with geographic limits and population. The total population of the test site is 235 630 inhabitants.

6.6.1. Disagreement at ixel level and at census section level

For this exercise census sections and density maps have been rasterised with a resolution of 25 m. The quality of a density map ${}^{k}Y_i$ can be assessed by comparing the disaggregated density Y_{cm} (commune data + CLC) with the density by census section ${}^{1}Y_i = {}^{1}Y_s = X_s / A_s$ (assumed to be the truth), where X_s and A_s are the population and the area of the census section *s* and *i* is any pixel in section *s*. The comparison has been made in two different ways:

- per pixel: $\delta_{1k}^{pix} = \sum_i \left| {}^{1}Y_i - {}^{k}Y_i \right|$
- per census section: $\delta_{1k}^{\sec t} = \sum_s \left| X_s - {}^{k}X_s \right|$

  where ${}^{k}X_s$ is the population attributed to section *s* by map ${}^{k}Y_i$

The following density maps have been compared with the truth' ${}^{1}Y$

k=2: Density map closest to ${}^{1}Y$ fulfilling the condition of having the same value for pixels in the same commune and the same CORINE land cover class with the classification into 17 classes (*k=3* for CLC in 7 classes). These maps cannot be computed using only commune data and CLC. They give an idea of the best possible result that can be achieved by maps with this condition (scale effect).

$$ {}^{2}Y_{cm} = \frac{{}^{2}X_{cm}}{S_{cm}} = \frac{\sum_j {}^{1}Y_j \, S_{cj}}{\sum_j S_{cj}} $$

k=4 disaggregation with the suggested coefficients (Table 11, Figure 4 c).
k=5: disaggregation with the initial coefficients (Table 7);
k=6: no disaggregation, i.e. attributing uniform density in each commune (Figure 4 a);

Disagreement between the population density map per census section and different maps per CLC class x commune Table 13.

	$\delta_{1k}^{\sec t}$	δ_{1k}^{pix}
k=2	205 766	235 042
k=3	206 737	237 661
k=4	201 946	240 380
k=5	298 676	307 115
k=6	383 998	425 274

Table 13 indicates that, **for this articular site**:

- There is a major disagreement between the density maps per commune and per census section.
- The disagreement is still strong with the disaggregation obtained with the coefficients of Table 13.
- Most of this disagreement is due to the different scales of CORINE land cover and the census sections:
  - Many census sections correspond to a small built area and are not represented in CLC
  - Different census sections in the same commune and same CLC class (specially urban) have very different population densities.
- The map we have obtained at EU-13 level (k=4) is not too far from the optimum that can be got combining commune data and CLC.

The population attributed to the section can be compared with the actual population: the ratio $\left(X_s / {}^4X_s \right)$ represented in Figure 5, seems to indicate that population in urban classes is underestimated.

Systematic over/underestimation for CLC classes can be also assessed by comparing:

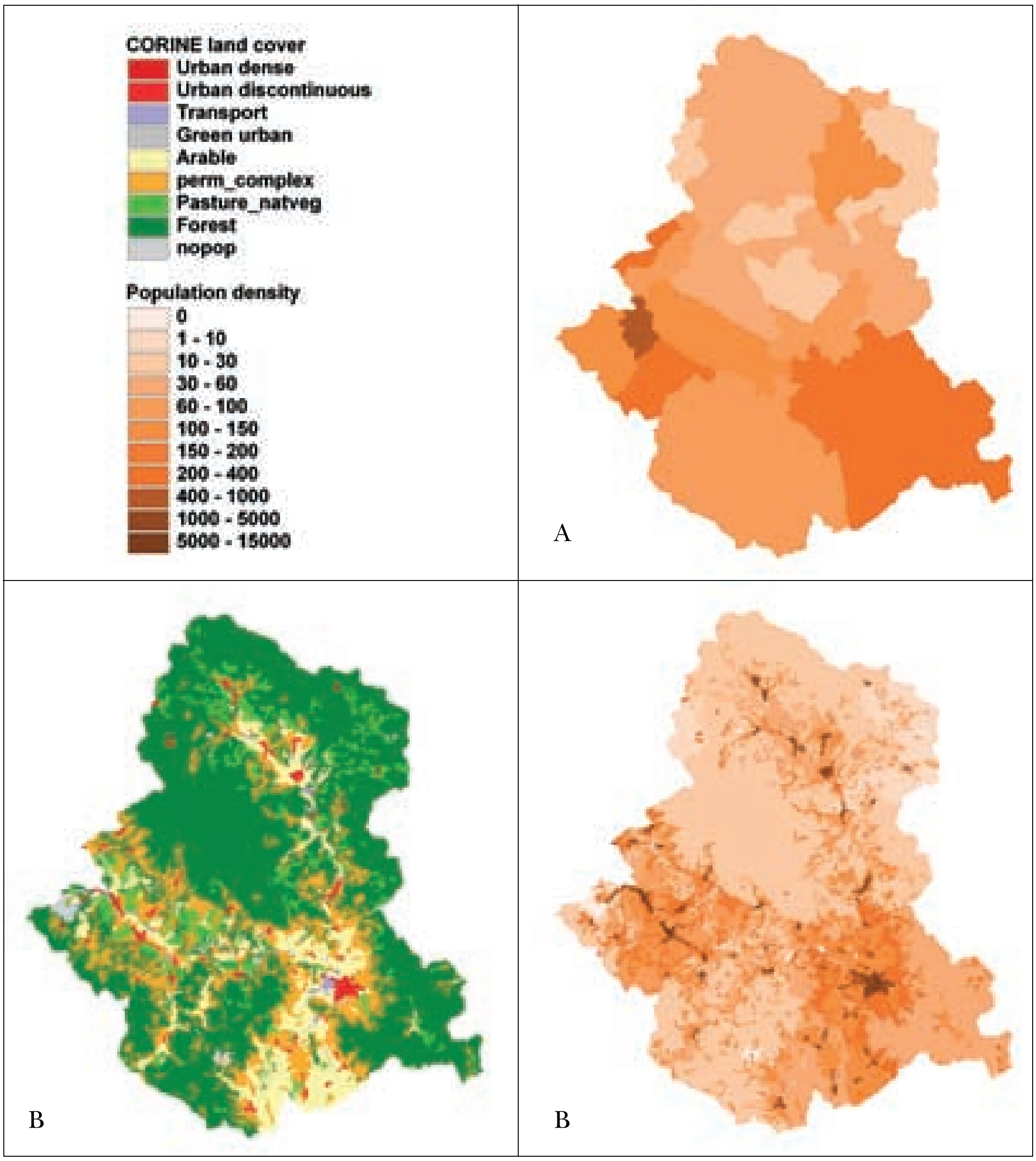
- the total population that has been attributed to pixels in each CLC class, and
- the population appearing in the same pixels when population density is mapped by census sections.

The result of this comparison (see Table 14) indicates that the amount of population attributed to CLC urban classes is approximately in agreement with the data by census sections. The coefficients seem to be too high for pasture, forest and natural vegetation, and slightly underestimated for arable land, permanent crops and heterogeneous agriculture.

Table 14. **% of population appearing in each CLC class in two density maps**

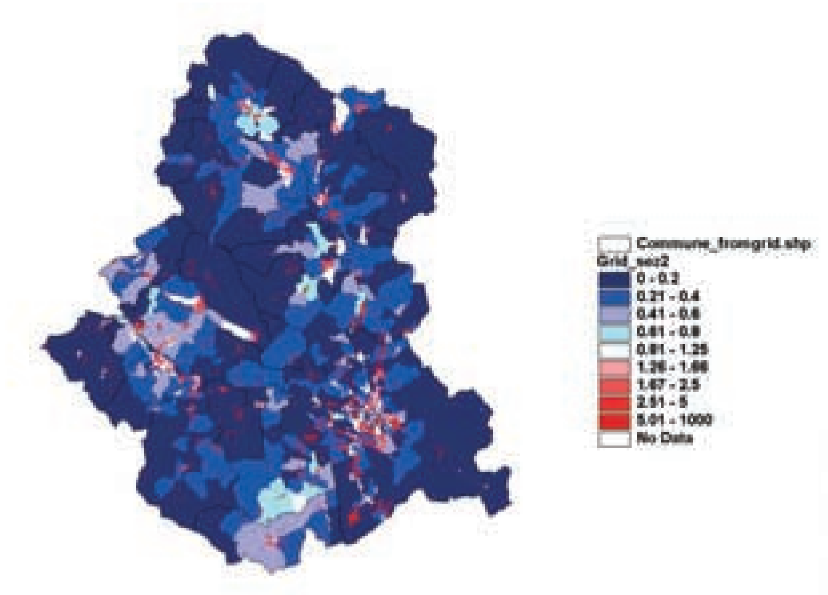
CORINE	Census sections	Attributed (communes +CLC)
Urban dense	9.2	8.7
Urban discontinuous	43.3	45.0
Transport infrastructure.	3.2	2.2
Gree urban, leisure	0.2	0.2
Arable land	10.1	7.5
Permanent crops and heterogeneous	24.1	22.3
Pastures	3.5	4.5
Forest and natural vegetation	6.4	9.6

Arezzo site | Figure 4.



Arezzo site: a) Population density per commune, b) CORINE land cover with simplified legend, c) Population density after disaggregation.

Figure 5. **Arezzo site. Ratio for each census section between actual population and attributed population**



6.7. Conclusions

CORINE land cover can be used to improve the mapping of population density available to the European Commission at commune level. An algorithm has been developed that combines two levels of administrative units (NUTS2 and communes in this paper) to estimate reasonable weighting coefficients for each land cover class. Results can be improved if each region is stratified grouping communes with similar characteristics. The estimation of coefficients becomes more robust if the nomenclature is simplified.

A first assessment, comparing with more detailed data for a test site in Arezzo (Italy), suggests that the estimated coefficients are approximately correct for this site, but additional checks are necessary for sites with different styles of population settlement.

References

Ambroise C., Govaert G., 1998, Convergence of an EM-type algorithm for spatial clustering, *Pattern Recognition Letters* v 19 n 10. p 919-927.

Anderson N H, Titterington D M,1997, Some methods for investigating spatial clustering, with epidemiological applications. *Journal of the Royal Statistical Society Series A, Statistics in Society* Vol. 160; No. 1; pp. 87-105.

Cerioli A, Riani M., 1999, The ordering of spatial data and the detection of multiple outliers, *Journal of Computational and Graphical Statistics*, Vol 8, No 2, pp. 239-258.

Coleman D., Dong X., Hardin J., Rocke D.M., Woodruff D.L., 1999, Some computational issues in cluster analysis with no a priori metric, *Computational Statistics and Data Analysis* Vol. 31 No 1. pp. 1-11.

Diggle P J; Chetwynd A G, 1991 Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics; Vol. 47; No. 3; pp. 1155-1163.

Everitt, B.S., 1980, Cluster Analysis, London: Heineman.

Fisher P.F., Langford M., 1995, Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment & Planning A* Vol. 27 page 211-224.

Flowerdew R., Green M., Kehris E., 1991, Using areal interpolation methods in GIS; *Papers in regional science*, Vol 70 No 3 pp.303-315.

