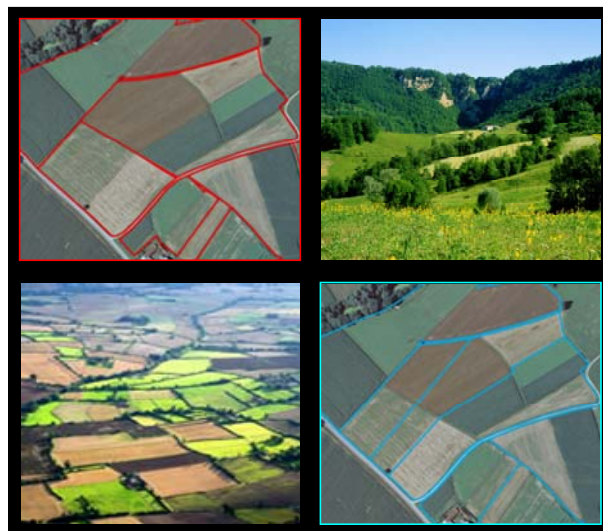




# Land Parcel Identification System (LPIS) Anomalies' Sampling and Spatial Pattern

Towards convergence of ecological methodologies and  
GIS technologies

David Grandgirard and Rafal Zielinski



EUR 23484 EN - 2008

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission  
Joint Research Centre  
Institute for the Protection and Security of the Citizen

**Contact information**

Address: Via Enrico Fermi 2749, I-210227 Ispra (VA), Italy  
E-mail: [david.grandgirard@jrc.it](mailto:david.grandgirard@jrc.it), [rafal.zielinski@jrc.it](mailto:rafal.zielinski@jrc.it)  
Tel.: +39 0332 78 5503  
Fax: +39 0332 78 9029

<http://ipsc.jrc.ec.europa.eu/>  
<http://www.jrc.ec.europa.eu/>

**Legal Notice**

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

***Europe Direct is a service to help you find answers  
to your questions about the European Union***

**Freephone number (\*):**

**00 800 6 7 8 9 10 11**

(\* ) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

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/>).

JRC 46971

EUR 23484 EN  
ISBN 978-92-79-09701-0  
ISSN 1018-5593  
DOI 10.2788/91912

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 2008

Reproduction is authorised provided the source is acknowledged

*Printed in Italy*

## ***Summary***

To date, the Land Parcel Identification System (LPIS) has often been proposed as the foundation for effective spatial management of agriculture and the environment and many land managers have suggested incorporating it in most of the instruments for sustainable agriculture. The LPIS is originally used for registration of agricultural reference parcels considered eligible for annual payments of European Common Agricultural Policy (CAP) subsidies to farmers. Its intrinsic quality depends on the frequency and magnitude of the discrepancies in area, since some parcels can be under- or over-declared by farmers compared with reference registered within the LPIS. General application of the LPIS therefore depends on our capacity to – first identify and explain the causes of these area discrepancies perceived as anomalies by national CAP payment agencies –second, to propose future improvements in its overall quality.

From a set of images used during the 2005 Control with Remote Sensing (CwRS) campaign, using the geographic information system (GIS) and ecological methodologies we assessed the quality of the LPIS by identifying the diversity of the existing anomalies. To that end, the ecological sampling method was adapted to the specific case of image-based detection of anomalies. The observed anomalies assemblages obtained from a set of European Member States representing the four types of LPIS were analysed to establish the spatial pattern of the anomalies.

We showed that the twelve zones surveyed can be grouped into four different clusters, each individually correlated with the presence of certain categories of LPIS anomaly. Some clusters were more particularly related to the presence of natural and anthropogenic landscape features, whereas others were typified by anomalies which stemmed from the process for creating and updating the LPIS, which accounted for 20% of the anomalies detected. Finally, we also showed that, even if useful for establishing procedures to manage the LPIS, the LPIS typology used in the European Union had no effect on the anomalies assemblage or on the spatial pattern; consequently, the type of LPIS no longer needs to be considered and LPIS anomalies assemblages could be pooled across Europe.

In the light of the results obtained, different proposals are made to improve LPIS quality by:

- identifying the critical points along the LPIS management chain;
- using landscape ecological methodologies to explain the causes of the clusters observed; and
- extrapolating the whole results in the CwRS risk analysis to perform *ex-ante* LPIS anomalies risk map.

## ***Keywords***

Land Parcel Identification System, Control with Remote Sensing, orthophoto, quality assessment, diversity, spatial pattern, landscape structure

## *Preface*

This report attempts to present, in a coherent form, to European national authorities in charge of the Single Payment Scheme (SPS) under the CAP regulatory framework or more generally to CAP managers the advantages of making simultaneous use of both ecological and geographic information system (GIS) methodologies to address the current quality of the Land Parcel Identification System (LPIS) from which payments to millions of farmers every year are decided.

The reason for conducting a survey of the LPIS across the European Union (EU) was not dictated by any future revision of the regulatory framework, but that agricultural and rural development managers wanted to make general use of the LPIS. Indeed, the LPIS has recently come to be considered the central geographic database able to provide the most accurate and up-to-date information about land occupation and land-use changes. Today, for European administrations and scientists alike, the LPIS is acknowledged as the foundation for tackling the sustainability of European agriculture cost-effectively. Because of the volume of georeferenced information it conveys concerning land occupation, cultivated biodiversity and areas identified as of high natural value, the LPIS is perceived as a means to address land management, biofuel and biomass production, preservation of agri-environmental resources or even problematic biotechnological crops. Combined with information and communication technologies, the LPIS then becomes an indispensable instrument for monitoring, reporting, certifying or assessing agri-environmental practices and possible effects of CAP reforms.

Consequently, by studying LPIS quality and proposing potential improvements, we expect to play an active part in its general application to everyone concerned by agri-environmental sustainability. To that end, we assessed LPIS quality by looking at the diversity of the anomalies it could contain and by analysing the reasons for them. The results are reported here in an easy-to-understand, but very detailed manner to allow anybody else to repeat the methods used here or adapt them to their specific efforts.

This quality assessment of the LPIS was made possible by the multidisciplinary nature of the JRC GeoCAP staff and by the flexibility we were allowed to develop ideas and conduct experiments. For that, we are grateful to our Head of Unit Jacques Delincé and to our respective project leaders Simon Kay and Philippe Loudjani. We also wish to thank Florence Carré and Ivan Buck for reviewing this document. Finally, we want to thank Charles J. Krebs for the wealth of knowledge he made available in his reference work: *Ecological Methodology*.

David Grandgirard and Rafal Zielinski

May 2008

## *Contents*

List of figures and tables

List of equations

List of abbreviations

Part A. European LPIS: implementation and utilisation framework	p. 1
Introduction	p. 1
1. Regulatory framework	p. 2
2. LPIS quality	p. 2
3. Creation and diversity of the LPIS	p. 3
4. LPIS updating	p. 5
5. LPIS irregularities	p. 6
Part B. European LPIS appraisal: ecological approach and subsequent objectives	p. 7
6. Ecological approach	p. 7
7. Objectives	p. 8
7.1. Typology of LPIS anomalies	p. 8
7.2. Validation of a sampling method for the survey of LPIS anomalies	p. 9
7.3. Description of LPIS anomalies assemblages	p. 10
7.4. Relationship between LPIS anomalies assemblages and landscape structure and LPIS management	p. 10
Part C. Data used and preparation	p. 12
8. LPIS dataset selection and preparation	p. 12
Part D. Sampling method selection and calibration	p. 15
9. <i>Ex-ante</i> assumptions and sampling method selection	p. 15
9.1. Initial assumptions	p. 15
9.2. Variables of interest	p. 17
10. Sample size	p. 17
11. Quadrat shape	p. 20
12. Quadrat size and sampling design	p. 21
12.1. Wiegert's method	p. 21
12.2. Sampling designs	p. 21
12.3. Wiegert's cost calculations	p. 24
12.4. Results of Wiegert's method	p. 24
12.5. Sampling design control method	p. 25
13. Sampling decision based on comparison of LPIS anomalies assemblages	p. 27
14. Spatial distribution of anomalies in zone FB2	p. 30
14.1. Statistical assessment of the spatial distribution of anomalies in zone FB2	p. 30
14.2. Metric assessment of the spatial distribution of anomalies in zone FB2	p. 30
14.3. Results of the statistical estimation of the spatial distribution of anomalies in zone FB2	p. 32

14.4. Results of the metric estimation of the spatial distribution of anomalies in zone FB2	p. 34
15. Preliminary conclusions and decisions concerning the final sampling method	p. 35
Part E. Analysis of diversity in LPIS anomalies	p. 36
16. Validation of sampling method	p. 36
16.1. Anomalies accumulation curves theory	p. 36
16.2. Non-parametric estimators	p. 38
16.3. Anomalies accumulation curves for completeness of the sampling method	p. 38
16.4. Non-parametric estimators for completeness of the sampling method	p. 42
16.5. Final statement on the comparability of anomalies assemblages	p. 42
17. Description of samples	p. 42
17.1. Areas sampled	p. 42
17.2. Anomalies occurrence and abundance	p. 45
17.3. Relationship between anomalies abundance and reference parcel area	p. 47
18. Description of anomalies assemblages	p. 49
18.1. Single anomaly detectability and $\gamma$ -diversity	p. 49
18.2. Single anomaly occurrence and abundance	p. 49
19. Diversity of LPIS anomalies	p. 51
19.1. $\alpha$ -diversity	p. 51
19.2. $\beta$ -diversity	p. 53
19.3. Anomalies clusters and spatial pattern	p. 54
Part F. Final discussion and prospects	p. 58
20. Discussion of the results	p. 58
• The sampling method	p. 58
• Validation of the sampling method	p. 59
• LPIS quality in the EU	p. 60
• The diversity of LPIS anomalies	p. 61
• Typology of LPIS anomalies and causes of the anomalies assemblage	p. 61
21. Prospects	p. 62
• The LPIS updating process	p. 62
• The landscape ecological approach	p. 63
Bibliography	p. 65
Appendices	p. 70
• Appendix A: LPIS anomalies typology proposed <i>a priori</i>	p. 70
• Appendix B: Anomalies accumulation curves from Clench and LD asymptotic models	p. 71
• Appendix C: LPIS anomalies assemblages from the twelve zones studied	p. 72

## *List of figures and tables*

Figure 1: “75/90%” LPIS quality rule

Figure 2: Workflow of LPIS creation and updating processes

Figure 3: Description of the four types of LPIS existing in Europe

Figure 4: General workflow for the 2005 LPIS quality survey

Figure 5: Distribution of the control zones for the CwRS 2005 project

Figure 6: Decision-making tree for choosing a method for estimating absolute abundance

Figure 7: The three different sampling designs applied

Figure 8: Changes in product of (relative variance \* relative cost) depending on quadrat size and sampling design

Figure 9: Relative bias, imprecision and  $\sqrt{MSE}$  of the mean and the standard error of anomaly occurrence in the LPIS for FB2

Figure 10: NMDS plot of Morisita’s similarity index calculated between the true FB2 anomalies population and the anomalies population sampled and replicated

Figure 11: Observed and expected frequency distributions of the number of LPIS anomalies per quadrat

Figure 12: Observed against estimated values of anomalies richness

Figure 13: Type-by-type relative percentage of anomalies occurrence and abundance for each of the twelve selected zones

Figure 14: Distribution of the mean number of anomalies per reference parcel for all twelve selected zones

Figure 15: Mean reference parcel area for the eight different classes of anomaly abundance observed during the survey

Figure 16: Classification of the product of (relative abundance \* relative occurrence) for the sixteen anomalies detected

Figure 17: Non-metric multidimensional scaling and hierarchical ascendant classification of similarity measurements between zones

---

Table 1: Replicates and corresponding acronyms of the four LPIS types surveyed

Table 2: *A priori* estimate of the minimum sample size for each LPIS replicate in the survey

Table 3: Expected uniformity, aggregation and randomness values of indices of dispersion

Table 4: Predicted maximum richness values from anomalies accumulation curves and non-parametric estimators

Table 5: Primary results concerning anomalies occurrence and abundances in the twelve zones selected for the survey

Table 6:  $\alpha$ -diversity metrics zone by zone

Table 7: Relative abundance of single anomalies in the identified clusters of zones