Gale N., Halperin W. C., 1982, A case for better graphics the unclassed choropleth map. *The American statistician*, Vol 36, No 4, pp. 330-336.

Gebhardt F., 1999, Cluster tests for geographical areas with binary data, *Computational Statistics and Data Analysis* Vol. 31 No 1 . pp. 39-58

Hunting Technical Services, 1999, *GIS application development*. Report to Eurostat May 1999.

Le Jeannic Th., 1996, Une nouvelle approche territoriale de la ville. *Economie et Statistique* 294-295 pp. 25-45.

Macchiato-M-F; LA-Rotonda-L; Lapenna-V; Ragosta-M., Time modelling and spatial clustering of daily ambient temperature: an application in southern Italy. *EnvironMetrics*; Vol. 6; No. 1; pp. 31-53; 1995.

Tobler, W.R., 1979, Smooth pycnophylatic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, pp. 519-530.

7. Agricultural statistics spatialisation by means of CORINE land cover to model nutrient surpluses

Philippe Crouzet*, Chris Steenmans** (*IFEN, **EEA)

7.1. Introduction

The nitrogen and phosphorus surpluses resulting from agricultural activities constitute one of the links of the chain of enrichment of inland waters in nutrients. Surpluses also constitute a crucial stage in the DPSIR ([38]) model (2000) of environmental assessment. The most accurate estimate at the scale of whole countries and reported by drainage basin, both of the tonnages and of the local surplus densities, is therefore needed for the environmental assessment to support public policies.

The calculation of the surpluses was developed from proven methodologies, mixing modelling and statistical approaches that require a minimum dataset as a response to an aim of being able to apply the method to the totality of the land areas concerned in the long-term.

The major disadvantage of the statistical sources available is their weak degree of geographical resolution. The use of CORINE *land cover* to spatialise statistical data made it possible to calculate the nitrogen surpluses from agricultural sources and their breakdown by drainage basin on two large areas: the Loire-Bretagne Water Agency district (France) and the basin of Elbe (Germany and Czech republic), covering together more than 300 000 km^2.

The idea of using CORINE *land cover* in order to better assess diffuse sources is not completely new. A systematic correlation test between the types of land cover and the nitrate contents of the watercourses was tried at the European scale (NERI, 1998). Results showed that this relation could not be established because causality between the losses of nitrate and the type of occupation of the land is not direct. On the other hand, this test showed that CORINE *land cover* constitutes a powerful tool to spatialise statistics, which appeared extremely promising.

In another section of this publication, CORINE *land cover* is used to help reassignment of the structure surveys carried out in EU countries (Kayadjanian and Vidal, cf. infra).

7.2. Methodology

7.2.1. Definition of nutrient sur lus

The nutrient surplus' is the difference between the total quantity of nutrients that enter the process of agricultural production and the outgoing quantity that results from production. It is therefore above all an accounting and statistical concept. In the DPSIR context, the agricultural surpluses constitute a key stage, analogue of raw pollution, dependent on the yield of the activity. It therefore constitutes a pressure. The resulting state, measurable or calculable is the quantity of this raw pollution that takes part in the variation of the nutrient stock of the soils of arable land. Lastly, the impact is, for example, the concentration found in the receiving aquatic media.

The term of surpluses is used instead of the term of remainder' used in agronomy to stress an accounting approach, applicable on a large scale and not claiming to agronomic representativity, even if its objective is to be an operational estimator of it.

7.2.2. The sur lus model used

The surplus model is derived from the model of the SCEES (SCEES, 1995) that was developed by the French Ministry of Agriculture. This model was intended

(38) DPSIR is acronym for Driving forces, pressures, State, Impact, Response, which constitutes the assessment framework to assess the environmental issues defined by the European Environment Agency (EEA). This framework is in particular used in all the evaluation reports submitted by the EEA and quoted in this publication. Details are available on the Internet site of the EEA (http://www.eea.eu.int), click 'SEARCH', and type DPSIR.

initially to help the drawing-up of the balance sheet of fertilizers at the farm level based on the methodology by the CORPEN ([39]) (Ministère de l'Agriculture, Ministère de l'Environnement *et al.*, 1998). Initially developed to be calculable by means of a spreadsheet (SCEES, 1997). Ifen and BETURE-CEREC modified and integrated it in the NOPOLU *Système 2* package used by Ifen to build the database of inland waters information and modelling.

Due to its origin, the initial model was completely dependent on the classification used for the French agricultural census. The first adaptation therefore consisted in replacing this classification by a European provisional classification (Joint Eurostat/ EFTA Group, 1997), to be able to deal with the data from any country, despite a certain loss of accuracy due to a smaller number of variables taken into account. This classification has however had to be re-examined because crops with similar economic functions behave very differently with respect to fertilisation. It is in particular the case of the fodder crops, for which was maize forages (fertilised) was separated from lucern and other leguminous crops (not or little fertilised).

All details regarding the equations used are in the source publication of this chapter (Crouzet, 2000).

7.2.3. The core role of CORINE land cover

The modelling of data on a large territory is necessarily the result of a compromise between the degree of detail of the calculation equations and the existence of data to be calculated. In the case of surplus modelling, the specific aim of reporting the results by drainage basin had in addition to be achieved, including the assessment of the diversity of the surpluses in each surface unit of reporting. The diversity of surplus is a key issue for further non-point source assessment, since the transfer factor between surplus and emission is not linear and is soil-dependent (Behrendt, Bach *et al.*, 1999, page 71 and following).

In practice, the best homogeneous and available data over a large territory are those resulting from agricultural censuses. This data is by definition, aggregated by statistical unit of survey (in France, by commune, NUTS5 for decennial census, NUTS3 for the annual Farm structure surveys). This unit is seldom related to catchment boundaries. In complex agricultural areas, land use is very uneven, and direct apportionment would produce erratic results.

The principle of disaggregating and re-aggregating using the land cover data is simple. Agricultural activity greatly defines the land cover; conversely, it is possible to apportion, in relevant regions, agricultural activities to certain types of land cover. Consequently, the cross-assignment between the activities (for example, wheat cultivation) and the types of land cover (for example, arable land) makes it possible to calculate surpluses on a reasonably fine scale, the total and distribution of which are allotted to each element of drainage basin. Obviously, this breakdown method applies only to field balances', and not to farm level' modelling techniques.

The choice of the ad hoc administrative level is dictated by legal and technical constraints. The application of the statistical secrecy rule results in the masking of the elements of aggregates coming from three individuals or less. For example, if only three farms of a commune breed pigs, then the total number of pigs of this commune will not appear. In addition, the geographical attachment of information is the administrative residence of the holding. In other words, if the address of the holding of a farmer is in commune A while his fields are in commune B, the crops land areas will be counted in commune A.

Consequently, and in a rather paradoxical way, the best calculation accuracy on a drainage basin will not necessarily be obtained from communal data if these are protected, but for a slightly larger aggregation unit. In practice, the breakdown from CORINE *land cover* areas was applied to cantonal aggregates ([40]). This process is in conformity with the recent scientific developments relating to the aggregation of data on various geographical scales (Launay, 1997).

---

(39) CORPEN is acronym of *Comité d'orientation pour la réduction de la pollution des eaux par les nitrates, les phosphates et les produits phytosanitaires provenant des activités agricoles* (Steering Committee for the reduction of water pollution by Nitrates, Phosphates and Pesticides from Agricultural source, French Ministry of the Environment).

(40) In fact, the pseudo-canton is used, since it is a strict aggregate of communes. The true 'canton' is a pooling entity, specific to France, clustering small agricultural communes and a fraction of urban communes, not representing an actual area.

Ifen carried out a systematic intersection of the three geographical layers referred to above. This crossing produced a quantified database, comprising 46 fields: code of the commune (NUTS5), codes of the hydrographical area, land areas of the 43 CORINE terrestrial *land cover* land classes in the intersection, and sum as a control. This base is named HYDROSOL, and the intersection methodology was used for France as well as for the Elbe catchment.

7.2.4. Data requirements

The data necessary for calculation is limited to five sets:

1. the CORINE *land cover* layer of land cover;
2. the administrative and hydrographical layers, according to projection compatible with CORINE *land cover*;
3. information, of census type of agriculture, according to the smallest possible administrative entity, in view of the remarks made in the next section regarding the accuracy of the data and the absence of bias resulting from survey techniques and of the application of the statistical confidentiality rules;
4. agronomic information on actual fertilisations and on the yields of the crops. It would be desirable to have this detailed information as time series and for small geographical areas. It is in fact available only in an overall way for large territories. Developing scenarios that allow corrections on any geographical scale mitigated this information deficit;
5. technical coefficient sets relating to crops and livestock, unit quantity of dejections per capita of cattle, content in nutrient of the crop products, etc.

7.2.5. Modelling scenarios

The correctness of the value of the calculated surplus depends mainly on the correctness of the three fundamental parameters of the assessment equations, namely the quantities of fertilisation supplied as fertilizers and as animal manure, the yield of the crops and finally the contents of the harvests in nutrients. This data is never available with a sufficient level of detail; they are even approximate at the level of a whole country.

Consequently, the only manner of correctly approaching the balances is to calculate scenarios taking account of agronomic knowledge, the local practices, etc. The values taken into account in the scenarios can always be improved by surveys to the various agricultural advisers and by exploiting literature. Within the very limited temporal and financial framework of the pilot project, only the values compiled by specialised organisations were retained to calculate two basic scenarios in addition to that of software validation (scenario 1'). Only two calculation scenarios therefore exist. The detailed values of the technical coefficients are reported in the basic publication (Crouzet, 2000).

7.2.5.1. Scenario 2: Calculation of nitrogen without symbiotic fixing

This scenario comprises a fertilisation value adjusted for each type of cultivation, according to the agronomic recommendations compiled by the CORPEN (Ministère de l'Agriculture, Ministère de l'Environnement *et al.*, 1992; Ministère de l'Agriculture, Ministère de l'Environnement *et al.*, 1998). The symbiotic fixing plant crops are not fertilised.

7.2.5.2. Scenario 3: taking into account the symbiotic fixing of nitrogen

For the crops plants capable of symbiotic fixing of atmospheric nitrogen, a value of 90 kg N ha^{-1} $year^{-1}$ was retained. It was also considered that these crops were not fertilised, contrary to a practice that develops, but which seems to have been anecdotic for the 1988 census year.

7.2.6. S atial levels of calculation and re orting

The principle of result apportionment is very simple, but its implementation becomes quickly complex in details, because of the very large number of overlaps between the administrative units, hydrological surfaces and land cover.

All surplus calculations are first made at the level of the smallest entity of collection of the census information. In the case of the pilot study, it is for example the pseudo-canton in France. These calculations are obviously carried out by reallocating all the figures by relevant CORINE land cover entity.

From this basis for calculation, all the required reassignments are carried out. That can lead in intermediate stage to calculating administrative entities smaller than the collection entity, this method being adopted to make the calculation procedure independent of the source data. Only the administrator of the application can accede to this level of calculation. It is possible to carry out aggregations of results for non-relevant levels with respect to environmental pressure, but for which comparison data exists.

The totality of the calculation system was established as a component of Ifen's working database for inland waters, developed within the framework of the NOPOLU Système 2 software. This working database makes it possible to carry out the necessary connections with the other elements of use or comparison with the results of the surpluses calculation model: integrated assessment of the emissions discharged to water, calculation of nutrient fluxes conveyed by watercourses, to quote only the modules having direct relationships with the surplus calculation.

7.3. Results in terms of nitrogen surplus

7.3.1. Main characteristics of source data
The Loire-Bretagne Water Agency district covers 156 217 km^2 and concerns three major hydrographical entities, the drainage basin of the Loire and the tributaries of its estuary (118 054 km^2), the Brittany basins (29 533 km^2) and the coastal basins of the south of the Loire (8 630 km^2). All these basins are located in France.

The smallest units of drainage basins of the data base CARTHAGE (RNDE, 1997) are sets of coding entities of elementary drainage basins, named hydrographical zones'. They are 6315 for the whole of France, at the time of update of the layer used. The study territory comprises 1402 zones. In a later stage, they were aggregated in 299 *sub-sectors*, according to the CARTHAGE terminology.

The data of agricultural use comes from the latest agricultural census (RGA), which unfortunately dates back to 1988. The next is scheduled for autumn 2000. Therefore more recent data is not available for a large area. The data used is the pseudo-cantonal data, already mentioned in the section dealing with methodology. The total values of the agricultural activity were recomputed according to the European classification used for the 1 379 pseudo-cantons of the departments of the studied area. In a second stage, a clustering in 103 entities was done, for the purposes of comparison with the calculations carried out on the Elbe basin.

The Elbe basin drains a total land area of 147 635 km^2. The largest part, downstream, (65.4 %) is in Germany. The upstream part is very unequally distributed between the Czech Republic, 34.2 %, and Austria that owns only 0.4 % of the basin. The latter part, which moreover is covered mainly by forests, was neglected in the study.

The structure of the statistical data provided for the German part is very different from those of the French RGA. They comprise 37 variables, of which 14 do not concern agricultural activities. It was necessary to add 5 variables, calculated from the 23 above, after validation by the data supplier. The technical coefficients published by Eurostat for Germany were used initially (Joint Eurostat/EFTA Group, 1997).

Except for the administrative layer, bought from the MEGRIN organisation, and comprising 56 units, all the Czech data was gathered and prepared by the CORINE *land cover* PHARE Topic Link, in particular the 104 drainage basins. The use of CORINE layer *land cover* posed individual problems, due to the absence of polygons for the type 211 (non-irrigated arable land). Consequently, the later HYDROSOL' type crossings (c.f., page 3) have had to be partly carried out manually.

The agricultural census data comprising 24 variables relating to the crops and 4 variables for livestock, was therefore already strongly aggregated. It was corrected and transformed according to the European classification, whenever possible.

The distribution of the field crops shows a preponderance of cereals in the Elbe basin. On the French basins, meadows are in a majority on the basin of the Loire, but the share of the fodder crops (including temporary meadows) is much more marked than on the basin of Elbe. A regular distribution of moderate nutrient surpluses on the basin of Elbe, and small islands of strong surpluses, as well as sectors in apparent deficit on the Loire-Bretagne Water Agency district can reasonably be expected, as consequence of the presence of a large number of animals and broad areas of permanent meadowland as well.

7.3.2. Results on the Loire-Bretagne Water Agency district

7.3.2.1. Surplus by catchment

The model was checked against results previously presented for Brittany (ENSAR, 1995). Despite several differences between this approach and the pilot study, it was possible to carry out comparisons and to validate the software used for modelling.

Table 1. **Aggregated results relating to scenario 3 (with symbiotic fixing). Results for an average year with practices of 1990 approximately**

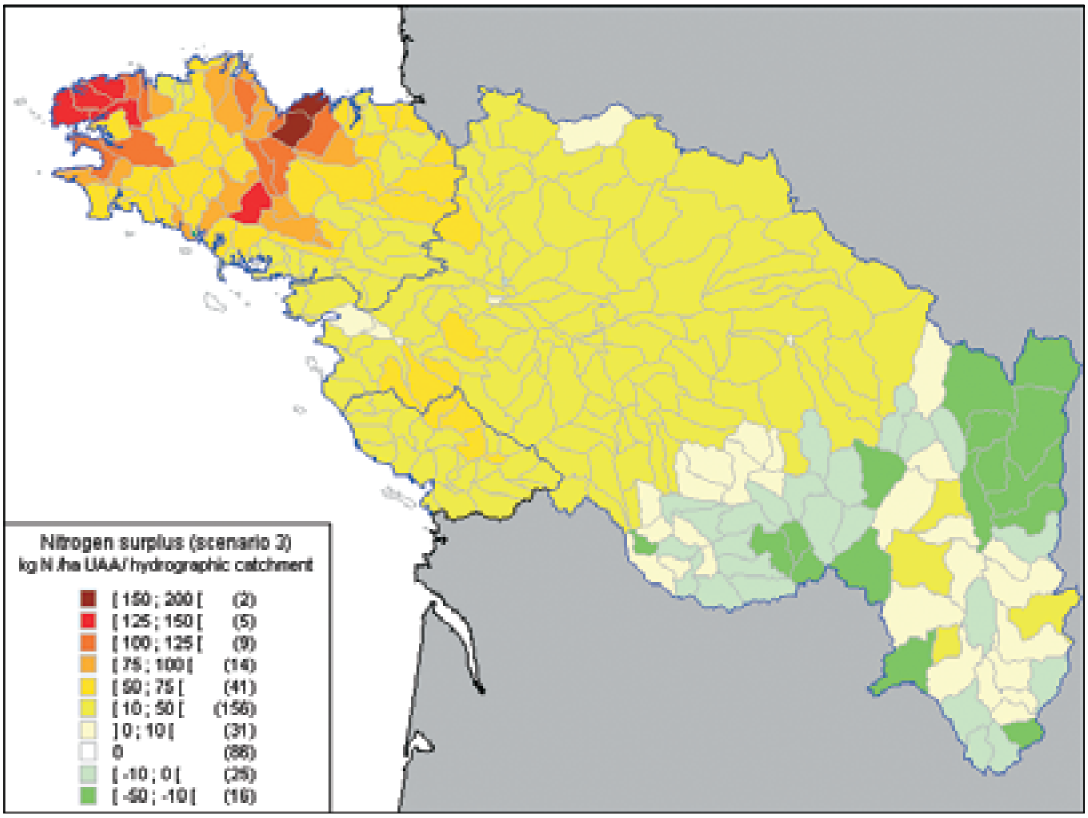
Source: Ifen and BETURE-CEREC, 1998.

	Total surplus Tons N	Area with surplus ha	Average surplus kg/ha/year	Total of deficiencies Tons N	Areas with deficiencies ha
Loire river basin	144 675	5 915 275	24	17 952	1 797 870
Brittany	157 299	2 077 393	76	85	857
South-Loire	23 544	831 090	35	0	0
TOTAL	**325 518**	**8 823 758**	**37**	**18 037**	**1 798 727**

The taking into consideration of the large areas of leguminous crops gives results that compare well with other sources of information. Hence, only results computed with scenario 3 hypothesis are discussed in this paper and are reported in the table above.

Figure 1. **Geographical distribution of the average nitrogenous surpluses per hectare, according to the scenario 3**

Source: Redrawn from (Ifen and BETURE-CEREC, 1998).



One points out that Brittany, numerous areas of which are considered in structural surpluses (as defined by regulation), generates a surplus virtually equal to that of the basin of the Loire, about 150 000 tons of nitrogen a year each one. The total for the Loire-Bretagne Water Agency district is therefore approximately 320 000 tonnes of nitrogen a year, for an average year centred over 1990.

The surplus is simply the difference between the input and output quantities of the farming system. Summing up the various items makes it possible to quantify the percentage of surpluses, as calculated by scenario 3, with those items.

Assessment of the input-outputs for the Loire-Bretagne Water Agency district (scenario3) — Table 2.

All figures in thousand T N/year	**Loire basin**	**Brittany**	**South-Loire**	**Total**
Total chemical fertilisers	565	157	54	776
Total inputs from symbiotic fixing	94	58	15	166
Total animal waste	279	176	27	481
Sub total	**938**	**391**	**95**	**1 423**
Total outputs	811	233	71	1 115
Raw balance (algebraeic input-output)	127	157	24	308
Net balance (sum of positive surplus)	154	167	24	345
in % of input	**16 %**	**43 %**	**25 %**	**24 %**
in % of outputs	**19 %**	**72 %**	**34 %**	**31 %**

Source: Ifen and BETURE-CEREC, 1998.

The figures of this table show that where the crops are the majority, the surplus rate (here 16 %) is very close to the high value of the average considered on a pan-European scale (Crouzet, 2000, page 9). By contrast, in the areas receiving a great deal of nutrient inputs from livestock sources, the surplus seems more marked, reaching 43 % of total nutrient input in Brittany.

7.3.2.2 Comparison with nitrate fertilizer sales

The sales of artificial fertilizer constitute the only source of independent information that can be usefully compared with the estimates of the model. These fertilizers constitute in addition an important source of fertilizers, or even the majority contribution where livestock is not very numerous. Fertilizer sales being published in the form of NUTS3 statistics, aggregation at this level was calculated for all the NUTS3 units intersected by the Loire-Bretagne Water agency district limits.

The comparison of results demonstrated that modelling based on agronomic constants gives much better results than the use of only sales statistics. Sales data should however be used to frame overall values calculated from agronomic constants. Hence, it is very important that the aggregation method permits the production of *ad hoc* aggregates; this is particularly the case when using CORINE *land cover* to carry out these operations.

7.3.2.3. Contribution of the surpluses to the fluxes monitored in the downstream part of the water-courses

The nutrients transported by the watercourses result from direct inputs from the cities and industries of non-registered inputs, of which diffuse inputs from agricultural sources account for the main share. Riverine fluxes constitute the only ultimate reference to emission assessment, provided correct accounting of retention is made (Behrendt, Bach *et al.*, 1999; Kirk McKlure Morton, 1999).

Two methodological developments were tested on a pilot basis on the Loire-Bretagne Water Agency district, covering respectively the calculation of nutrient fluxes at the mouth of the rivers (BETURE-CEREC, 2000a) and the assessment of nutrient emissions, all sources taken together (BETURE-CEREC, 2000b). Results for the Loire and the Vilaine are presented in the following table.

Table 3. **Nutrients inputs and fluxes measured on the basins of the Loire and the Vilaine rivers**

Source: see text above. Studies carried out on behalf of the Loire-Bretagne Water Agency and Ifen.