## *List of equations*

**Equation 1: Anomaly occurrence**

**Equation 2: Anomaly abundance**

**Equation 3: Normal approximation to the confidence interval**

**Equation 4: Desired margin of error**

**Equation 5: Sample size for an infinite population**

**Equation 6: Correction of sample size for finite population**

**Equation 7: Wiegert's sampling cost**

**Equation 8: Morisita's similarity index**

**Equation 9: Component  $\lambda_1$  of Morisita's similarity index**

**Equation 10: Component  $\lambda_2$  of Morisita's similarity index**

**Equation 11: Green's coefficient of dispersion**

**Equation 12: Morisita's index of dispersion**

**Equation 13: Uniform dispersion index**

**Equation 14: Clumped dispersion index**

**Equation 15: Declination 1 of the standardised Morisita's index of dispersion**

**Equation 16: Declination 2 of the standardised Morisita's index of dispersion**

**Equation 17: Declination 3 of the standardised Morisita's index of dispersion**

**Equation 18: Declination 4 of the standardised Morisita's index of dispersion**

**Equation 19:  $\chi^2$  goodness-of-fit of the Poisson distribution of anomalies per quadrat**

**Equation 20: Probabilistic determination of the negative binomial distribution of anomalies per quadrat**

**Equation 21: Clench model**

**Equation 22: 95% and 90% estimators of the Clench model**

**Equation 23: Linear dependent model**

**Equation 24: 95% and 90% estimators of the linear dependent model**

**Equation 25: Gompertz model**

**Equation 26: 95% and 90% estimators of the Gompertz model**

**Equation 27: Simpson's reciprocal index of diversity**

**Equation 28: Camargo's index of evenness**



## *List of abbreviations*

<b>1/D</b> = Simpson's index of heterogeneity	<b>IACS</b> = Integrated Administration and Control System
<b>2D</b> = 2-dimensional representation	<b>Jack1</b> = First-order Jackknife non-metric estimator of richness
<b>3D</b> = 3-dimensional representation	<b>Jack2</b> = Second-order Jackknife non-metric estimator of richness
<b>ANOVA</b> = ANalysis Of VAriance	<b>JRC</b> = Joint Research Centre
<b>A<sub>obs</sub></b> = OBServed Abundance of anomalies	<b>Km<sup>2</sup></b> = Square kilometre
<b>AP</b> = Agricultural Parcel	<b>LD model</b> = Linear Dependent model
<b>A<sub>true</sub></b> = TRUE Abundance of anomalies	<b>LFA</b> = Less Favoured Area
<b>BRA<sub>i</sub></b> = Class of Boundary-Related Anomalies (i = from 1 to 4)	<b>LPIS</b> = Land Parcel Identification System
<b>CA</b> = Correspondence Analysis	<b>MRA<sub>i</sub></b> = Class of Man-made Related Anomalies (i = from 1 to 5)
<b>CAP</b> = Common Agricultural Policy	<b>MSE</b> = Mean Square Error
<b>CAPI</b> = Computer-Assisted Photo-Interpretation	<b>NMDS</b> = Non-metric MultiDimensional Scaling
<b>Chao1</b> = First-order Chao non-metric estimator of richness	<b>OTS</b> = On-The-Spot check
<b>Chao2</b> = Second-order Chao non-metric estimator of richness	<b>PB</b> = Physical Block
<b>CLU</b> = Systematic CLUster sampling design	<b>PCA</b> = Principal Component Analysis
<b>CP</b> = Cadastral Parcel	<b>POP</b> = Total anomalies assemblage of the pre-test zone (FB2) as the "true" anomalies census
<b>CwRS</b> = Control with Remote Sensing	<b>r<sup>2</sup></b> = Linear regression coefficient
<b>E'</b> = Camargo's index of evenness	<b>RAND</b> = RANDom sampling design
<b>E<sub>45</sub></b> = Completeness percentage of the true anomalies assemblage when sampling 45 km <sup>2</sup>	<b>RDR</b> = Rural Development Regulation
<b>E<sub>90%</sub> (E<sub>95%</sub>)</b> = Number of quadrats necessary to sample 90% (95%) of the estimated maximum richness in an anomalies assemblage	<b>SD</b> = Standard Deviation
<b>EC</b> = European Council	<b>SE</b> = Standard Error
<b>EU</b> = European Union	<b>SMR</b> = Statutory Management Requirements
<b>F</b> = F statistic from ANOVA	<b>S<sub>obs</sub></b> = OBServed anomalies richness
<b>FB</b> = Farmer Block	<b>SPS</b> = Single Payment Scheme
<b>GAEC</b> = Good Agricultural and Environmental Conditions	<b>S<sub>true</sub></b> = TRUE anomalies richness
<b>GIS</b> = Geographic Information System	<b>SYS</b> = SYStematic sampling design
<b>GPS</b> = Global Positioning System (d-GPS = differential Global Positioning System)	<b>TRA<sub>i</sub></b> = Class of Tree-Related Anomalies (i = from 1 to 4)
<b>Ha</b> = hectare	<b>VHR</b> = Very High Resolution
<b>HAC</b> = Hierarchic Ascendant Classification	<b>WRA<sub>i</sub></b> = Class of Water-Related Anomalies (i = from 1 to 4)
	<b>WTO</b> = World Trade Organization

## ***PART A. European LPIS: implementation and utilisation framework***

### **Introduction**

Examination of the successive major Common Agricultural Policy (CAP) reforms in Europe clearly reveals the consequences of the trade negotiations undertaken at World Trade Organization (WTO) level and of citizens' expectations concerning the sustainability of agriculture and rural development. Indeed, from the productive agriculture of the '90s when farmers' entitlements were directly proportional to the production yields set to safeguard food security to the latest 2003 "Fischler" reform, environmental and social considerations have been progressively integrated into CAP regulations. To date, CAP reforms have involved significant reductions in most of the main forms of support by decoupling them from production and by implementing the Single Payment Scheme (SPS) (Buckwell, 2007). Nowadays entitlements are mostly calculated from the eligible farm area, whatever its production. At the same time, environmental and other non-commodity based support for agriculture, forestry and rural development has been restructured in the form of the new 2007-2013 Rural Development Regulation (RDR) (Commission Regulation (EC) No 817/2004). Together, the first and second pillars in the CAP Regulation (concerning, respectively, management of support for agricultural production and environmental/rural sustainability) now encourage integration of environmental, social and economic concerns into agriculture.

Under the first pillar of the CAP, this was done by (i) setting up a new Integrated Administration and Control System (IACS) involving mapping for management of all agricultural land and (ii) making the SPS payments due to over seven million farmers in Europe conditional on meeting cross-compliance standards (Good Agricultural and Environmental Conditions – GAEC – and Statutory Management Requirements – SMRs) (Council Regulation (EC) No 1782/2003) in order to justify farmers' CAP subsidies. On the other hand, the second pillar of the CAP proposes voluntary agri-environmental schemes to support agricultural and rural activities in less favoured areas (LFA), to promote marketing of agricultural products and conversion to organic farming or to assist farmers starting up or taking early retirement (Council Regulation (EC) No 1257/1999).

Even if it could be thought that current economic concerns could negatively counterbalance the environmental benefits expected from agriculture (Bennet *et al.*, 2006), cross-compliance under the CAP Regulation appears the minimum necessary to limit and control risks. To that end, cross-compliance standards address (i) food market safety by allowing registration and identification of animals, animal welfare and human, animal and plant health; and (ii) protection of environmental resources (soil, air, water, biodiversity, etc.) to ensure the sustainability of agriculture. From cross-compliance, where GAECs cast an environmentally sensitive light on farms by targeting internal characteristics of agricultural holdings and cropping systems, SMR integrate the surrounding environment in which agricultural holdings are located (the Nitrates and Birds/Habitats Directives are two good examples of SMRs). Today, each individual plot of agricultural land is taken into account, by considering the uses made of the land and its surrounding landscape. Scientists and policy-makers alike use the term "agri-environment" to refer to this new level of integration within the CAP. Council Regulation (EC) No 1257/1999 (Chapter IV) defines the concept "agri-environment" as "*support for agricultural production methods designed to protect the environment and to maintain the countryside ... it shall promote ways of using agricultural land which are compatible with the protection and improvement of the*

*environment, the landscape and its features, natural resources, the soil and genetic diversity.”* Based on that, “agri-environment” appears to be the crossroads where the environmental performance of European agriculture can be assessed, not only at farm but also at higher levels when evaluating, respectively, cropping system sustainability and the effectiveness of CAP management.

## **1. Regulatory framework**

CAP performance can be partly assessed by combining agricultural parcels, defined as “*a continuous area of land on which a single crop group is cultivated by a single farmer*” (Council Regulation (EC) No 972/2007), with the particular landscape features within or next to them. Performance assessment can be considered relatively feasible locally (at agricultural parcel or farm levels), but becomes more complex when undertaken on a larger scale. To solve this difficulty, an integrated system is needed for monitoring and reporting on farmers’ compliance with CAP requirements. Such a system should make it possible to merge and manipulate datasets containing not only agricultural but also environmental information.

This system is already available. It has been clearly defined as the Integrated Administration and Control System (IACS) in the CAP Regulation. As required by Council Regulation (EC) No 1782/2003 (Article 17) and Commission Regulation (EC) No 796/2004, the IACS manages farmers’ applications and direct support payments at national level in EU Member States. The IACS is a GIS-assisted (geographic information system) computerised database containing at least:

- An identification system for location and measurement of “reference parcels” declared in applications for agricultural aid, i.e. the “Land Parcel Identification System (LPIS)”. “Reference parcel” means a geographically delimited area with a unique identification code under which it is registered in the Member State’s GIS identification system (Council Regulation (EC) No 796/2004); besides reference parcels, features of the surrounding landscape are often registered in the LPIS as supplementary elements to be excluded from individual calculations of farmers’ area-based aid.
- Further modules for identification of entitlements, registration of aid applications, identification of farmers and an integrated control system.

Despite its primary objective of supporting CAP schemes, the LPIS is increasingly coming to be considered a very accurate, cost-effective system for land-management purposes (Leteinturier *et al.*, 2006). Therefore, an increasing number of rural agencies or local agricultural companies propose to have recourse to the LPIS and related datasets to build, assess, monitor, report and even certify management activities (Milčinski *et al.*, 2007). However, effective use of the LPIS for monitoring and reporting on the environmental (and economic) performance of agriculture strongly depends on the intrinsic quality of the LPIS.

## **2. LPIS quality**

For the purposes of the CAP Regulation, LPIS quality is high when the reference parcel system in use is adequate. Adequacy is demonstrated by complying with the “75/90%” rule. This is the case if the reference parcels in the 75th percentile of the whole reference parcel population at national level present an “up-to-date reference area/total LPIS reference area” ratio equal to or higher than 90%. The “up-to-date reference area” is measured from up-to-date satellite imagery and/or with certified GPS devices and is considered the “true” eligible area. The “total LPIS reference area” is the total area actually registered in the LPIS and considered the reference eligible area. This rule

calls for obtaining regular up-to-date images or measurements of the “true” eligible area to be compared with the reference eligible area within the LPIS.

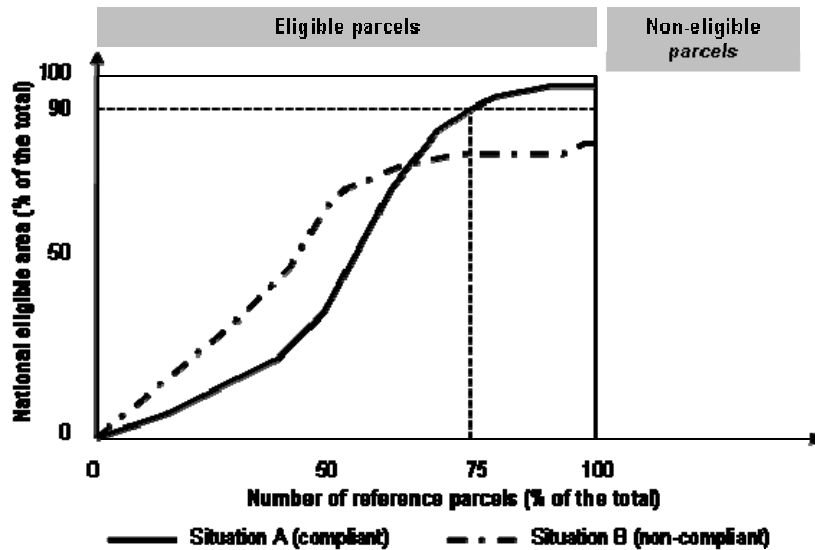


Figure 1: “75/90%” LPIS quality rule

Because the discrepancy between the “reference” area (registered as eligible within the LPIS) and the “true” area (as measured from satellite imagery or in the field with a GPS device) for exactly 75% of the reference parcels at national level does not exceed 10% of the total national eligible area, situation A is considered compliant with the “75/90%” LPIS quality rule; by contrast, situation B is non-compliant.