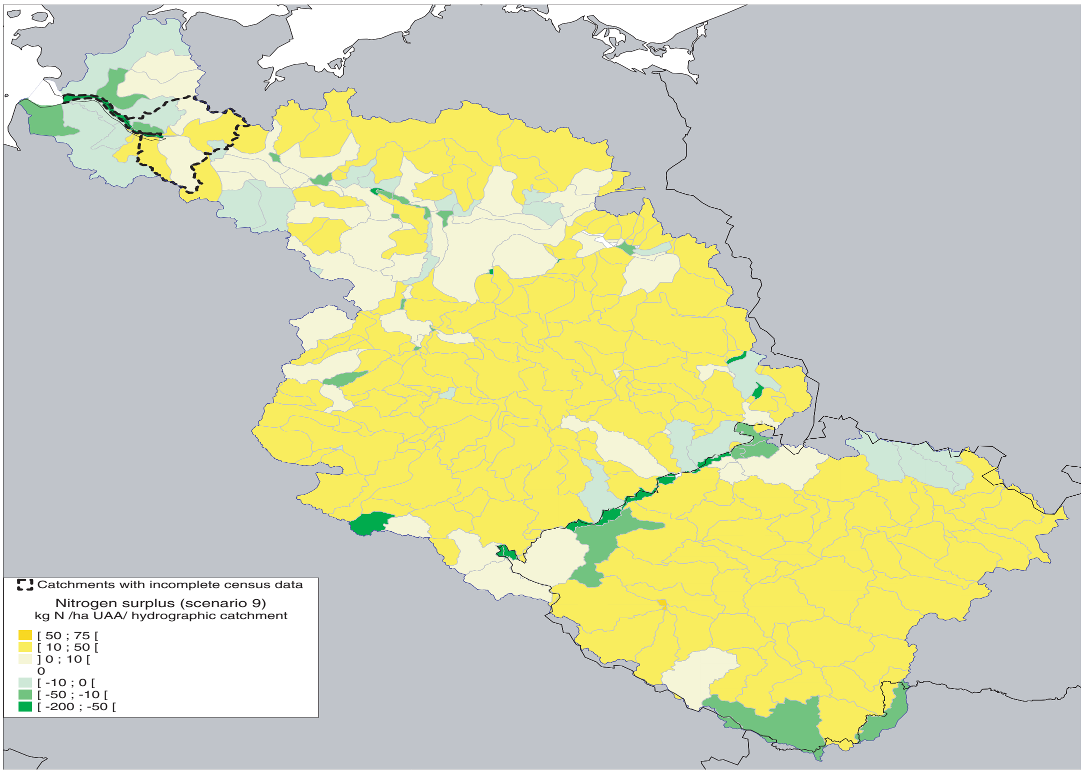
All figures are in thousand Tons N	**Vilaine**	**Loire**
Net surplus	43.3	130.0
Retention hypothesis	70 %	30 %
Non-point agricultural inputs (net surplus *(1-retention))	13.9	91.0
Other non-point sources	1.0	9.2
Direct emissions (urban and industrial)	1.8	26.0
Total of emissions to rivers	**16.7**	**126.2**
Calculated flux (interannual average)		
Inorganic nitrogen	14.9	87.0
Organic nitrogen	1.9	44.2
Total flux	**16.8**	**131.2**
	Hypothesis of strong retention, according to local surveys	*Hypothesis of weak retention, according to local surveys and soil types*

7.3.3. Results on the Elbe basin

Adjusted scenarios were used, after result assessment. The final scenario (derived from scenario3) after adjustment of the technical coefficients proposed for Germany according to Eurostat's compilation (Joint Eurostat/ EFTA Group, 1997) and technical coefficients relating to fodder and to animal feed is called scenario 9.

Figure 2. **Results on the German and Czech parts of the Elbe basin, by catchment 'scenario 9'**

Source: Redrawn from (Ifen and BETURE-CEREC, 1998).



In the case of the Czech part, result proofing by basin and by administrative entity reveals distribution differences particularly marked next to the German border which meaning is discussed in the source report.

Summary of the results obtained on the basin Elbe (scenario 9 only). Table 4.

Elbe		Total surplus (1000*T of N)	Average surplus (kg N/ha/ year)	Total deficit (1000*T of N)	Average deficit kg N/ha/year	Deficit area % of Agricultur al area (A.A.)	A.A. (1000*ha)	Total area (1000*ha)
Scenario 9	Czech Republic	42.0	16.9	4.1	19.4	8 %	2 695	5 046
	Germany	93.8	19.5	6.7	8.6	14 %	5 586	9 647
	Total	**135.9**	**18.6**	**10.8**	**10.9**	**12 %**	**8 281**	**14 693**

Summary of the results obtained on the basin of the Loire Table 5.

Loire	Total surplus (1000*T of N)	Average surplus (kg N/ha/year)	Total deficit (1000*T of N)	Average deficit kg N/ha/year	Deficit area % of A.A.	Agricultural area (A.A.) (1000*ha)	Total area (1000*ha)
Scenario 7	147.9	25.8	25.2	12.3	26.4	7 804	11 770

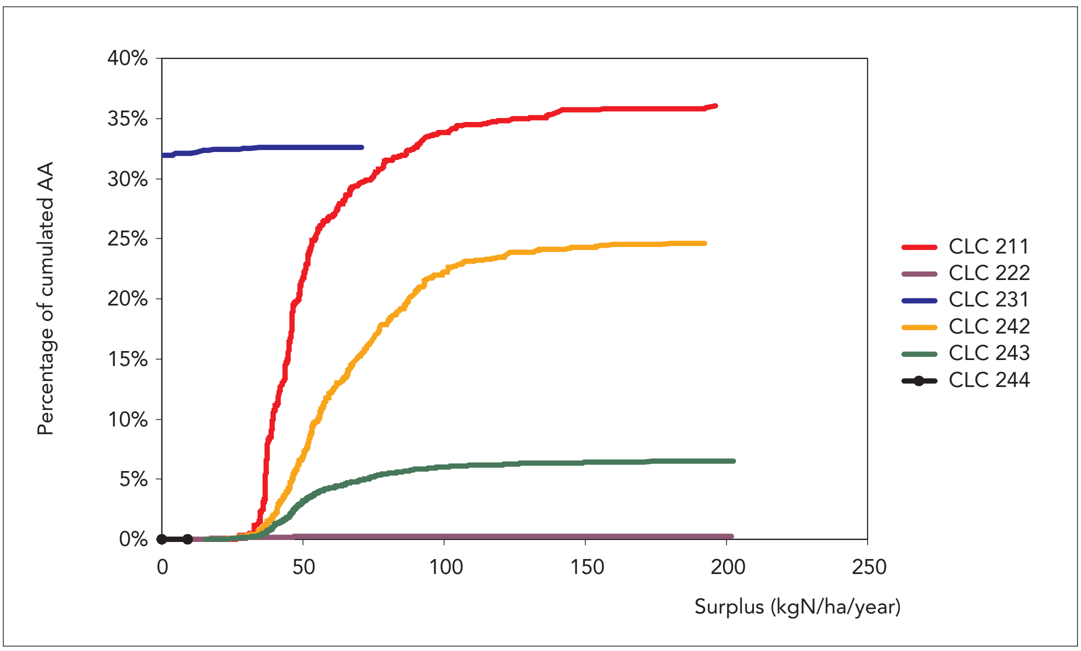
Source: of Table 4 and Table 5: Ifen./ BETURE-CEREC, NB: The reported values are so with all the figures calculated, for the sake of consistency with the final tables and aggregates. This implies no information about the accuracy of the results

7.3.4. Relations between the sur luses and the ty e of occu ation of the land

Only six types of land cover contribute to surplus production. These are classes 211 (*Non-irrigated arable land*), 222 (*Fruit trees and berry plantations*), 231 (*Pastures*), 242 (*Complex cultivation patterns*), 243 (*Land principally occupied by agriculture, with significant areas of natural vegetation*) and in a very marginal way, the occupation of the type 244 (*Agro-forestry areas*).

The figure below presents the cumulated percentages of areas ofagricultural land, according to the value of the surplus, for the entire the Loire-Bretagne Water Agency district. It can be seen that the meadows, which constitute a considerable fraction of the land areas, do not exceed 50kg N/ha surpluses in practice.

Distribution of the cumulated land areas, according to the surplus (the Loire-Bretagne Water Agency district) Figure 3.



Source: EEA and Ifen/ BETURE-CEREC. NB CLC XXX refer to the CORINE land cover types mentioned in the paragraph above.

Only three types contribute significantly to the surplus (211,242 and 243). It is instructive to note that the activities over the classes 211 (*Non-irrigated arable land*), and 242 (*Complex cultivation patterns*) have a notably different impact. Whereas the class 211 extends over a larger proportion of the total area, the high values of the surpluses are less frequent than it is for class 242Tjhis second class is finally at the source of a double load. On this type, as for the type 243 (Land principally occupied by agriculture, with significant areas of natural vegetation) sectors with low surplus are rarely found.

7.4. Discussion of main findings

7.4.1. The ositive in ut of CORINE land cover

In the source study, the impact of source data, livestock feeding hypothesis, differences in statistical nomenclatures and degree of spatial aggregation were computed. In this summary report, only the degree of spatial aggregation with respect to CORINE *land cover* is discussed.

The calculation model was developed in order to apply it to very wide areas. Consequently, it is essential to evaluate the possible effect of the heterogeneity of the aggregation unit of source data likely to be mobilised on the results, knowing that it cannot be expected to find homogeneous statistics on a pan-European scale.

In this pilot application, this heterogeneity is clear; it concerns mainly the following groups of information:

1. Basic administrative and hydrographical divisions being used to calculate the CORINE *land cover* areas of the intersections. The assumption, that was not fully confirmed, is made that the CORINE *land cover* layer is homogeneous for all the territories concerned. With only regard to communes and drainage basins of the Loire-Bretagne Water Agency district, there are18 354 intersections with an average land area of 8.5 km^2. In the case of the basin Elbe, the geographical layers available made it possible to calculate only 1233 intersections, of average land area 120 km^2.
2. The size of the minimum territorial units for supplying agricultural statistics. In France, the cantonal RGA is available. It comprises 1 379 units for the departments (NUTS3) of the studied area, the basin itself by including only 996 units. On the other hand, for the basin of Elbe, only 138 statistical units were obtained for the German part and 56 units for the Czech part, i.e. 194, after deduction of a German unit (obviously urban) lacking in agricultural information.

To evaluate the impact of these differences on the results, the only standard available was the Loire-Bretagne Water Agency district, by applying various aggregation rules to the different available datasets.

The most spectacular check consisted in aggregating the combinations involving CORINE *land cover* in larger units, the 30 402 initial intersections being merged into 1 133 intersections, of average land area, close in their size to the intersections available on the basin of Elbe. The results obtained were tested in parallel, with and without use of CORINE *land cover*.

These aggregations and calculations were carried out for the canton, the arrondissement' (the 103 units referred to above) and the department (36 units of NUTS3 level).

The maps in the Figure 4 below distinctly show a gradient of evenness of the results from cantonal calculation spatialised with CORINE *land cover* until departmental calculation without spatialisation.

The analysis of the results was carried out by correlating the results of surpluses by sub-sector, by density and by absolute value, while taking as reference the calculation that were obtained by canton with the use of CORINE *land cover*.

Obviously, the difference between the results spatialised or not is expected to be all the larger as basic statistics are aggregated over a greater land area. If statistics were available on equal or smaller units than units CORINE *land cover,* then its use would obviously be unjustified.

Since the aim is to check the equality of the results, the correlation and the regression are calculated with a linear model, the constant of which is forced to zero. In this case, the value of the correlation coefficient indicates the consistency of the series, while the slope indicates the difference between the calculated values, if it deviates from the unity.

Figure 4. **Values of the surpluses by sub-sector, according to the level of aggregation of the statistical data and of the taking into account or not of CORINE land cover**

Source: Ifen/BETURE-CEREC

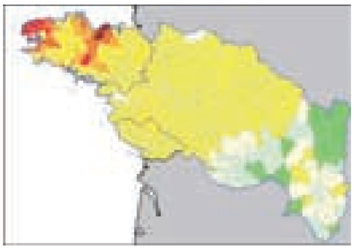
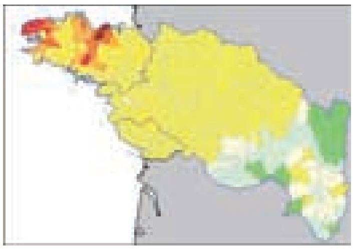
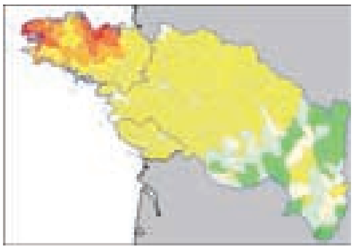
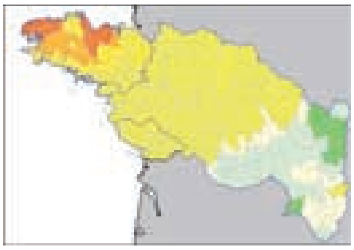
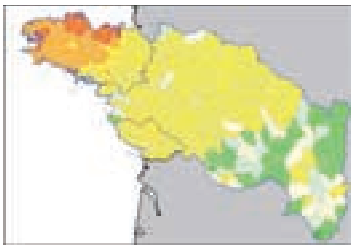
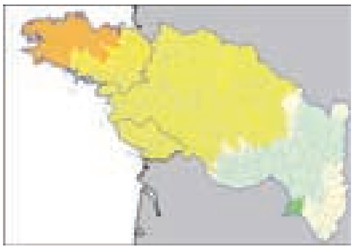
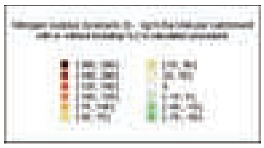
	With spatialisation CORINE *land cover*	Without use of CORINE *land cover*
Canton		
Arrondissement		
Department		
		

Figure 4 clearly show the superiority of modelling based on spatialisation. Although the correlation coefficients relating to the departments calculated with CORINE *land cover* are not as good as those of the arrondissements' calculated without it, the slope is closer to 1, which suggests a better estimate of the total value, despite a more substantial dispersal. In the case of the departments, the gain given by CORINE *land cover* is clear in the case of the calculation of the total surplus. It appears less obvious in the case of the surplus densities. In fact this comes from a de facto standardisation of the surpluses to a value of ~100 kg N /ha/year, which created a very strong discrepancy between the reference figures and the values tested in this range of values.

Table 6. **Statistics on the validity of the results modelled with and without the aid of CORINE land cover**

Source: EEA on Ifen/ BETURE-CEREC

Results related to:		**Cantons (without CLC)**	**'Arrondisse-ments' (with CLC)**	**'Arrondisse-ments' (without CLC)**	**Departments (with CLC)**	**Departments (without CLC)**
Surplus (1000 T/year)	r2 coefficient	0.99	0.93	0.91	0.86	0.82
	Slope	0.98	0.97	0.91	0.94	0.85
Density (kg/ha/year)	r2 coefficient	0.98	0.86	0.84	0.75	0.71
	Slope	0.98	0.99	0.95	0.99	0.93

Table 7. **Overall results on the Loire-Bretagne Water Agency district (reference scenario 3)**

Source: EEA on Ifen/ BETURE-CEREC

Results related to:	**Cantons (with CLC reference)**	**Cantons (without CLC)**	**'Arrondisse-ments' (with CLC)**	**'Arrondisse-ments' (without CLC)**	**Depart-ments (with CLC)**	**Depart-ments (without CLC)**
Surplus (1000 T/year) (sum of positive values)	326	323	328	319	331	316
Surplus (1000 T/year) (sum of negative values)	18	16	26	13	35	8

7.4.2. Sources of uncertainty related to in ut data and constants

The conclusions related to issues other than CORINE *land cover* were that results obtained by modelling the surpluses with the use of CORINE *land cover* are mainly sensitive, by order of **decreasing im ortance**, firstly to the *technical coefficients and to the values of yield from the crops,* secondly to the *degree of aggregation of basic statistics* and finally, but in a very marginal way, to the more or less major *aggregation* of the collected *statistical variables.*

The impact of livestock feeding sources deserves mentioning. Since, under maximalist hypothesis, based on non-limited supplies of fodder to the livestock, the total surplus calculated on the Loire-Bretagne Water Agency district, these figures rise from 325 thousand tonnes a year to 492 thousand tonnes. The area in surplus passes then to 9.6 million ha (62 % of the total land area) instead of 8.8 million ha.

In the basin of Elbe the maximum scenario relating to the feeding of livestock raises the surplus from 140 thousand tonnes/year to 325 thousand tonnes, distributed among 8.2 million ha (56 % of the total land area) instead of 6.9 million ha.

7.5. Conclusions

The results obtained are very encouraging because they show the solid nature of the approach and its capacity to provide usable results even by using censuses, which results are aggregated over tens of km^2 and which number of variables is limited to about twenty.

Certainly, the results obtained are even more dispersed as the basic variables and the technical coefficients available are lumped. On the other hand, results are only slightly biased, i.e. the total values, in tonnage, as well as the density of the surpluses are preserved. The result is that the use of a geographical layer compensates for to a certain extent the loss of information introduced by the increase in the degree of aggregation of statistics. **In other words, overall distribution is well reserved, as well as the quantity of the sur lus even if the location becomes increasingly vague as the degree of aggregation increases.**

The good sensitivity of the model to the values of the technical coefficients does not present only disadvantages. The structure of the model also allows a fine adjustment of calculations, there only where it is necessary. It is therefore a tool for evaluating the relevance of the coefficients.

This means that the use of CORINE *land cover* makes it possible to lower an uncertainty factor, which is the heterogeneity of the statistics of the agricultural censuses. Consequently, it only now remains to solve the questions related to the technical coefficients and the crop yields. These questions are obviously of importance, but they pertain to the field of the agronomists, of experts and of the analysis of literature. It therefore becomes possible in the near future to produce homogeneous and comparable results with limited efforts, because the questions to be solved come within the domain of engineering and not within the basic statistical data gathering. The method therefore reduces the dependence of those who have to produce the results with respect to the organisations responsible for the production of the basic data, making it more sustainable.

In addition, the reliability brought to the model by the introduction of the CORINE *land cover* layer makes it possible to envisage the production of correct results with acceptable accuracy at the scale of a few hundred to a few thousands km^2. The related calculations would be quite independent of the degree of aggregation of basic agricultural statistics scale, which can go from the cantonal level (units from 100 to 150 km^2) as far as the departmental level (units from 3 000 to 6 000 km^2).

Hence, the use of CORINE *land cover* makes it ossible to com ensate mainly for the absence of availability of s atially homogeneous statistics. Des ite that, it becomes ossible to roduce usable results, because they are sufficiently exact and com arable. Spatialisation by means of CORINE *land cover* makes it possible to evaluate simply and quickly the range of the agricultural surpluses in Europe, even using the heterogeneous data currently available. Similarly, it seems possible to calculate intermediate states between two censuses of agriculture, and therefore to evaluate better the trends.

This will be helped by the improvement of statistical relationships between land use and farming surveys and land cover.

References

Behrendt, H., Bach, M., Huber, P.*, et al*, 1999. *Nährstoffbilanzierung der Flußgebiete Deutschlands.* Umwelbundesamt, Berlin. In: *Nährstoffbilanzierung der Flußgebiete Deutschlands* CD-Rom by Geodaten Integration und Analyse, 1999.

BETURE-CEREC, 2000a. 'Evaluation des flux'. (*report type: final, written by Guillaume Le Gall, ordered by* 'Ifen, Agence de l'eau Loire-Bretagne') Orléans, 72 pages, (5 annexes), availability: total.

BETURE-CEREC, 2000b. *Evaluation intégrée des émisisons. Application pilote au BV de l'Agence de l'eau Loire-Bretagne.* Ifen, Orléans. In: *Evaluation intégrée des émissions polluantes* 2 CD-Rom by BETURE-CEREC, St Quentin en Yvelines, 2000.

Crouzet, P., 2000. *Calcul des surplus de nutriments d'origine agricole. Spatialisation des statistiques grâce à Corine land cover. Etudes et travaux n° 31.* Orléans. Ifen, EEA and JRC. 64 pages.

EEA, European Environment Agency, *'How we reason'*, Copenhagen.
URL: http://org.eea.eu.int/documents/brochure/brochure_reason.shtml, (in 2000).

ENSAR, 1995. 'Les bassins versants de Bretagne et leur charge polluante'. (*report type: final, writen by Pierre Aurousseau, Marie-Chantal Baqué et al., ordered by* 'DRAF Bretagne') Rennes, 30 pages, (23 cartes annexes), availability: limit.

Ifen and BETURE-CEREC, 1998. 'CORINE L.C. as basic layer to non-point source emissions assessment. Agricultural emissions comparison between the Loire (F) and another European basin'. (*report type: final, writen by Philippe Crouzet, Guillaume Le Gall et al., n° 13432-97-11 F1ED ISP F, ordered by* 'EEC Joint Research Centre') Orléans, 27 pages, (7 annexes), availability: total.

Joint Eurostat/EFTA Group, 1997. 'Soil surface nitrogen balances in EU countries'. (*report type: final, writen by Claude Vidal and Maria Pau-Vall, n° Agrienv/97/4, ordered by* 'Eurostat, Statistics on the Environment. Sub-Group on Nitrate balances') Luxembourg, 18 pages, (9 annexes), availability: limit.

Kayadjanian, M. and Vidal, C. 'Reassignment of the Farm Structure Survey's data', cf. *Infra*.

Kirk McKlure Morton, 1999. 'Low level trial — Implementation of Draft OSPAR Guidelines for harmonised Quantification and Reporting procedures for Nutrients. Contribution by Ireland'. (*report type: final, ordered by* 'OSPAR Secretariat. London') Dublin, 83 pages, (2 annexes), availability: limit.

Launay, M., 1997. 'La pollution agricole diffuse par l'azote sur le bassin versant de l'Elorn: diagnostic du risque par agrégation de données à différentes échelles'. (*Thèse de Doctorat en Géographie*, Université de Rennes — Haute Bretagne, Rennes), 325 pages.

Ministère de l'Agriculture, Ministère de l'Environnement and Mission Eau-Nitrates (CORPEN), 1992. 'Recueil des bases de préconisation de la fertilisation azotée des cultures'. (*report type: final, written by Bernard Carlotti, ordered by* 'Ministère de l'Environnement') Paris, 136 pages, availability: total.

Ministère de l'Agriculture, Ministère de l'Environnement and Mission Eau-Nitrates (CORPEN), 1998. 'Bilan de l'Azote à l'exploitation'. (*report type: final, written by Collectif, ordered by* 'CORPEN') Paris, 36 pages, availability: total.