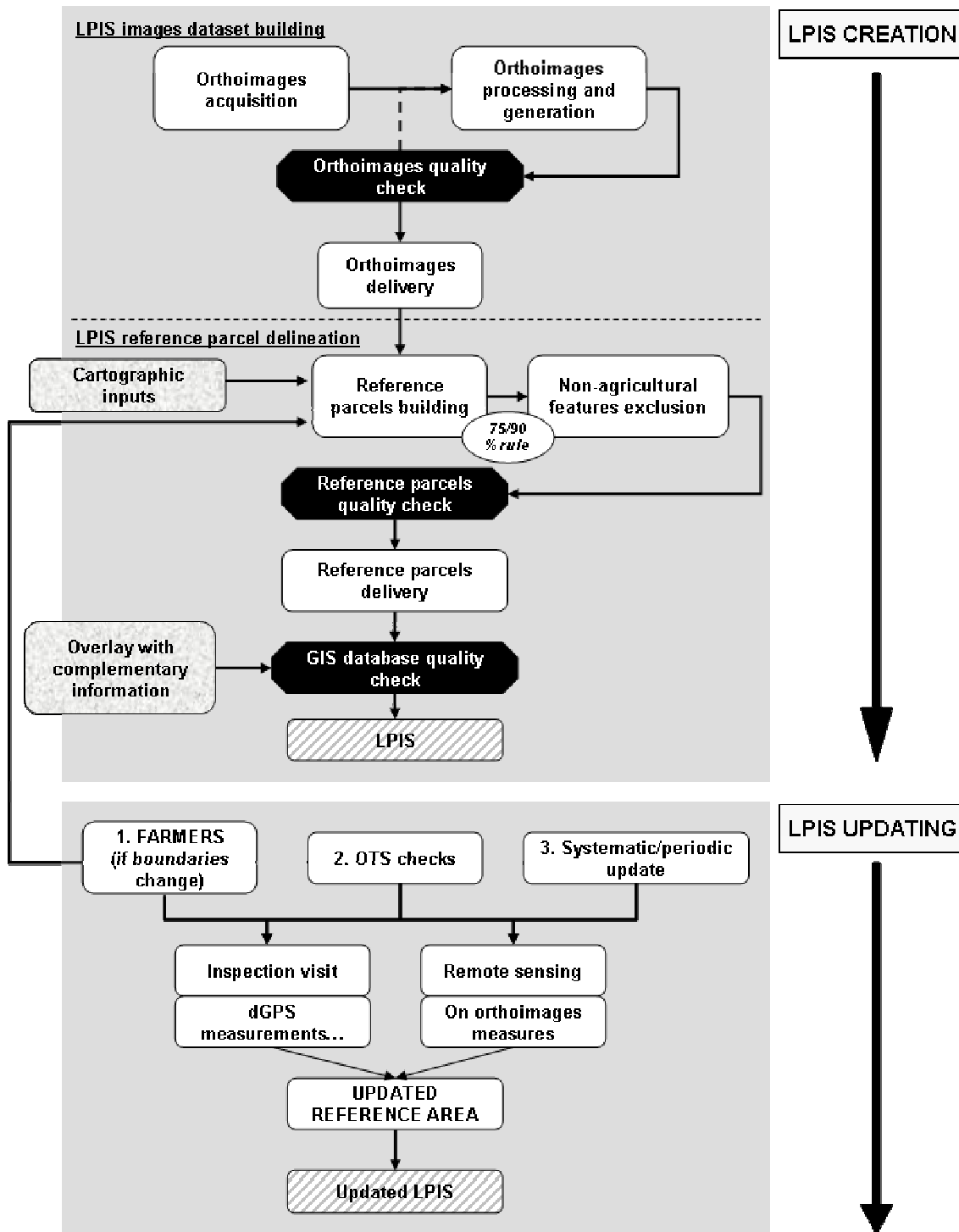
At parcel level, the “75/90%” rule could be interpreted as the need for each reference parcel within the LPIS to present a ratio between the true and the reference values higher than 0.9. In fact, this rule applies to the total national eligible area and non-conform small parcels (discrepancy in the area > 10% of the reference parcel area) can easily be compensated by conform larger reference parcels. This leaves national administrations a degree of flexibility to comply with this requirement.

Considering this, LPIS quality can be assessed by measuring the number of anomalous parcels presenting a discrepancy in area higher than 10% and the discrepancy in area of the anomalous parcels population. These two indicators of the overall LPIS quality could be useful for identification of the critical points in (i) LPIS creation (initial registration of reference parcels in the LPIS) and/or (ii) LPIS updating (from up-to-date reference area measurements) to decide on possible improvements and achieve global correctness.

### **3. Creation and diversity of the LPIS**

Although the processes for LPIS creation and updating can vary between Member States, there are some common features.

First, almost all the Member States rely on more or less up-to-date orthophotos to create reference parcels. However, reference parcels can also be created from other datasets, such as cadastral maps, land distribution plans, topomaps, etc. In both cases, the reference parcels are then checked and validated against archives or new orthophotos (see Figure 2). Second, although LPIS reference parcels are usually created at national level, some Member States leave the regions the possibility of establishing LPIS in accordance with local preferences.







**Figure 2: Workflow of LPIS creation and updating processes (general description of the processes; this figure does not claim to match exactly all the situations existing in Europe)**

Finally, the setting-up of the LPIS could be closely related to the available information technologies and investment decisions on the date of the country's accession to the EU or simply to historical preferences. Airborne imagery is widely used in Western Europe, whereas Eastern Europe tends to use very high resolution (VHR) satellite imagery. In addition, the historical

relationship between farmers and land occupation or land ownership has led to slightly different procedures for LPIS creation, resulting in a range of different types of LPIS in the EU.

Today, four major types of LPIS co-exist in Europe (see Figure 3 for descriptions and characteristics). The most commonly used LPIS in the EU is (i) the “physical block (PB)” because it is stable over time and, consequently, easy to update. Member States also use (ii) the “agricultural parcel (AP)” or (iii) “farmer block (FB)” as reference parcel systems. These are considered appropriate choices to facilitate administrative CAP checks, but FB and AP updating is considered more complex and time-consuming. Finally, Member States which base their LPIS on the land register (parcel ownership) use “cadastral parcel (CP)” as the reference (Milenov and Kay, 2006).

				
<b>LPIS types</b>	<b>Agricultural parcel (AP)</b>	<b>Farmer block (FB)</b>	<b>Physical block (PB)</b>	<b>Cadastral parcel (CP)</b>
<b>Content/coverage</b>	Single crop group	One or more crop groups	One or more crop groups	Does not match agricultural patterns
<b>Applicant</b>	Single farmer	Single farmer	One or more farmers	One or more farmers
<b>Timeframe</b>	Annual	Multi-annual	Semi-permanent	Not applicable

**Figure 3: Description of the four types of LPIS existing in Europe (Sagris *et al.*, 2007)**

#### **4. LPIS updating**

To ensure correct distribution of annual agricultural subsidies to farmers, it is necessary to maintain LPIS quality over time. This is achieved by an updating process to demonstrate that the LPIS complies with the regulatory requirements and its capacity to integrate the latest changes to farmers’ aids applications. Figure 2 illustrates the three main regulatory paths (Commission Regulation (EC) No 796/2004) currently proposed for integrating land-use changes and updating the LPIS:

- (1) Farmers could correct pre-printed forms by indicating the new boundaries of each reference parcel and, when necessary, by updating their area. The new reference parcel boundaries could be then automatically checked by the GIS to pinpoint and correct overlaps (de Laroche, 2007). This method could allow annual updating of up to 100% of the reference parcels and is generally well accepted by farmers.
- (2) National payment agencies conduct annual on-the-spot (OTS) checks to legitimise individual farmers’ aid applications, based on the reference parcels in the LPIS. During these checks, field inspectors have recourse to validated remote-sensing procedures and/or certified dGPS devices to record the up-to-date situation of the reference parcel areas. Unfortunately, although OTS checks provide very precise information and correction of irregularities,

complete updating of the LPIS every year by OTS checks is unfeasible for economic reasons. Instead, a minimum of at least 5% of farmers' applications are expected to be checked every year (for instance, in the 2007 campaign, 590 000 OTS checks in the EU-27, corresponding to an average of 6.7% of the reference parcels).

- (3) An overhaul driven by monitoring anomalies in the LPIS is required; accordingly, national CAP payment agencies ask their contractor(s) responsible for image acquisition and processing to ensure a complete renewal of orthophoto coverage over a period of five years and to update the reference parcels delineation.

Today, one or more of these possibilities has been introduced in each Member State actively participating in the ongoing LPIS updating.

However, because of the day-to-day changes in the landscape (undeclared anthropogenic changes such as new buildings, construction of facilities, deforestation, etc.), misdeclaration (over- or even under-declaration of reference parcel areas) or obsolete orthophotos, an exact match between the reference parcels area registered in the LPIS and the up-to-date reference parcel area is utopian. Several years after introduction of the IACS, the declared area still often does not reflect the actual situation (Oesterle and Hahn, 2004). Moreover, a complete and regular check on all the reference parcels by national authorities is impossible. Consequently, irregularities persist within the LPIS. For all Member States, in 2004 an average of 40% (28.9% in 2006) of the applications checked on the spot contained over-declarations of the area claimed (Court of Auditors Annual Report, 2005); this was equivalent to 2.1% of the reference area verified by the European payment agencies.

## **5. LPIS irregularities**

According to Council Regulation (EC) No 796/2004, "irregularities" means "*any non-respect of the relevant rules for the granting of the aid in question*". Since the 2003 CAP reform, aid has become area-based. Consequently, LPIS irregularities usually take the form of differences between the area declared in the application (by the farmer or others) and the reference area identified as eligible in the LPIS. Most irregularities are over-declarations. However, the discrepancy is often below the 5% buffer tolerance accepted and, consequently, does not lead to any reduction in payments. The buffer tolerance is currently set at a maximum of 1.5 m applied to the perimeter of the reference parcel and the area discrepancy should not exceed 1.0 ha (Council Regulation (EC) No 972/2007). This buffer tolerance is partly explained by the accuracy of today's imagery and of the dGPS used (Pluto-Kossakowska *et al.*, 2007).

No survey has yet been undertaken of the irregularities found in the European LPIS. Consequently, a preliminary description of the nature and frequency of irregularities in the LPIS would greatly help administrations with general application of the LPIS for sustainable agri-environmental management and would provide meaningful information for cost-effective updating of the LPIS.

## ***PART B. European LPIS appraisal: ecological approach and subsequent objectives***

Strictly speaking, LPIS irregularities are defined as a significant discrepancy between the area stated in the farmer's annual declaration and the reference area in the LPIS, calling for withdrawal of undue payments. This implies control procedures, measurement and monetary valuation of the discrepancy in area triggering a reduction/exclusion/correction process. Because neither individual farmers' subsidies nor economic valuations of discrepancy in area are widely available public information, this survey did not focus on these aspects.

Moreover, because no area discrepancy measurements were planned during the survey, it was decided not to use the regulatory term "irregularity". Instead, "anomaly" was substituted for "irregularity". Anomaly is a more general term covering reference area discrepancies observed (but not measured). In accordance with the accuracy of the images used during the survey, "anomaly" corresponded to area discrepancies equal to or higher than 0.1 ha.parcel<sup>-1</sup>, whatever the cause.

### **6. Ecological approach**

During the survey, LPIS anomalies were detected and analysed by computer-assisted photo-interpretation (CAPI). "LPIS anomaly" meant nothing other than the existence of an inconsistency between the area registered in the LPIS and the area observed on the up-to-date orthorectified image of the real situation.

Whatever the size of the discrepancy observed, the anomaly was identified by the nature of the cause of the discrepancy; object in cause was then considered as sub-area to be excluded from the reference parcel. Area discrepancies could have many causes and could be differentiated by the nature of the object partly or totally embedded in the reference parcel. Anomalies could then be conceived as spatially (i) clumped (aggregative) in the case of non-eligible marshlands surrounding several contiguous reference parcels situated along a river, (ii) systematically distributed across an area in the case of a shift in parcel boundaries due to a systematic orthorectification error during processing of images or (iii) randomly distributed in the case of new buildings constructed by farmers who have decided to invest since the last LPIS update. In this study the spatial distribution of anomalies was perceived as similar to the spatial distribution of many biological organisms, with individual anomalies and classes of anomalies treated like species and taxa respectively. The ecological approach was consequently considered a suitable method for analysing diversity of LPIS anomalies.

Over the last 80 years, numerous ecologists addressing biological diversity have needed to answer the central question "*How many are there?*", meaning the biological diversity in one specific area or the diversity between different areas. Local measurements of the number of species, i.e. species richness, by a complete survey of local diversity ( $\alpha$ -diversity<sup>1</sup>) is generally out of the question due to limited human and financial resources. To overcome this limitation, extrapolation from samples has proven relevant (Sanders, 1968; Heltshe and Forrester, 1983; Colwell and Coddington, 1994). Unfortunately, as underlined by Colwell *et al.* (2004), species richness is

---

<sup>1</sup> Alpha diversity ( $\alpha$ -diversity) is the biodiversity within a particular area, community or ecosystem, generally habitat, and is measured by counting the number of taxa (distinct groups of organisms) within the habitat (e.g. families, genera or species).



notoriously dependent on sample size due to sampling effects and depends intrinsically on sample size when data from habitats are successively pooled, due to species turnover ( $\beta$ -diversity<sup>2</sup>). In most cases, an optimum sample size exists that allows detection of the maximum species richness. Once this is reached, any additional sampling will identify no extra species. Consequently, an approximate inventory of species is produced mainly by (i) measuring or estimating the species richness of each local species assemblage ( $\alpha$ -diversity) and (ii) measuring or estimating the complementarity (i.e. distinctness) of different local inventories ( $\beta$ -diversity) in order finally to extrapolate it to the overall diversity on global scale ( $\gamma$ -diversity) (Whittaker, 1972).

Combined with the development and everyday application of spatial devices such as GIS to manipulate and analyse orthophoto and associated data, the ecological approach was seen as a meaningful solution to analyse the diversity of LPIS anomalies. It was adapted to a large range of scales, from local (a group of reference parcels in a given zone) to distinct sites on different continents (between Member States or in the whole EU). With this in mind, the general approach taken was to combine the well accepted *ecological methodology* and *geographic information system* in order to propose a complete, replicable method to assess LPIS quality.

## **7. Objectives**

The goal of this study was to assess the consistency and robustness of LPIS for actual and future users or decision-makers responsible for the CAP by surveying **the diversity and the spatial pattern of LPIS anomalies**.

### **7.1. Typology of LPIS anomalies**

Each LPIS anomaly corresponds to a single reference area discrepancy and a single cause. A typology of LPIS anomalies reflecting all possible causes of the observed discrepancies was compiled *ex ante*.

First, LPIS anomalies could be the consequence of inaccurate orthophotos used when creating the LPIS. They could also stem from weaknesses in image processing (e.g. misuses of orthorectification) or be due to computer-assisted photo-interpreters' inexperience. Second, during LPIS updating, reference area boundaries can be corrupted, whether intentionally or not. Third, if the shape of the reference parcel is complex or if irregular landscape features surround it, correct delineation of its area can be difficult. Finally, addition of anthropogenic landscape elements, such as buildings, roads, electricity facilities, etc., or changes of land use continuously change reference parcels and require up-to-date delineation of the area.

Accordingly, an *a priori* classification of possible LPIS anomalies was proposed. This classification was used for this survey. In parallel, it was also used to address automatic detection in the LPIS of 2/3D objects considered elements to be excluded from the LPIS reference parcels (Zielinski, 2009). All the possible LPIS anomalies were organised into four distinct classes based on their plausible cause (see Appendix A):

- Tree-related anomalies (TRA<sub>i</sub>): this class covers reference areas where a high density of trees was erroneously declared as an eligible part of the parcels. Four different types of anomaly make up this class in which 3D objects are considered.

---

<sup>2</sup> Beta diversity ( $\beta$  diversity) is a measure of biodiversity which works by comparing the species diversity between ecosystems or along environmental gradients. It is the rate of change in species composition across habitats or among communities.

- Man-made related anomalies (MRA<sub>i</sub>): this class covers reference areas where new man-made objects were erroneously declared as an eligible part of the parcels. Five different types of anomaly make up this class in which 3D objects are considered.
- Water-related anomalies (WRA<sub>i</sub>): this class covers reference areas where water bodies, such as rivers, ditches or ponds, were erroneously counted as part of the reference area. Four different types of anomaly make up this class in which 2D objects are considered.
- Finally, boundary-related anomalies (BRA<sub>i</sub>): this class covers anomalies explained by incorrect positioning (i.e. shifting) of the reference boundaries in the LPIS. It also includes overlapping of boundaries in the LPIS from two different contiguous reference areas. A final category of anomaly is uses of the reference parcel other than for agriculture; these were considered as BRA<sub>i</sub> anomalies because of the need to exclude them or to redefine the true reference area boundaries. Four different types of anomaly make up this class in which 2D objects are considered.

→ ***The first objective was then to assess the validity of the typology of LPIS anomalies against the census of anomalies within the LPIS.***

#### 7.2. Validation of a sampling method for the survey of LPIS anomalies

Diversity analysis initially depends on obtaining significant estimators of diversity. This means obtaining samples representative of the real situation. Consequently, an appropriate sampling method for gaining an approximate picture of the diversity in LPIS anomalies was required.

Neither the technical reports provided from national authorities in charge of the LPIS nor the scientific literature have explicitly described the diversity and spatial pattern of LPIS anomalies. It was also difficult to find examples of sampling methods adapted to detection and description of LPIS anomalies from Member States conducting yearly reference parcel sampling from orthophotos. As it was necessary to propose a complete methodology to sample anomalies assemblages within the LPIS, it was important to validate use of the method for the range of types of LPIS existing and the multiple causes of LPIS anomalies all along the LPIS creation and updating chain (see Figure 2).

→ ***The second objective was to propose and calibrate an easy-to-understand, cost-effective and replicable method to sample anomalies within the LPIS accurately in order to compare LPIS anomalies assemblages.***

### 7.3. Description of LPIS anomalies assemblages

In line with the typology proposed above, diversity in LPIS anomalies can be addressed by considering anomaly diversity in a single LPIS type ( $\alpha$ -diversity), between LPIS types ( $\beta$ -diversity or spatial pattern) and at EU level, considering all LPIS types ( $\gamma$ -diversity) (Whittaker, 1972). The different levels of diversity obtained from the relative occurrence (i.e. the frequency of anomalous parcels) and/or relative abundance (i.e. density of each type of anomaly) of each of the 17 anomalies listed (see Appendix A) can (i) provide a reliable identification of common and rare anomalies regarding each LPIS type and (ii) allow description and comparison of the LPIS anomalies assemblages obtained.

→ *The third objective was to produce a description of the  $\alpha$ - and  $\beta$ -diversity of the relevant LPIS anomalies.*

### 7.4. Relationship between LPIS anomalies assemblages and landscape structure and LPIS management

From the typology of anomalies proposed, LPIS anomalies assemblages could be related to the conditions for LPIS creation and updating and/or to the landscape organisation. Certain candidate variables describing the workflow for LPIS creation/updating or landscape should be of interest to depict the anomalies assemblages. Identification of such relationships could provide potential indicators of LPIS quality; it could also provide suitable information to map “at-risk” LPIS situations and to help Member States to decide the zones to be included in LPIS risk analyses each year.

→ *The last objective was to discuss possible relationships between LPIS anomalies assemblages and (1) LPIS creation/updating conditions and (2) landscape structure descriptors.*

The workflow followed during the LPIS survey is illustrated in Figure 4. This shows the successive steps taken during the survey and can be perceived as a guideline for everyone who would like to conduct or improve the general method in particular situations.

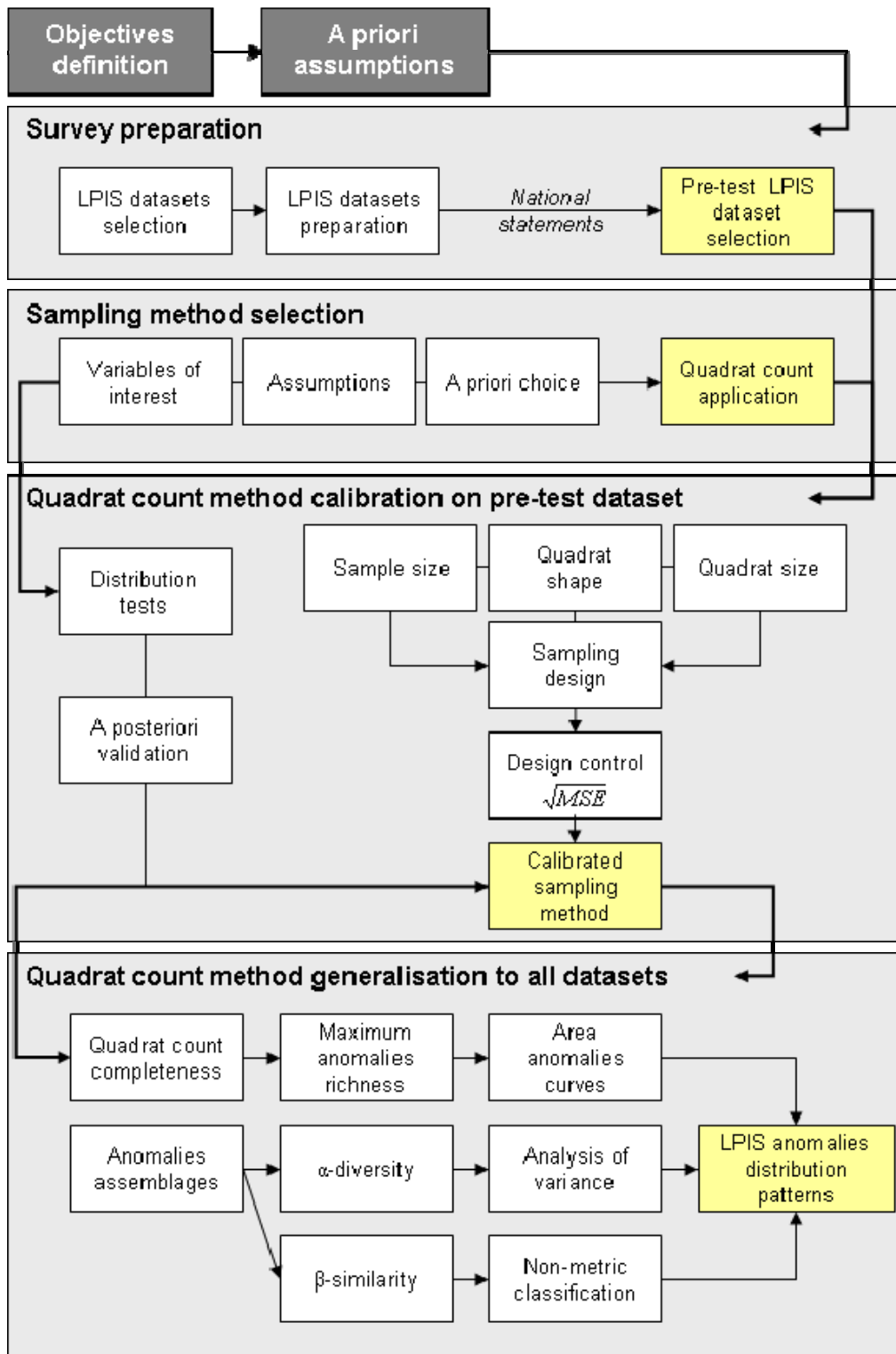
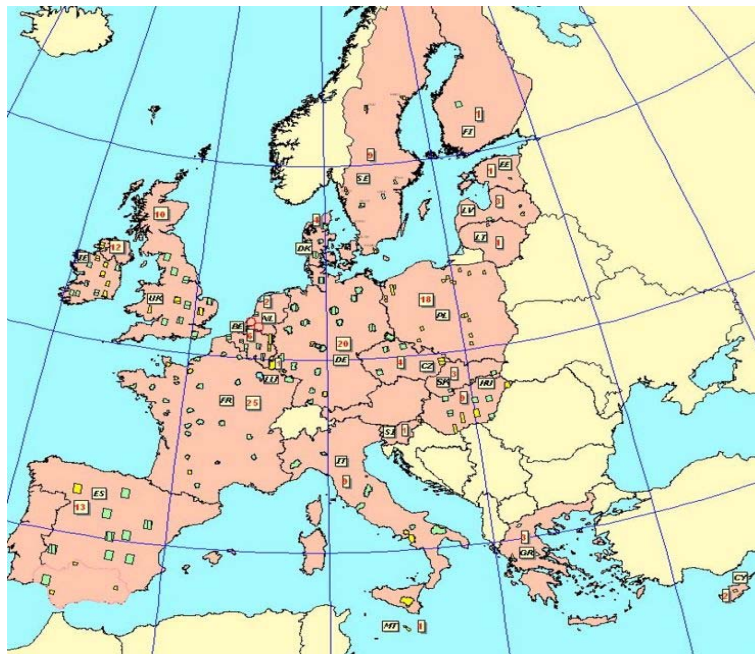


Figure 4: General workflow for the 2005 LPIS quality survey

## ***PART C. Data used and preparation***

### **8. LPIS dataset selection and preparation**

Annually, the satellite imagery for the Control with Remote Sensing (CwRS) to check farmers' declarations is provided by the European Commission's Joint Research Centre (JRC) and delivered to each Member State participating in the project. The satellite data acquired are geometrically and radiometrically corrected so that they can be used in the computer-assisted photo-interpretation phase for parcel determination and irregularity diagnosis. Due to budgetary and technical limitations, checks are conducted on only a limited number of zones in each Member State (see Figure 5). These zones are selected by the national administrations based on pre-established rules, including random sample and risk analysis criteria. They are then distributed to each contractor responsible for the CwRS in a Member State.