NERI, 1998. ‘Pilot-study on the potential use of CORINE land cover information to interpret mean nitrate concentrations in rivers at the European scale’. (*report type: draft, written by D.I. Müller-Wohlfeil, O Sortkjær et al., ordered by* 'European Topic Centre on land cover') Silkeborg, 21 pages, availability: limit.

RNDE, 1997. ‘Notice de constitution de la BD CARTHAGE. Version 1.1 du 12/05/97’. (*report type: final, written by OIEau, ordered by* 'RNDE') Limoges, 46 pages, availability: limit.

SCEES, 1995. ‘Logiciel de calcul du bilan de l'azote à l'échelle cantonale’. (*report type: final, writen by Claude Vidal, ordered by* 'Ministère de l'agriculture. Service central des enquêtes et études statistiques') Toulouse, 24 pages, availability: limit.

SCEES, 1997. ‘Guide d'utilisation du classeur de calcul BILAN.XLW’. (*report type: final, writen by Michel Poiret, ordered by* 'Ministère de l'agriculture. Service central des enquêtes et études statistiques'), 23 pages, availability: limit.

8. Comparing CORINE land cover with a more detailed database in Arezzo (Italy)

Javier Gallego (JRC)

8.1. Introduction

When CORINE land cover (CLC) is compared with land cover information at a different scale, premature conclusions could be drawn if a straightforward comparison is seen as a quality assessment. If the comparison between two geospatial land cover data layers takes into account their different scales, the conclusions change. We give a procedure to perform such a comparison trying to remove the effects of different scales and collocation inaccuracy.

We also illustrate the fact that CLC should not be directly used for area estimation by polygon area measurement. However CLC, used as covariable, is extremely useful for land cover area estimation (Gallego et al, 1999)

8.1.1. The Arezzo ILC database and CORINE land cover

ISTAT, the National Statistical Institute of Italy, has developed a land use/land cover database on an area of 200,000 hectares in the Arezzo province, in central Italy, with the geometric accuracy of 1:25,000 scale. Below we call this land cover map ILC. The main purpose of this database was testing the possible extension at country level (ISTAT, 1998). The pilot project was partially funded by the European Union (EU).

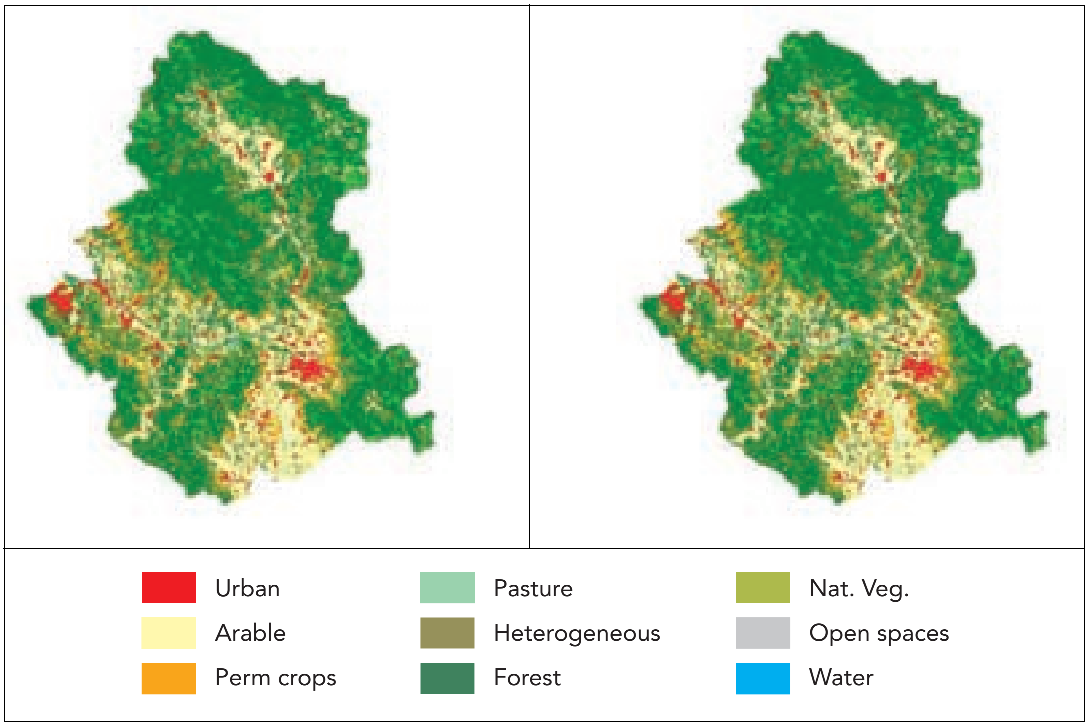
The territory is divided into polygons classified into land use/land cover classes. The nomenclature has five levels; the three first levels coincide with the CORINE land cover nomenclature (EC, 1993); a fourth level is added for urban areas and a forth and a fifth level are added for forests and semi-natural areas.

The project used digital black and white ortho-photos with a resolution of one meter and three Landsat-TM quarter-scenes, taken on the following dates: 10/05/97, 26/10/97, 06/05/98. The photo-interpretation was carried out drawing polygons directly on screen with data at the level of census sections as ancillary information. The minimum mapping unit was 1,56 hectares (1 hectare for urbanised areas).

The polygon layer was integrated with a vector layer composed of linear and point features, referring to railways, highways, major roads and rivers and topographic names. Linear and point features were digitised on screen taking into account the digital ortho-images geometry, after identification on 1:250 000 scale maps. The database has undergone several quality checks, including independent photo-interpretation of samples of arcs, polygons and points (Napolitano et al, 2000).

Figure 7 gives a visual overview of both land cover maps (ILC and CLC) and Table 1.1 shows their main differences. In the pilot area there is a dominant presence of forest with significant agricultural and urban areas in the valleys.

CORINE and ILC land cover maps recoded into 9 classes — Figure 7.



Main differences between CLC and ILC — Table 1.1.

	CORINE land cover	Arezzo ILC
minimum mapping unit	25 ha	1.56 ha (1 ha for urban areas)
location accuracy	100 m	25 m
Area covered	>3 Million km^2	2 000 km^2 . Extension to 300 000 km^2 proposed
Nomenclature	3 levels, 44 classes	same with 2 additional levels

8.2. Comparing total area per class

land cover maps are sometimes used for area estimation of a land cover class by simply adding the area of the polygons labelled as belonging to that class. This approach is rather naïve and can lead to a serious bias if the mapping scale is not detailed enough or the thematic accuracy is not very high (Gallego et al, 1999). We can compare the total area obtained by this method from CORINE land cover and the ILC map for some groups of land cover classes (Table 2.1).

On the other hand **inter reting the strong differences that a ear for some classes as a sign of oor accuracy of CLC would be also naïve**. These differences appear mainly because CLC respects its scale specifications. For example urban areas under 25 ha do not appear in CLC but may appear in ILC and certainly should be counted for statistics. Interpreting the total area of ILC as statistics is also inadequate. Isolated buildings under 1 ha should not appear in ILC and should be counted as artificial for land cover statistics.

Table 2.1. **Total area in km2 per land cover class from CORINE and ILC land cover maps**

	ILC	CORINE
Artificial	84.9	57.1
Arable land	375.9	234.8
Permanent crops	141.0	99.7
Pastures	21.8	66.9
Heterogeneous agriculture	110.8	328.2
Forest	1117.6	1029.5
Natural vegetation	132.4	167.2
Open spaces and wetland	1.7	5.5
Water	5.3	2.3

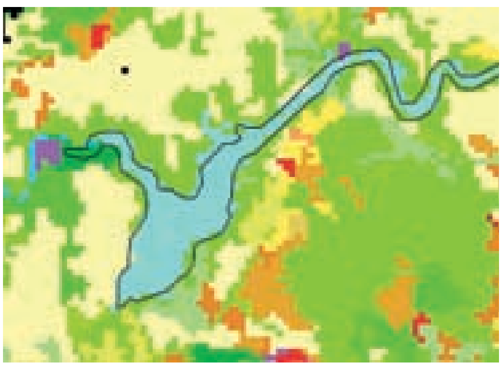
Part of the disagreement can be attributed to the fact that the images were taken on different dates, part can be due to the scale effect and part to different thematic accuracy levels. Compared with CORINE land cover, the area labelled as heterogeneous agriculture' is significantly reduced and attributed to pure' classes. This would explain, at least partly, the increase of arable land, permanent crops and forest. However, if we remove the polygons coded in CORINE land cover as heterogeneous', we still have 26.000 ha classified as arable land by ILC, i.e. 11 % more than CORINE land cover. We shall see below that the thematic disagreement for arable land is very low; therefore a significant part of the difference in polygon area comes from the scale effect.

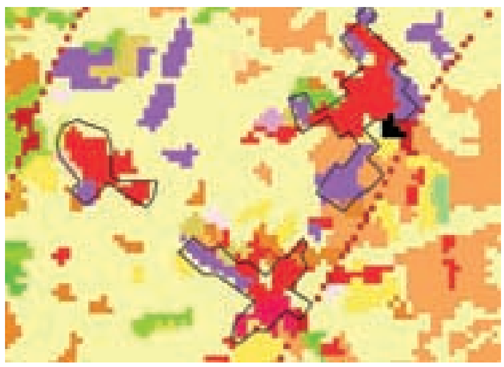
Another part of the disagreement may be due to a different interpretation of the nomenclature or to photo-interpretation errors, but it is difficult to know the impact of each source of disagreement without a suitable ground survey. We try below to resolve the impact of each source of disagreement.

8.3. Statistical analysis of the differences between the two maps

Both land cover maps have been rasterised with a resolution of 50 m for the present analysis. These 50m × 50m cells are called pixels'. A visual inspection of the overlay of both land cover maps in UTM 32 indicates that the co-location accuracy is generally within 50 m (Figure 8), and nearly always within 100 m, i.e. within CORINE land cover specifications.

Figure 8. **CORINE land cover poygons with ILC map rasterised at 50 m.**





If we disregard co-location inaccuracy, we can describe the disagreement through a matrix (often named confusion matrix) with elements $A_{cc'}$ = area of the pixels with code c in CLC and c' in ILC. This confusion matrix would be a naïve description of the disagreement. To eliminate the part of the disagreement due to co-location tolerance (100 m), we eliminate a buffer 2 pixels wide on each side of the CORINE land cover polygon borders. This corresponds to eliminating a 200m wide corridor around the polygon limits (Figure 9).

Buffer of 2 pixels eliminated around CORINE land cover limits with rasterised CORINE land cover Figure 9.

CORINE land cover | ILC

8.3.1. Pixelwise (naïf) commission and omission disagreement

The accuracy of a rasterised land cover map can be assessed through the confusion matrix, that gives the number of pixels $A_{cc'}$ for which the true land cover is c and the land cover given by the map is c'. The omission error for class c refers to the proportion of pixels for which the truth is c and the map gives a different class. The commission error is the proportion of pixels represented as c by the map, but having a different true land cover.