**Figure 5: Distribution of the control zones for the CwRS 2005 project**

Once the CwRS is completed, the data from the zones are then sent back to the JRC following a clearly defined procedure, so that it can assess the quality and efficiency of the CwRS performed by the Member States. The data consist of the basic details among them LPIS normally used by the contractor to carry out the ordinary check (except for the “dossiers”, i.e. the irregularities detected and resultant withdrawals of subsidies). Consequently, an extensive database of year-to-year images and orthophotos from all the Member States is available at JRC level.

From these datasets, in line with the objectives of this study, it was decided to concentrate on a single year – 2005 – to obtain preliminary knowledge about the diversity of LPIS anomalies before planning broader or longer-term studies. In 2005 the CwRS project covered 161 zones (see Figure 5) in 23 Member States and 27 national contractors were employed to perform LPIS quality assessment. This year was chosen because relatively complete LPIS datasets (imagery and

reference LPIS data) were available for 21 Member States. For two other MS the national datasets (image or vector data) sent to the JRC were found to be incomplete (e.g. parts of the LPIS or landscape exclusion layers were missing). From these 21 available datasets, the zones were classified in accordance with the four existing European LPIS types (Milenov and Kay, 2006). Next, the completeness and usability (computer incompatibility could limit use of some complete datasets) were verified for each LPIS type, the goal being to select randomly three replicates (three different Member States, with one CwRS zone each) from each LPIS type. In 2005, three replicates were available for three of the four different LPIS types (AP – agricultural parcel, CP – cadastral parcel and PB – physical block). For FB (farmer block), due to lack of availability, two replicates (zones) from the same Member State had to be included to obtain the same statistical validity. This was possible because the zones in this particular Member State were highly distant and managed by two different CwRS contractors. First, it was assumed that, despite the common European technical specifications they had received, their internal procedures, skills and management could vary widely and lead to different LPIS anomalies assemblages; second, high distance between zones allowed for sampling very different landscapes.

The zones selected as replicates for the survey are listed in Table 1.

**Table1: Replicates and corresponding acronyms of the four LPIS types surveyed**  
(Member States’ denominations are not indicated to respect the initial anonymity terms; however, they could be communicated to the Member States concerned by this survey individually on request)

LPIS types	Agricultural parcels	Cadastral parcels	Farmer blocks	Physical blocks
Acronyms	AP	CP	FB	PB
Replicate 1	AP1	CP1	FB1	PB1
Replicate 2	AP2	CP2	FB2	PB2
Replicate 3	AP3	CP3	FB3	PB3

Since the goal of this survey was to address LPIS quality for the range of LPIS types existing in the EU and not to compare LPIS quality between Member States, the names of the Member States from which zones were sampled were not indicated. However, on request, the national authorities concerned could obtain results concerning their territory.

The selected zones and corresponding images were subjected to several procedures to ensure that each test zone would be elaborated in the same manner, e.g. by a unified file structure and processing. First, the entire dataset for each zone was individually established in the GIS environment. Second, several GIS vector operations were used to select from the dataset delivered only parcels which could be checked against the image content (exclusion of incomplete parcels on the edge of the image). In addition, vector layers (anomalies) with defined attributes (cause of anomalies) were assigned to each zone.

One of the twelve zones selected (FB2) was chosen for adaptation (calibration) of the sampling method. Zone FB2 was chosen on the basis of several considerations:

- Zone FB2 is in a Member State where a very limited number of irregularities are generally reported, implying that the quality of the LPIS was *a priori* high.
- Because it was assumed that landscape characteristics and reference parcel area could have an effect on the composition of LPIS anomalies, a zone was selected for which (i) the mean reference parcel area was close to the EU average and where (ii) the landscape structure (and farming system) were representative of the European situation. In this way, this sampling method adapted to the difficult circumstances of zone FB2 (low number of anomalous

parcels) and its medium landscape structure was expected to prove suitable for any other landscape and/or more anomalous situations.

## ***PART D. Sampling method selection and calibration***

This part explains the successive steps to decide and adapt the ecological sampling method to the specific case of LPIS quality assessment. Adaptation and calibration of the sampling method were performed on zone FB2, which had first been selected as the pre-test zone (see above).

### **9. Ex-ante assumptions and sampling method selection**

#### **9.1. Initial assumptions**

Regarding the typology of anomalies proposed, LPIS anomalies could stem from landscape features intersecting the reference parcel or from GIS misuse during LPIS creation/updating. Then absolute<sup>3</sup> abundance of LPIS anomalies could be partly correlated with the structure and level of complexity of the landscape in a zone. Landscape complexity is defined as the fragmentation of the habitat (Fahrig, 2003) into different classes of land use, spatially distributed in accordance with driving forces. The five principal types of factor determining the land-use pattern are political, socio-economic, natural, cultural and technological (Brandt *et al.*, 1999). Recently, Sklenicka and Salek (2007) showed that land ownership patterns, expressed as mean parcel size and the proportions of private/public land ownership, have a positive impact on landscape fragmentation and reciprocally. This suggests that the probability of TRA<sub>i</sub> or MRA<sub>i</sub> (see Appendix A) landscape-related anomalies could increase in proportion to reference parcel area: larger reference parcels should have more chance of intersecting with tree-related or man-related landscape features. On the other hand, in zones of high natural value, the ban on deforestation could be expected to limit groupings of contiguous reference parcels and favour small reference parcels. In this case, the presence of a highly complex network of ditches and hedges around the parcels could favour a high proportion of tree- or water-related anomalies as the predominant natural driving force.

It was therefore assumed that:

- Land-use classes present in a zone, or next to a reference parcel, could explain the anomalies detected.
- The magnitude of each anomaly detected could be correlated with the level of fragmentation of landscape around the reference parcels.
- Finally, reference parcel area could also have an impact on both the occurrence and abundance of anomalies, because of the higher probability of being surrounded by more numerous and complex landscape features, i.e. the “parcel area” effect.

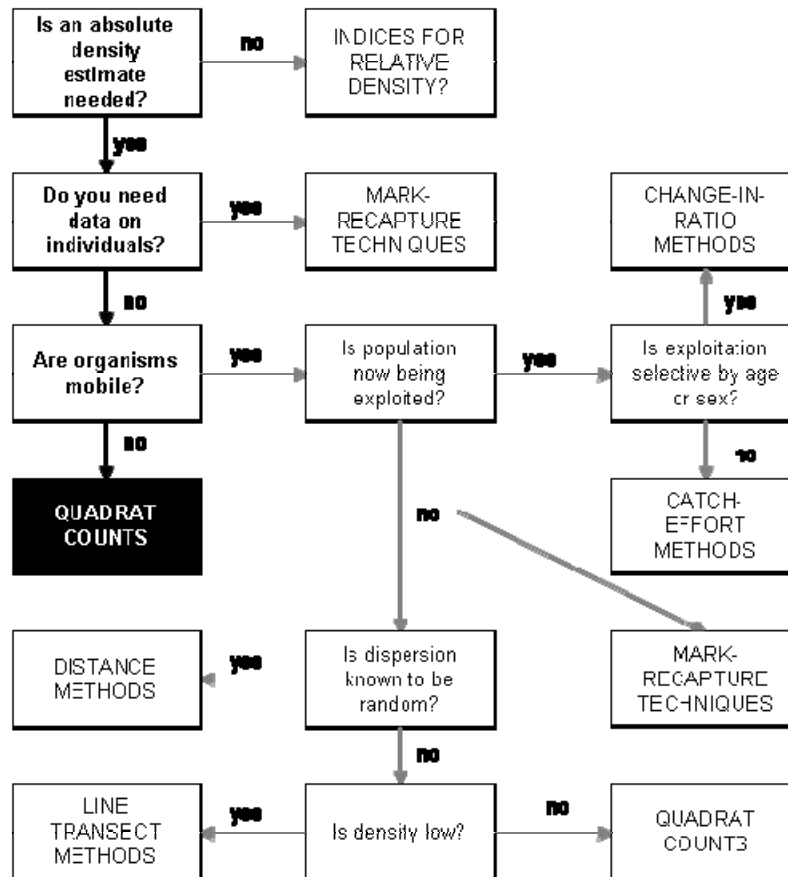
Because both the occurrence and abundance of anomalies could be consequences of landscape fragmentation and/or of reference parcel area, direct sampling of reference parcels as the sampling unit was out of the question. Sampling of a set range of reference parcel areas would provide a biased estimate of anomalies occurrence and abundance because of the assumed “parcel area” effect. To free the sample of this potential bias, it was decided to use a method independent of parcel area to sample anomalies in the LPIS. According to Krebs (1999) (see Figure 6), the facts that (i) absolute abundance is expected, (ii) no supplementary information, such as area

---

<sup>3</sup> Absolute occurrence or abundance means the real number of individuals of a species of interest, mainly expressed per unit area (i.e. absolute density); by contrast, relative occurrence or abundance means the number of individuals of species A per unit area relative to another species B.



discrepancy, is measured and (iii) LPIS anomalies are immobile on orthophotos, at least during the current campaign, led to opting for quadrat counts.



**Figure 6: Decision-making tree for choosing a method for estimating absolute abundance (Krebs, 1999)**

Quadrat counts are widely used for evaluating the diversity of immobile organisms such as plants (Moora *et al.*, 2007; Singh *et al.*, 2007). The main idea is to sample a unit area, i.e. quadrat, instead of sampling individuals. The species richness is then expressed as the number of individual specimens per unit area (quadrat) and, consequently, corresponds to a density. There are two basic requirements for applying this technique: (i) that the sampling unit area is known and (ii) that the individual specimens are immobile during the counting period.

During the LPIS survey, LPIS anomalies were the individuals to be detected in the reference parcels in the quadrats sampled. Reference parcels were considered as belonging to the quadrat if (1) the highest proportion of their area fell within the quadrat (in cases where more than two quadrats intersected the reference parcel) or (2) more than 50% of the parcel area fell within the quadrat (in cases where a parcel intersected a maximum of two quadrats).

Once the sampling method had been decided, the focus turned to the different steps to adapt it to the LPIS survey. This adaptation stage was all the more important since no references to this topic were available in the literature.

## 9.2. Variables of interest

To allow analysis of the diversity of LPIS anomalies, two variables of interest were calculated at quadrat level:

- Anomaly occurrence was expressed as the percentage of anomalous reference parcels within the  $j^{th}$  quadrat:

$$O_j(\%) = \frac{\sum_{i=1}^i O_i}{n_j} * 100 \quad (\text{equation 1})$$

where  $O_j(\%)$  = the anomaly occurrence in the  $j^{th}$  quadrat;

$O_i$  = the anomaly occurrence in the  $i^{th}$  reference parcel in the  $j^{th}$  quadrat; this was assigned the value 0 in the absence and 1 in the presence of at least one anomaly in the  $i^{th}$  reference parcel, whatever the cause of the anomaly detected;

$n_j$  = the number of reference parcels in the  $j^{th}$  quadrat.

- Anomaly abundance was determined for each anomaly type and for each parcel in a certain quadrat; it was expressed as the total number of anomalies by quadrat and/or as the total number of one of the seventeen possible anomaly types listed (see Appendix A) by quadrat. Because it corresponded to a number of individuals per unit area (quadrat), abundance was also a density of anomaly. The formula is as follows:

$$A_{j,l} = \sum_{i=1}^i A_{i,l} \quad (\text{equation 2})$$

where  $A_{j,l}$  = the total abundance of anomaly type  $l$  in the  $j^{th}$  quadrat;

$A_{i,l}$  = the abundance of anomaly type  $l$  counted in the  $i^{th}$  reference parcel of the  $j^{th}$  quadrat.

## **10. Sample size**

Sample size, i.e. the minimum recommended number of quadrats to consider when assessing the diversity of LPIS anomalies, is determined *a priori* from a complete census of all the anomalies. It consequently required measurement of the “true” occurrence and abundance of anomalies. Within an orthophoto, the number of reference parcels is high, as could be the number of anomalous parcels; because of the limited time and human resources, a complete census of all twelve zones selected was impossible. The sample size was determined from the pre-test zone (FB2) for which a complete census of anomalies was conducted for the whole orthophoto. The total anomalies assemblage was considered the “true” anomalies assemblage within zone FB2 and was called “POP”.

Within POP, 11 of the 17 types of anomaly proposed were detected. The table set out below shows the absolute abundance detected for each of the 11 anomalies from the initial LPIS anomalies typology. The six remaining anomalies were not detected in zone FB2.

Anomaly category	TRA1	TRA2	TRA3	MRA1	MRA2	MRA3	MRA4	MRA5	WRA1	WRA2	BRA3
Abundance	328	8	7	121	20	58	43	56	15	16	13

The true number of anomalous parcels was 572 out of the 9 936 reference parcels in zone FB2. Hence, the anomaly occurrence was 5.76%. This means that 5.76% ( $p$ ) of the reference parcels in zone FB2 displayed at least one anomaly. The remaining 94.36% ( $1-p$ ) were not anomalous. From this, the sample size needed to estimate a population to a specified margin of error within a specified level of confidence was calculated. This corresponded to probabilistic determination of sample size when sampling an infinite population where the expected sample will be less than 10% of the population (Remenyi *et al.*, 2007). To that end, it was assumed that anomaly occurrence, as a discrete variable, was described statistically by the binomial distribution with only two parameters:

$p$  = the proportion of anomalous parcels with at least one anomaly in zone FB2, and

$q = 1 - p$ , the proportion of non-anomalous parcels in zone FB2.

The sample size was then determined by considering an acceptable margin of error ( $d$ ) in the estimate of  $p$  and the probability ( $\alpha$ ) of not achieving this margin of error. This led to use of the normal approximation to the confidence interval given by the formula:

$$\hat{p} \pm t_{\alpha} s_{\hat{p}} \quad (\text{equation 3})$$

where  $\hat{p}$  = observed proportion of anomalous reference parcels (0.0576),

$t_{\alpha}$  = the value of the Student's  $t$ -distribution for  $n-1$  degrees of freedom and

$s_{\hat{p}}$  = the standard error of  $\hat{p}$ .

The desired margin of error is then:

$$d = t_{\alpha} s_{\hat{p}} = t_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (\text{equation 4})$$

Solving for  $n$ , the sample size required for an infinite population is:

$$n = \frac{t_{\alpha}^2 \hat{p}\hat{q}}{d^2} \quad (\text{equation 5})$$

Finally, if the sample size equals more than 5% of the initial population, the procedure is to calculate the sample size from equation 3 above and then to correct it with the following *finite population correction*:

$$n' \cong \frac{n}{1 + (n/N)} \quad (\text{equation 6})$$

where  $n'$  = estimated sample size required for finite population  $N$ ,

$n$  = estimated sample size required for an infinite population,

$N$  = total size of the finite population.

For the specific zone FB2 where  $\hat{p} = 0.0576$  ( $\hat{q} = 0.9424$ ),  $p$  was estimated within an error limit of  $\pm 0.02$  (2%) with  $\alpha = 0.1$  ( $t_{\alpha} = 1.645$ ). From equation 3:

$$n_{FB_2} = \frac{(1.645)^2 (0.0576)(0.9424)}{(0.02)^2} = 367$$

Because the 367 reference parcels to be sampled ( $n$ ) equalled less than 5% of the total number of reference parcels ( $N$ ) in zone FB2 ( $R_{FB2} = 367/9\ 936 = 3.68\%$ ), no finite population correction was applied. From the mean reference parcel area in zone FB2 ( $5.04 \pm 9.13SD$  ha.parcel<sup>-1</sup>), the minimum orthophoto area to sample was 18.5 km<sup>2</sup>.

From this, it was assumed that anomalous parcel occurrence in the eleven other zones selected was close to the occurrence measured for zone FB2. Consequently, the minimum number of reference parcels to be sampled obtained from zone FB2 was applied to all eleven other zones by multiplying  $n_{FB2}$  by the mean reference parcel area for each of the eleven zones. Finite population correction was applied solely to the zones for which  $n/N(\%) \geq 5\%$ . The minimum area to be sampled by zone is shown in Table 2.

**Table 2: *A priori* estimate of the minimum sample size for each LPIS replicate in the survey**

LPIS types	Replicates	N (equation 6)	Total area (ha)	Mean area (ha)	N (equation 6)	$R_{(n/N)}$ (%)	n'	km <sup>2</sup>
AP	AP1	18 307	87 872.96	4.80	367	2.00	367	17.62
	AP2	5 647	6 621.64	1.17	367	6.50	344	4.04
	AP3	6 507	24 249.38	3.73	367	5.64	347	12.95
CP	CP1	172 540	96 275.56	0.56	367	0.21	367	2.05
	CP2	44 784	168 494.04	3.91	367	0.82	367	14.35
	CP3	69 170	28 635.84	0.41	367	0.53	367	1.52
FB	FB1	5 365	35 714.45	6.66	367	6.84	343	22.87
	FB2	9 936	50 074.62	5.04	367	3.69	367	18.50
	FB3	3 302	16 910.00	5.12	367	11.11	329	16.92
PB	PB1	3 068	13 535.63	4.41	367	11.96	327	14.47
	PB2	39 118	71 190.80	1.82	367	0.94	367	6.68
	PB3	2 144	82 632.59	38.54	367	17.12	313	120.82

The minimum area to be sampled on the orthophoto was between 2.05 km<sup>2</sup> (AP1) and 120.82 km<sup>2</sup> (PB3). To obtain comparable results on the distribution of LPIS anomalies between zones, the sample size should be the same for every zone and should correspond to the maximum value obtained, in this case PB3  $\approx$  121 km<sup>2</sup>. However, because of limited time and human resources, it was inconceivable to apply a sampling size of 121 km<sup>2</sup> to all the zones. Instead, it was decided to apply the second highest minimum sample size (FB1: 22.87 km<sup>2</sup>) and to double it bearing in mind the total of four weeks available for the GIS-assisted anomalies census. A sample size of **45 km<sup>2</sup>** was therefore applied to each of the LPIS zones. This sample size, at a constant anomaly occurrence of between 4% and 7%, should at least make it possible to obtain accurate (to 2%) estimates of anomalies assemblages for the first eleven zones and an approximate idea of the anomalies assemblage in zone PB3. This will be verified in Section 16.

### **11. Quadrat shape**

In practice, “quadrat” means any sampling unit, whether circular, hexagonal or even irregular in outline. However, the *shape* of the quadrat introduces a degree of bias due to the “edge effect” (= ratio between the length of the edge and the area inside). An individual anomaly on the edge of a quadrat can be considered either inside or outside the quadrat and, consequently, counted or not. If it is counted, this leads to possible overestimation of the density. The higher the edge effect of the quadrat, the greater the potential bias. To reduce the bias related to the edge effect, a circular

shape could be chosen as it displays the lowest edge effect. At the same time, however, use of circular quadrats cannot cover the whole area of interest without overlaps. For this LPIS survey a **square quadrat was chosen**, as this shape seemed the best compromise between a rectangular quadrat (with a high edge effect but complete coverage) and a circular quadrat (with the lowest edge effect but incomplete coverage).

## **12. Quadrat size and sampling design**

It is never easy to determine the quadrat size to apply. The simplest approach is to refer to the literature and use the same size as everyone else (Krebs, 1999). Unfortunately, no references to sampling of reference parcels from orthophotos were available at the time of the survey. The second solution is to test the optimum quadrat size for the particular study. “Optimum” means statistically and/or logistically optimal. Because the experiment was GIS-assisted, there was no logistical difficulty in setting and using any kind of quadrat size. Consequently, “optimum” meant “statistically” only, i.e. a quadrat size giving the highest statistical precision (i.e. the lowest standard deviation) for a given amount of available time. However, detection of anomalous reference parcels within the LPIS from orthophoto surveys can be seen as the first step in LPIS quality assessment. Following the GIS-assisted census of area discrepancies, direct measurements of area discrepancies in the field could be planned to evaluate technical or financial sanctions. In that case, optimum quadrat size could also mean statistically optimum in terms of the shorter direct route which a field inspector would have to take to on-the-spot check anomalies.

Cost-effective and statistical assessment of the optimum quadrat size and sampling design is expected to produce a robust sampling method which makes it possible to obtain an accurate figure for anomaly occurrence (at 2%) and abundance (at least 90% complete).

### **12.1. Wiegert's method**

The method applied to determine the optimum quadrat size is based on the method proposed by Wiegert (1962). Two factors were considered crucial for deciding the optimum quadrat size: (i) the relative variance obtained for the variables of interest, i.e. anomaly occurrence and abundance, and (ii) the relative cost of spatial detection and registration of anomalies in the GIS database, the goal being to optimise measurements (to reduce variance) in terms of sampling time (cost).

This method requires setting up different experiments to obtain samples from different sized quadrats at constant sample size. The results should help to decide the optimum quadrat size and not to analyse anomalies diversity. Hence, although the 45 km<sup>2</sup> sample size (see above) was to apply to all twelve zones for assessing anomalies diversity, in this case it was possible to use any other sample size facilitating setting up quadrats in accordance with the sizes and designs tested.

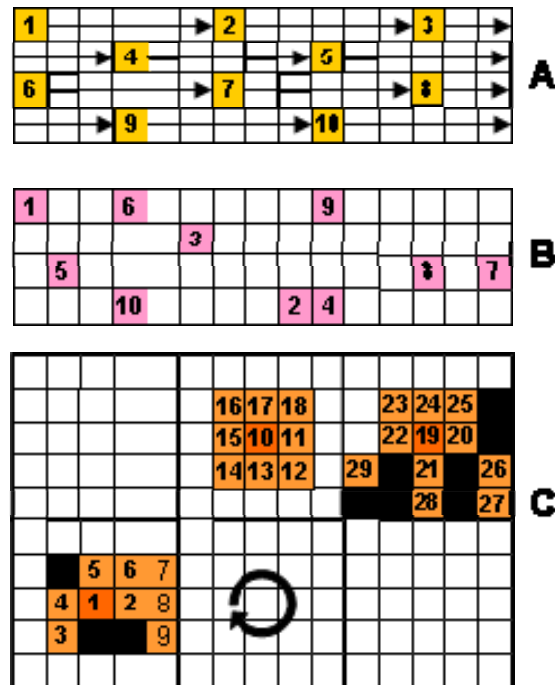
This was done for constant sampling of 75km<sup>2</sup> for square quadrats of six different sizes (0.1, 0.25, 0.5, 1, 3 and 5 km<sup>2</sup>) applied to zone FB2 for which the “true” anomalies assemblage was known.

### **12.2. Sampling designs**

From the orthophoto of zone FB2, from which external reference parcels within a boundary strip of 1 km width were excluded to avoid any incomplete reference parcels on the border, six different grids were built with quadrat sizes ranging between 0.1 and 5 km<sup>2</sup>. Each quadrat (in each grid) was numbered from 1 (top left-hand corner) to  $N$  (bottom right-hand corner). Within the GIS, for each grid, quadrats which did not contain at least one reference parcel (more than 50% of its area) were considered empty and were excluded. The remaining  $N'$  (non-empty) quadrats were then renumbered to ensure continuity of quadrat identification and selection during sampling design.

From each grid, the number of quadrats ( $n$ ) necessary to obtain a sample size of 75 km<sup>2</sup> were sampled by applying three different sampling designs: systematic (SYS), random (RAND) and systematic cluster (CLU) (Krebs, 1999; Mier and Piquelle, 2008) (see Figure 7).

Each “quadrat size (6)/sample design (3)” combination was replicated twenty times to assess variability. A total of 360 anomalies assemblages were obtained to be later compared to POP, the true anomalies assemblage in the zone FB2.