Here the usual terms confusion matrix’ and commission and omission errors’ are substituted by disagreement matrix’ and commission and omission disagreement’ to stress the fact that two different land cover maps are compared: none of the maps can be considered as the truth and part of the disagreement can be explained by the different scales; in this case it should not be considered as an error.

The agreement between two land cover maps can be computed ast he overall % of pixels with coinciding codes, i.e. in the main diagonal of the disagreement matrix. The kappa statistic (Bishop et al, 1975) is a better measure of agreement: its value is close to 0 when the coincidences are random and 1 when both maps coincide perfectly. Table 3.1 shows both agreement measures before and after removing buffers. The comparison gives an idea of the part of the disagreement that can be attributed to co-location inaccuracy.

Agreement from a confusion matrix between CORINE and ILC Table 3.1.

	Total area (x 1 000 ha)	% coincident	kappa
no buffer	199	56	0.42
buffer 200 m	112	67	0.52

Table 3.2. **Commission and omission disagreement by pixel (in %)**

	Commission		Omission	
	No buffer	**buffer 200m**	**No buffer**	**buffer 200m**
Urban dense	89	87	46	37
Urban discontinuous	59	62	60	57
Industrial-commercial	64	62	36	21
Road-rail	96	99	90	88
Mines	40	32	17	3
Arable non irrigated	51	39	24	19
Vineyards	73	76	54	43
Fruits	46	47	30	22
Olive	49	41	35	24
Pastures	76	78	92	92
Arable+permanent crops	80	81	94	94
Complex agriculture	59	55	87	87
Agriculture + natural vegetation	78	75	92	92
Forest broad-leaved	27	16	14	10
Coniferous	60	58	33	18
Forest mixed	63	64	87	85
Natural grassland	86	88	60	67
Moors-heath	69	70	63	55
Wood-shrub	87	85	96	95
Water bodies	45	63	30	0

Table 3.2 reports commission and omission disagreements with and without buffer for categories that have at least 100 ha for both maps. We should stress again that **ixelwise disagreement can be very misleading:** both maps can be perfectly consistent with each other, and have a high % of disagreement by pixel because they represent the same reality at different scale. We shall see below that both maps have a very good agreement when the scale effect is removed.

8.3.2. Recoding into 9 classes

To simplify the analysis of different types of disagreement, we have recoded both maps into 9 major classes. **Figure 7** shows that both land cover maps have a similar pattern, with the obvious smoothing effect for CORINE land cover. Tables 5, 6 and 7 below report main agreement parameters when a pixelwise overlay is performed.

Table 3.3. **Agreement between CORINE and ILC with 9 classes**

	Total area (x 1 000 ha)	**% coincident**	**kappa**
no buffer	199	68	0.51
buffer 200 m	125	77	0.61

Commission and omission disagreement with 9 classes (%) Table 3.4.

	Commission		Omission	
	No buffer	buffer 200m	No buffer	buffer 200m
Urban	47.4	44.4	22.3	8.2
Arable	50.4	40.2	20.1	14.7
Permanent crops	51.6	49.9	33.5	23.6
Pastures	76.4	79.5	92.3	92.3
Heterogeneous	43.5	35.8	80.8	80.4
Forest	17.3	9.2	10.0	6.0
Other natural vegetation	61.5	63.6	69.5	66.3
Open spaces	80.2	93.8	42.2	18.4
Marsh and water	63.9	75.4	35.0	32.0

Pixelwise disagreeement matrix with 9 classes and 200 m buffer Table 3.5.

CORINE

	ISTAT land cover									
ha	Urban	Arable	Perm. crops	Past-ures	Hete-roge-neous	Forest	Other nat. veg.	Open spaces	Water	Total
Urban	2601	105	10	0	46	34	29	0	10	2834
Arable	532	13355	292	89	513	545	200	5	120	15650
Perm. crops	126	258	3160	18	219	299	53	0	4	4136
Pastures	66	1597	25	201	277	179	273	5	2	2623
Heterogeneous	1043	6442	2223	406	3162	2188	660	7	9	16141
Forest	180	393	527	163	435	72071	2916	16	8	76710
Other nat.veg.	129	178	68	104	262	3960	2405	25	3	7133
Open spaces	0	0	0	0	0	1	0	4	0	5
Water	0	1	0	0	0	0	17	6	51	74
Total	4677	22329	6305	981	4914	79277	6553	68	207	125306

Off-diagonal figures correspond at first sight to disagreements, but they do not take into account the scale effect. For example the 2 ha in the case *pastures×water*’ correspond to a small pond in a pasture area, and CORINE land cover respects its specifications not reporting it; therefore it cannot be considered as a real disagreement. There are hundreds of other false disagreements in the test site that distort Table 3.5.

8.3.3. Scale-corrected thematic disagreement by CORINE land cover olygon.

If a CLC polygon has 80 % of arable land, it is correctly classified according to the specifications of this land cover map. Similar criteria apply for the main CLC categories, excepting the ones that correspond to heterogeneous landscapes.

We have followed the procedure described below to remove the part of the disagreement between both land cover maps due to the different scales:

- The assessment is made on the nomenclature grouped into 9 classes. The class heterogeneous agriculture’ does not follow the rules below. We eliminate as well the class burnt areas’;
- A buffer of 200 m is removed around CORINE land cover polygon borders (100 at each side) to avoid counting location inaccuracy as disagreement;
- If more than 70 % of the pixels have the same code for the ILC map, the whole polygon is considered in agreement;
- If the % of different code is between 30 and 70, the polygon is considered partially in disagreement;
- If less than 30 % of the pixels have a different code for the ILC map, the whole polygon is considered in disagreement.

For example if a polygon of 100 ha (after removing the buffer) has been coded arable’ in CORINE land cover and ILC reports:

– 85 ha arable and 15 permanent crops, it will no contribute to the disagreement
– 60 ha arable and 40 permanent crops, it will contribute as 40 ha disagreement
– 20 ha arable and 80 permanent crops, it will contribute as 100 ha disagreement

Table 3.6. **Thematic disagreement with a grouped nomenclature**

CORINE land cover	area excluding buffer (Kha)	% disagreement
Urban	2.8	2.2
Arable	15.2	1.4
Permanent crops	3.9	6.4
Pastures	2.3	93.6
Heterogeneous	13.0	.
Forest	76.2	0.3
Other nat. veg.	6.8	68.6
Open spaces	0.2	0
Marsh and water	0.1	32

Table 3.6 reports the disagreement rates. This table confirms several expected facts:

- The agreement is very good for urban, arable, forest, and open spaces.
- There are major discrepancies on pasture and natural vegetation, that represent 7.5 % of the area (excluding buffers). This divergence probably comes from the difficulty to interpret the nomenclature.

Other figures are more surprising but may have some explanation:

- The disagreement for marsh and water' comes from a single marsh polygon.
- The low disagreement of permanent crops increases our confidence in CLC for this land cover class, especially difficult to photo-interpret.

References

Bishop Y., Fienberg S., Holland P., 1975, *Discrete Multivariate Analysis*, M.I.T. press, Cambridge, Ma.

EC, 1993, *CORINE land cover; guide technique*, Report EUR 12585EN. Office for Official Publications of the European Communities. Luxembourg,. 144 pp.

Gallego, F.J. Carfagna, E., Peedell S., 1999, The use of CORINE land cover to improve area frame survey estimates. *Research in Official Statistics*, Vol 2, no 2, pp. 99-122.

ISTAT, 1998, — *Capitolato tecnico per il progetto pilota del database sull'uso e la copertura del suolo in scala 1:25.000 su un'area test.* Rome.

Napolitano P., Carbonetti G., Gallego J., 2000, Accuracy Assessment and Validation of a Land Use-land cover Database in Arezzo (Italy). Proceedings of *Accuracy 2000*, Amsterdam, July 12-14, 2000, pp. 233-236.

9. Conclusions

This joint publication Towards agri-environmental indicators: Integrating statistical and administrative data with land cover information' is a follow-up to the work presented in the report From land cover to landscape diversity in the European Union' published by the Agriculture Directorate-General in 2000.

This new work takes into consideration the three main conclusions expressed in the previous report, related to:

- the feasibility and relevance of automatic computation of diversity indices using CORINE (Coordination of information on the environment) land cover data;
- the limitations and confusions encountered due to heterogeneous classes and working scale;
- the availability of alternative/ complementary material.

In addition, it is in line with the recommendations of the Commission communication COM(2000) 20 on Indicators for the integration of environmental concerns into the CAP', especially concerning the better use of already available information.

Topics covered

- Most of this second publication concentrates on the problem of spatial transformation of data. Input and output space characteristics are discussed. The role of the CORINE land cover covariable is identified in terms of mean value, possible modelling of the bivariate spatial distribution, resolution and scale.
- In terms of thematic applications, various additional economic sectors are involved. The Natura 2000 database is explored in terms of its land cover content and comparisons established with CORINE land cover. Spatial redistribution of NUTS (Territorial units for statistics) population statistics is also covered. Through modelling, strategies of calculation of nitrate surpluses are evaluated.
- Statistical data, traditionally available at administrative level, are merged with georeferenced land cover data and maps are presented pointing out the main differences between aggregates of agricultural land cover classes. Proposals are made to reallocate Farm Structure Survey data at relevant geographical levels.
- The integration of administrative data is explored through examples taken from the Integrated Administration and Control System (IACS). The potential of using administrative data for the evaluation of agricultural diversity indicators is explored, and its application to further characterise the agricultural content of the CORINE land cover heterogeneous classes is analysed.
- In term of geographic coverage, applications deal with the usual EU-15 Member States as well as with selected cases referring to the candidate countries.

Main questions

The major question raised in this publication is related to the integration process of different data sources. Ideally, accurate source data should be merged with fine resolution spatial covariables having together a strong correlation, so that reliable spatialised data can be output. Compatibility of nomenclatures eases the comparisons but is in no way mandatory, as shown with the example of redistribution of population data with CORINE land cover data. Limitations of representation through administrative units are reported and compared with grid and thematic unit presentations.

In terms of boundaries in the disaggregation/aggregation process, it should be stressed that no new information is created during the process and that relevant accurate output statistics require accurate information from the outset, at least at the same resolution as this output.

The results

A first conclusion is that data coregistration does not pose serious problems. Even in the case of slightly different nomenclatures, links can be established between classification systems and geometrical matches obtained. On that basis, most papers conclude that the different data sources are unlikely to be equivalent but are complementary, e.g.: IACS data in Belgium and Italy behave rather differently from CORINE land cover data. At different mapping scales, differences are observed on land cover statistics derived from image interpretation.

The GIS for Natura 2000 can benefit from the CORINE land cover database. Disaggregation of land cover types should fit with the definitions in the habitats directive. Additional detailed information available in European Community programmes (e.g. LIFE (a financial instrument for the environment)) could fill in the level of detail required for the characterisation of designated sites.

Data spatialisation requires special attention in most cases and needs validation. CORINE land cover can be used to improve the mapping of population density. Population redistribution at commune level in the case of Arezzo province (Italy) leads to important localisation errors, but this is mainly due to a scale effect.

Administrative data offers encouraging possibilities and the extraction of agricultural area from the heterogeneous classes appears to be promising. In Perugia (Italy), estimates of agricultural share are around 34 %, but additional work is required to understand some divergences in the results.