**Figure 7: The three different sampling designs applied**

**A = systematic; B = random; C = systematic cluster (for the “systematic cluster” design, different examples of sampling non-empty quadrats inside blocks from the central quadrat are indicated; the dark quadrats represent “empty quadrats” which are not considered in the grids – any empty quadrat was excluded and the next quadrat clockwise was considered.)**

- Systematic sampling (SYS) without replacement meant sampling  $n$  quadrats out of the  $N'$  available in a grid by sampling quadrats constantly spaced at  $[\text{integer}(N'/n)-1]$  quadrats; for instance, if  $n = 25$  and  $N' = 610$   $\text{int}[610/25]-1 = 23$ , then the first quadrat was  $i$  and the second was  $i + 23$ ; for the twenty replicates,  $i$  was equal to 1, 2, 3, etc.
- Random sampling design (RAND) meant simple random sampling without replacement where  $n$  quadrats were sampled randomly out of the remaining  $N'$  non-empty quadrats.
- Systematic cluster design (CLU) was performed on the initial grids before excluding empty quadrats. The grids were divided into  $M$  blocks as primary units also called “clusters” in the relevant literature. Each block contained a varying number of quadrats (or secondary units), depending on the quadrat size tested:  $3^2$ ,  $5^2$ ,  $5^2$ ,  $7^2$ ,  $9^2$  and  $15^2$  respectively for quadrat sizes of 0.1, 0.25, 0.5, 1, 3 and 5 km<sup>2</sup>. Five blocks were randomly sampled within the grid. In each of the five blocks, starting from the central quadrat and depending on the quadrat size tested, 3, 5, 15, 30, 60 and 150 non-empty quadrats were sampled clockwise; whenever an empty quadrat was sampled, it was excluded and the next quadrat was considered in order finally to

obtain the necessary number of quadrats per block. This follows Cochran's aligned or "square grid" systematic sample in two dimensions (Cochran, 1977). First a primary unit is sampled, then the corresponding secondary units per primary unit in the same relative positions in each block.

Compared with adaptive sampling (Thompson, 1992), it was assumed that the systematic cluster design should permit local sampling of rare anomalies supposedly clustered in complex landscapes. This sampling method should use the hypothetical spatial pattern in the anomaly population assumed to be clustered in order to obtain more precise measurements of occurrence and abundance. Furthermore, it should greatly reduce the travel costs related to field inspections to verify and confirm area discrepancies observed previously during the GIS-assisted stage.

Then, for each twenty replicates of each quadrat size/sampling design combination, the total number of parcels sampled, total number of anomalous parcels, total anomaly abundance and total anomaly abundance per type of anomaly were calculated. These values were then standardised to a single unit area (per 0.1 ha). Wiegert's method was then applied to each anomalies assemblage to determine the cost-effectiveness of the quadrat size and sampling design.



### 12.3. *Wiegert's cost calculations*

The sampling cost was calculated as follows:

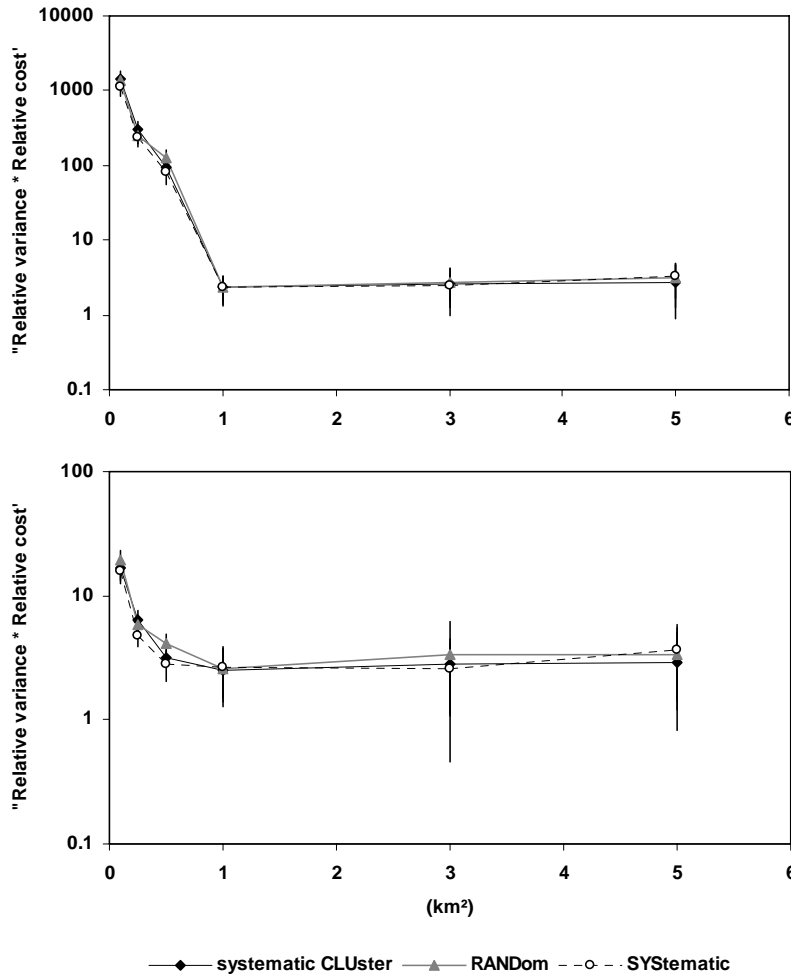
$$C_{sample} = (n_{parcel} * C_{parcel}) + (n_{anomalies} * C_{anomaly}) + (n - 1_{quadrat} * C_{quadrat}) \quad (\text{equation 7})$$

It was assumed that the sampling cost was strongly related to (i) the number of reference parcels contained within a quadrat, individually displayed and checked on screen, (ii) the number of anomalies detected and to be declared in the GIS table and (iii) the number of quadrats to check. The mean basic costs for these three components were measured during the pre-testing of zone FB2 and were approximated to, respectively, 3.5, 4.0 and 2.0 seconds.

Then, the product of (1) the “relative cost” (defined as the ratio between the time to take one sample of a given size divided by the minimum time to take one sample for all quadrat sizes considered) and (2) the “relative variance” (defined as the ratio between the square of the standard deviation for one given size divided by the square of the minimum standard deviation for taking one sample for all quadrat sizes considered) was calculated for each quadrat size for each replicate (Krebs, 1999).

### 12.4. *Results of Wiegert's method*

The results obtained from Wiegert's method are shown in Figure 8.



**Figure 8: Changes in product of (relative variance \* relative cost) depending on quadrat size for (a) anomaly occurrence and (b) total anomaly abundance (vertical lines indicate twice the standard deviation)**

Whatever the variable of interest considered, Figure 8 shows that the maximum precision at the lowest cost was observed from a quadrat size of 1 km<sup>2</sup>. Quadrat sizes smaller than 1 km<sup>2</sup> showed higher product values, indicating that the relative cost and/or relative variance were higher with smaller quadrat sizes. The product of (relative variance \* relative cost) increased slightly for quadrat sizes above 1 km<sup>2</sup> but, because of the logarithmic scale used in the figure, this is not visible. Second, whatever the quadrat size considered, the relative products for anomaly occurrence and anomaly abundance were not very different between sampling designs; this did not allow differentiation of the best sampling design. It just cleared the way to exclude small quadrat sizes (less than 1 km<sup>2</sup>) and indicated that a 1 km<sup>2</sup> square quadrat could be the appropriate quadrat size to use.

Further tests should be conducted to assess the bias and the imprecision of each remaining sampling design/quadrat size combination.

### *12.5. Sampling design control method*

Studies that have introduced methods to control sample size and sampling design include those by Lo *et al.* (1997), Christman (2003), Su and Quinn (2003) and Mier and Picquelle (2008). Like any other method, the ones used in this LPIS survey introduce bias into the estimators of the variables of interest (in this case, anomaly occurrence and abundance). This bias or systematic error is then measured by repeating sampling on the initial population and comparing the true values of the population with the repeated estimates.

The control method used in this study was adapted from Mier and Picquelle (2008). All the formulae to calculate estimators of mean population and variance and to compare properties of estimators between different survey designs against the true population of LPIS anomalies were carefully respected. The sole difference was the number of populations used in the control process: whereas Mier and Picquelle (2008) used 100 artificial populations simulated from ichthyoplankton surveys conducted between 1986 and 1998, this study used the single population of anomalies detected within pre-test zone (POP) FB2, where the global anomaly occurrence equalled 5.76%, which means that 5.76% of the LPIS reference parcels in zone FB2 contained at least one anomaly.

In this study, the comparison of survey design and remaining quadrat size ( $\geq 1$  km<sup>2</sup>) was based on the bias, precision and root square of the MSE (mean square error) of the estimated mean anomaly occurrence and standard error. The relative bias, imprecision and root square of the MSE expressed as a percentage of the true population anomaly occurrence were measured to allow comparison between sampling designs and quadrat sizes. MSE is often used to compare estimators showing different levels of bias and precision. In fact, the MSE corresponds to the sum of the bias and of the imprecision.

The results obtained for each of the nine remaining quadrat size/sampling design combinations are set out in Figure 9.

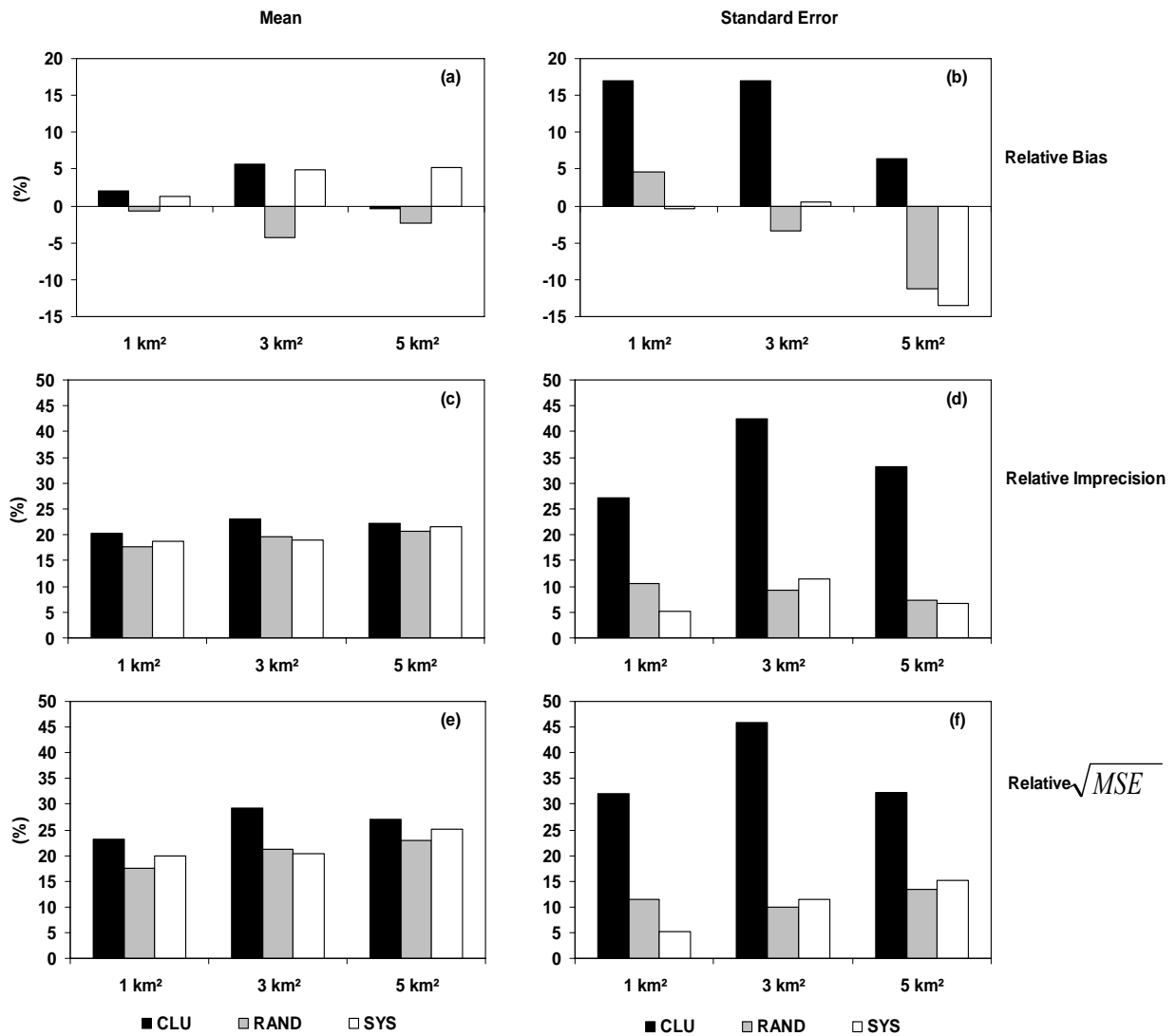


Figure 9: Relative bias, imprecision and  $\sqrt{MSE}$  of the mean and standard error of anomaly occurrence in the LPIS for FB2 for the nine remaining quadrat size/sampling design combinations averaged from twenty replicates from pre-test zone FB2

(a), (c) and (e) and (b), (d) and (f) are the relative bias, imprecision and  $\sqrt{MSE}$  for the mean occurrence and the standard error respectively.