The influence of the retained transformation process varies between applications. While it seems very important in the population redistribution example, the contrary appears to be the case with nitrogen surpluses evaluation. In the case of nutrient surpluses modelling, the use of CORINE land cover compensates for the absence of available spatially homogeneous statistics.

Future directions

Based on the results obtained, more effort is still needed to better exploit the available data. This implies the realisation of simple GIS applications to facilitate access to information and the comparison of independent data sources, so that the spatial integration of standard statistics with mapped land cover can be processed on significant datasets.

More work is certainly required on the modelling inherent in any future data merge. On the basis of validation procedures, strategies have to be compared and best practices defined. Additional effort should be devoted to the definition of the output space, especially with regard to the thematic stratification.

Compared to the available datasets, only a few of the possible computations have been initiated so far. Statistical and administrative data should be integrated with land cover information in large-scale programmes in order to start the real production of agri-environmental indicators.

Finally, these studies face the inherent limitations of CORINE land cover data, due to the adopted scale and nomenclature. The definition at the European level of large-scale digital maps, adapted to the needs of rural area management, should be considered.

Glossary

Agenda 2000:
The Agenda 2000 is a legislative package conceived at the Madrid European Council in December 1995.

At that meeting, the Commission was invited to prepare a communication on the Union's future financial framework, having regard to the prospects of enlargement.

The communication [41] highlighted a number of priorities in particular: the need to maintain the policy of economic and social cohesion, to pursue the reform of the common agricultural policy, to strengthen growth, employment and living conditions through the Union's internal policies and to allow the accession of new members, while maintaining budgetary discipline.

To translate these priorities into legal instruments, the Commission presented in March 1998 legislative proposals. The European Council, at its meeting in Berlin in March 1999, reached a political agreement on the Commission's proposals.

The resulting package of about twenty legislative measures covers four main areas: the reform of the common agricultural policy, structural policy reform, the pre-accession instruments and the new financial framework.

agri-environmental indicators:
Generic term designating a range of *indicators* aiming at giving synthesised information on complex interactions between agriculture and environment. Common agri-environmental indicators are those that provide an assessment of impacts of agriculture on water quality, climate change, soil or landscape structures.

aggregation (s atial):
Summarising of data of adjacent geographical units that are merged into units putting as far as possible like with like.

CAP:
Common Agricultural Policy. Since its creation in 1962 the CAP has played a key role in the EU's development. The progressive implementation of common market organisations which cover the overwhelming bulk of the EU's agricultural production has been accompanied by structural policies which reflect the various facets of the CAP including the important social role of agriculture in the European Union, its regional and national diversity and the need to take account of consumer and environmental concerns.

The reform of the CAP in 1992 and subsequently has shifted somewhat the previous dominance of market measures towards the provision of a greater role for rural development. The new focus is on meeting the challenges posed by the depopulation, abandonment of many rural areas and environmental impacts.

CARTHAGE (BD):
French geographical database, hydrographic object oriented (water streams and basins).

CORINE:
Coordination of Information on the Environment. A Program proposed in 1985 by the European Commission, aimed at gathering information relating to environment on certain priority topics for the European Union (land cover, Coastal Erosion, Biotopes, etc).

CLC:
Acronym for CORINE land cover.

diversity:
Various different objects. Conversely to *heterogeneity*, it does not consider relations between them. Diversity relies on two components: richness and evenness.

DEM:
Digital Elevation Model

A digital representation of a continuous variable over a two- dimensional surface by a regular array of z values.

(41) 'Agenda 2000: For a stronger and wider Union' [COM(97) 2000]

DG AGRICULTURE:
The General Directorate of Agriculture, service based in Brussels, is the Commission department responsible for the implementation of the European Union's policies on agriculture and rural development.

DG ENVIRONMENT:
The General Directorate of Environment, service based in Brussels, is the Commission department responsible among others for the integration of environmental concerns into EU policies.

disaggregation (s atial):
Spatial transfer of data from units to embedded sub-units. A common transfer is based on the principle of areal weighting.

DPSIR:
Driving Forces, Pressures, State, Impact, Response, which constitutes the conceptual model to approach the environmental problems defined by the European Environment Agency.

DTM:
Digital Terrain Model. Digital elevation model used to represent terrain relief.

EFTA:
The European Free Trade Association is an international organisation comprising four states, Iceland, Liechtenstein, Norway and Switzerland. It was established in 1960.

Euro ean Environment Agency:
EEA was launched by the European Union in 1993 with a mandate to orchestrate, cross-check and put to strategic use information of relevance to the protection and improvement of Europe's environment. The Agency, based in Copenhagen, Denmark, has a mandate (EEC-1210/90) defined to ensure the supply of objective, reliable and comprehensive information at European level, enabling its member states to take the requisite measures to protect their environment, to assess the result of such measures and to ensure that the public is properly informed about the state of the environment.

EUROFARM:
Storage system of the Farm Structure Survey's data.

Eurostat:
Statistical Office of the European Communities. Eurostat based in Luxembourg is one of the directorates general of the Commission. Its mission is to provide the European Union with a statistical information service. Eurostat uses uniform rules to collect statistical data from the official statistical services, in particular the National Statistical Institutes of each of the 15 Member States of the European Union.

Farm Structure Survey:
Farm Structure Survey is a European survey on agricultural holdings. It consists of a census organised every ten years to which are added intermediate surveys by sample survey every two or three years. The first survey, carried out in 1966/67, arose from the need to have harmonised information at the Community level. Since then, regulatory texts have defined the methodological framework and the contents of the survey's questionnaires.

GIS:
Geographic Information System. An organised collection of specific computer hardware, software, geographic data and personnel designed to efficiently capture, store, update, manipulate, analyse and display all forms of geographically referenced information (i.e. raster/vector) that can be drawn from different sources.

GISCO:
Geographic Information System of the Commission. The GISCO is based in Eurostat.

heterogeneity:
Diverse character of a designated group of different objects .

hydrosol:
Database developed at IFEN resulting from the combination of CLC, commune boundaries and watershed boundaries.

IACS:
Integrated Administration and Control System. It has been designed to monitor claims for area aid. The principle is the following: a farmer declares the parcels for which he claims aid payments. The IACS verify the truthfulness of the declaration in order to enable the payment.

The producer receives payments only for certain types of culture and for the declared parcels (aid is proportional to the area up to a certain theoretical amount of production).

IFEN
Institut Français de l'Environnement (French Institue of Environment).

indicator:
Observed value representative of a phenomenon to study. In general, indicators quantify information by aggregating different and multiple data. The resulting information is therefore synthesised. In short, indicators simplify information that can help to reveal complex phenomenon.

integration (data):
Process of unifying existing data sources into a single framework. It consists in general in the following phases:

– identification and analysis of the relevant data,
– harmonisation,
– incorporation and generalisation

ISTAT:
Istituto Nazionale di Statistica (Italian Statistical Institute)

JRC:
Joint Research Centre. The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of the EU policies. As a directorate general of the European Commission, the JRC functions as a centre of reference in science and technology for the European Union.

land cover:
Land cover corresponds to a (bio)physical description of the earth's surface. It is that which overlays or currently covers the ground. This description enables various biophysical categories to be distinguished - basically, areas of vegetation (trees, bushes, fields, lawns), bare soil, hard surfaces (rocks, buildings) and wet areas and bodies of water (watercourses, wetlands).

land use:
Land use corresponds to the socio-economic description (*functional dimension)* of areas: areas used for residential, industrial or commercial purposes, for farming or forestry, for recreational or conservation purposes, etc. Links with land cover are possible; it may be possible to infer land use from land cover and conversely. But situations are often complicated and the link is not so evident. Contrary to land cover, land use is difficult to "observe". For example, it is often difficult to decide if grasslands are used or not for agricultural purposes. Distinctions between land use and land cover and their definition have impacts on the development of classification systems, data collection and information systems in general.

landsca e:
Complex concept encompassing several definitions. The one adopted in this publication is considering landscape as an area containing a mosaic of land cover patches. Third dimension of space and its effect on visual perception are not taken into account. Only the spatial configuration is considered which however influences landscape physiognomy.

LIFE:
L'Instrument Financier pour l'Environnement.
LIFE is a financial instrument for three major areas of action: Environment, Nature and Third Countries. While all three areas aim to improve the environment, each has its specific priorities. LIFE is implemented since 1992.

LUCAS:
Land Use / Cover Area Frame Statistical Survey. LUCAS is an EU pilot project (1999 – 2003) for the application of area-frame survey to the agricultural statistics.

Natura 2000:
European network of protected sites for nature conservation.

nomenclature:
A nomenclature is a list of categories, summarising information in a highly reduced form while attempting to maintain a maximum information content. A nomenclature normally covers a particular field of interest.

NUTS:
Nomenclature of Territorial Units for Statistics. Established by Eurostat, provides a single uniform breakdown of territorial units. NUTS subdivides each Member State into a hierarchy of increasingly smaller administrative areas.

PHARE:
Poland and **H**ungary: **A**ction for the **R**estructuring of the **E**conomy.

The Phare Programme is the EU's financial instrument designed to assist and support partner countries in the reintegration of their economies and societies with Western Europe. The Phare Programme was established in December 1989 on economic aid to the Republic of Hungary and the Polish People's Republic. It now covers 14 eastern Europe countries.

ixel:
Picture element.

raster:
One of the two major types of internal data organisation used in *GIS* (*see also vector*). Raster systems superimpose a regular grid over the area of interest and associate each cell or pixel, with one or more data records. The value associated with each grid cell may represent either real values or any scalar or nominal data values associated with the cell coordinate.

SABE:
Seamless Administrative Boundaries of Europe. SABE is a data set compiled from source data provided by 29 of Europe's official national mapping organisations, members of EuroGeographics Association. It contains all administrative units from the country level down to commune (NUTS 5).

SCEES:
Service Centrale des Enquêtes et Études Statistiques. Statistical service of the French ministry of agriculture.

SCI:
Sites of Community Importance. One of the two types of Natura 2000 areas designated for the protection of wild fauna and flora.

Shannon (index):
The Shannon index quantifies the diversity of the countryside based on two components: the number of different patch types and the proportional area distribution among patch types. Commonly the two components are named richness and evenness. Richness refers to the number of patch types (compositional component) and evenness to the area distribution of classes (structural component).

SPA:
Special Protection Areas. One of the two types of Natura 2000 areas designated for the protection of wild birds.

standardisation:
Statistical technique allowing variables whose values are expressed in different measurement units to be compared. In general, the technique consists of dividing the difference of each value to the mean by the standard deviation.

SUPCOM:
Research Support Activities on a Competitive Basis. Since 1995 Eurostat has benefited from a specific research budget line of the 4th Framework Programme budget. Eurostat used this opportunity to launch almost 100 R&D studies concerning a wide variety of statistical topics such as statistical methodology, treatment of metadata, information technologies, transport, environment statistics. These projects were won through open tender by the contractors. Almost 60% of these projects have now been completed and the results are progressively made available on the EU web site.

vector:
One of the two major types of internal data organisation used in *GIS* (*see also raster*). Vector systems are based primarily on coordinate geometry and take advantage of the convenient division of spatial data into point, line and polygon types.