Whatever the quadrat size, systematic (SYS) and systematic cluster (CLU) design estimators overestimated the mean anomaly occurrence, whereas random sampling design (RAND) estimators generally underestimated it (Figure 9(a)). The bias was not constant between quadrat sizes for all the sampling designs, but the smallest bias was observed for a 1 km<sup>2</sup> quadrat, considering all designs combined, with a bias of around 2%. This indicated that, regardless of the sampling design, a quadrat size of 1 km<sup>2</sup> should be used to sample anomalous LPIS parcels from orthophotos more accurately.

Despite the negligible bias in the mean observed for a 1 km<sup>2</sup> quadrat, the bias in the standard error was high and exceeded the 10% negligibility threshold (Cochran, 1977) (Figure 9(b)). This indicated that, although the mean estimators from the three designs appeared relatively accurate, above all with a 1 km<sup>2</sup> quadrat size, they nevertheless displayed considerable variability which, in certain situations, could lead to misinterpretation.

When comparing the precision obtained for the mean occurrence, results ranged between 15% and 25% and were not very different between sampling designs. The imprecision of the mean increased slightly in parallel to the quadrat size, confirming that better precision should be obtained both for the mean and for the standard error with a quadrat size of 1 km<sup>2</sup>. In the case of standard error, the systematic cluster design performed significantly worse than the other two, confirming the lower precision of the standard error which can be expected from the systematic cluster design. The relative  $\sqrt{MSE}$  as the sum of the relative bias and of the relative imprecision painted the same picture: (i) a 1 km<sup>2</sup> square quadrat appeared the best size and (ii) even though three designs were allowed for sampling the equivalent mean anomaly occurrence, the systematic cluster design was again less precise and more biased than systematic or random designs.

The results set out above confirmed those obtained from Wiegert's method: all in all, the statistically and cost-effectively optimum quadrat size for the GIS was 1 km<sup>2</sup>. This quadrat size was therefore selected for the remainder of this study.

As regards the relative  $\sqrt{MSE}$ , random or systematic designs should be chosen. At constant cost-effectiveness (see Figure 8), these two designs appeared more precise and less biased (see Figure 9) than the systematic cluster design. However, Wiegert's method measured cost-effectiveness taking into account the sole costs of GIS-assisted detection and ignored the travel costs incurred by one field inspector for verifying, on the spot, the area discrepancies detected. For that purpose, the minimum distance (as the crow flies) to link all the quadrats sampled was taken as a proxy for the travel costs and was measured in the GIS for each of the 20 replicates for each sampling design. Then a, analyse of variance was performed.

The travel costs (i.e. minimum direct distances in km) differed significantly between sampling designs ( $F(2.59) = 158.4416$ ,  $p\text{-value} < .0001$ ) at  $126.0 \pm 38.0SD$ ,  $208.7 \pm 8.3SD$  and  $314.2 \pm 43.0SD$  kilometres for the systematic cluster, systematic and random designs respectively. It was consequently assumed that travel costs for the systematic and random designs would be 166% and 249% higher than for the systematic cluster design. From this, it was assumed that, despite its relative weakness in terms of MSE concerning the standard error, the systematic cluster design would be the most suitable to achieve both representativeness (MSE concerning the mean is relatively similar between designs) and cost-effectiveness at the same time for LPIS contractors and field inspectors. It was perceived as the best compromise to address GIS and on-the-spot LPIS quality cost-effectively.

It was therefore decided to apply systematic cluster design to all the zones selected.

### **13. Sampling decision based on comparison of LPIS anomalies assemblages**

In the previous section, the optimum sampling method was decided by considering LPIS anomaly occurrence and/or LPIS anomaly abundance. However, the anomalies assemblages obtained from each sampling combination were never verified. Anomalies assemblage concerns the presence/absence of each of the seventeen anomaly types proposed in the typology and the exact abundance of each type detected in all the 1 km<sup>2</sup> quadrats sampled. The completeness of the anomalies assemblages sampled was verified against the true anomalies population of pre-test

zone FB2 (POP). This corresponded to assessment of the *similarity* between the POP assemblage and sampled assemblages to determine the degree to which the structure of the original FB2 anomalies assemblage was preserved when diversely sampled.

Currently two kinds of methods are widely used to estimate the similarity between species assemblages: univariate and multivariate methods (Cheng, 2004).

- Univariate methods are used when species by site matrices are reduced to a single coefficient of diversity, such as Shannon-Wiener's, Simpson's or Pielou's index of diversity. These indices are then discriminated using univariate methods such as analyses of variance (Clarke and Warwick, 1994).
- Ordination entails multivariate methods; different multivariate methods exist, such as hierarchical clustering (Johnson and Wichern, 1992), non-metric multidimensional scaling (NMDS) (Kruskal, 1964), correspondence analysis (CA) (Jongmann *et al.*, 1995) or principal component analysis (PCA) (Wold *et al.*, 1987). They start from a triangular matrix of similarity indices between every pair of anomalies assemblages. All the methods are applied to reduce the complexity of multivariate information in the original matrices to a low-dimensional picture.

There are numerous measurements of similarity (Legendre and Legendre, 1983), and confusion exists about which similarity measurement to use. Two broad classes of similarity coefficient exist: (i) binary coefficients using presence/absence (1/0) data, such as Jaccard's coefficient (Chao, 2005) or Sorensen's coefficient (Sorensen, 1948); these coefficients are generally used when only the lists of species are available and comparisons are possible at this lower level of resolution, weighting rare species the same as common species; (ii) quantitative coefficients for which supplementary information such as species abundance in an assemblage is required; among these, Morisita's index of similarity (Morisita, 1959) is considered the best overall measurement of similarity for ecological use (Wolda, 1981), almost independent of sample size (unlike Sorensen's index).

From zone FB2, 60 anomalies assemblages (3 sampling designs (1 km<sup>2</sup>) \* 20 replicates) were grouped together in a single table. A supplementary sample corresponding to the "true" FB2 anomalies assemblage (POP) was added to the table. Then, Morisita's similarity coefficient for each pair of anomalies assemblages (non-transformed data) was calculated as follows:

$$C_{\lambda} = \frac{2\sum(X_{ij}X_{ik})}{((\lambda_1 + \lambda_2)n_jn_k)} \quad (\text{equation 8})$$

where:

$$\lambda_1 = \frac{\sum(X_{ij}(X_{ij} - 1))}{(n_j(n_j - 1))} \quad (\text{equation 9})$$

$$\lambda_2 = \frac{\sum(X_{ik}(X_{ik} - 1))}{(n_k(n_k - 1))} \quad (\text{equation 10})$$

where  $C_{\lambda}$  = Morisita's index of similarity between samples  $j$  and  $k$ ,  
 $X_{ij}$ ,  $X_{ik}$  = the abundance of anomaly  $i$  in samples  $j$  and  $k$ ,  
 $n_j$ ,  $n_k$  = the total numbers of anomalies in samples  $j$  and  $k$ .

First, analyses of variance of Morisita's similarity index between sampling designs (from the twenty replicates) were performed. The mean similarity index was significantly different for systematic clusters ( $F(2.59) = 5.286$ ,  $p\text{-value} = 0.008$ ). The mean Morisita similarity between the POP and systematic cluster designs was  $93.60\% \pm 6.11\text{SD}$ , whereas it was  $96.66\% \pm 2.49\text{SD}$  and  $97.42\% \pm 1.65\text{SD}$  respectively for random and systematic designs. Despite this difference, all three sampling designs allowed sampling of assemblages very similar to the true LPIS anomalies assemblage detected for zone FB2. Whatever the sampling design, the completeness of the true anomalies assemblage was preserved.

Second, because the choice of sampling method cannot be decided from this sole analysis of variance, two-stage similarity ordination (Cuffney *et al.*, 2007) was also used to confirm the greater variability of systematic cluster design. This was done by using PAST© freeware (Hammer, 2001). Non-metric multidimensional scaling is based on a distance matrix calculated with Morisita's similarity distance. The algorithm then attempts to place the data points in a two- or three-dimensional coordinate system so that the ranked differences are preserved. For example, if the original distance between points 4 and 7 is the ninth largest of all distances between any two points, points 4 and 7 will ideally be placed so that the Euclidean distance between them in the 2D plane or 3D space is still the ninth largest. Non-metric multidimensional scaling intentionally does not take absolute distances into account. The program may converge on a different solution in each run, depending on the random initial conditions. Each run is actually a sequence of 11 trials, from which the one with smallest stress is chosen. The stress value indicates the obtained versus the observed ranks, i.e. the quality of the result. Ideally, all points should be placed on a straight ascending line ( $x = y$ ) and stress should be around zero. The results of the NMDS are shown in Figure 10.

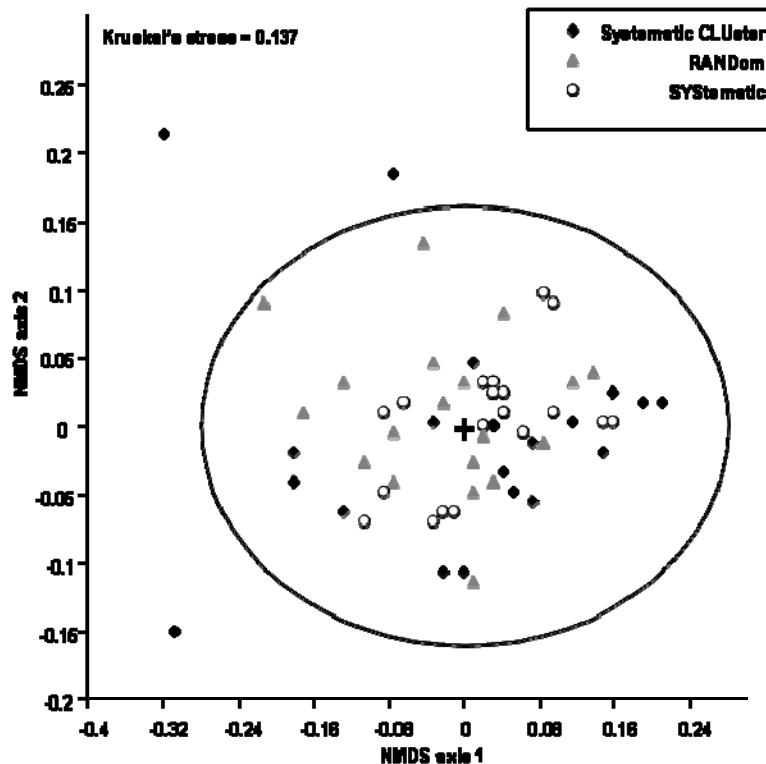


Figure 10: Non-metric multidimensional scaling (NMDS) plot of Morisita's similarity index calculated between the twenty 1 km<sup>2</sup> quadrat sampling design replicates and the true FB2 LPIS anomalies population (dark circle marks the 95% confidence interval)

From the correlation matrix of similarity, the ordination clearly and significantly (Kruskal stress = 0.137) showed that the three designs produced anomalies assemblages very similar to the true anomalies assemblage. Like systematic and random designs, and despite three replicates for which the similarity values were outside the 95% confidence circle, systematic cluster designs produced assemblages well grouped around the POP. These results confirmed the trend suggested earlier, i.e. that although CLU, SYS and RAND were adequate to sample LPIS anomalies assemblage from orthophotos and to provide good estimators of mean occurrence and abundance, CLU was less precise than the other two because of its higher variability.

Clearly, there is no obvious best design for sampling LPIS anomalies. Based on the MSE for occurrence, systematic cluster design would be a poor choice and systematic or random designs offer superior performance. However, the systematic cluster design's MSE for mean occurrence did not differ from the other two designs and was even smaller when using a 1 km<sup>2</sup> quadrat size (see Figure 9(a)). Second, the completeness of the true anomalies assemblage was correctly preserved by all three designs; CLU achieved a lower similarity value, but this value was still higher than the 90% completeness expected. Finally, systematic cluster design offered one significant benefit concerning post-management of anomalies detection by reducing by at least 166% the travel costs necessary to verify and confirm on the spot the area discrepancies observed. Consequently, the systematic cluster design was perceived as relevant for the successive steps; its relative imprecision and bias were taken into account by applying the interval of validity for anomalies occurrence results, as described later (see Table 5).

#### **14. Spatial distribution of anomalies in zone FB2**

As explained earlier, systematic cluster and other adaptive designs are generally used when species are expected to be spatially clumped. Here, even though CLU1 (systematic cluster design using 1 km<sup>2</sup> square quadrats) performed correctly (2% bias of mean occurrence and > 93% completeness of the true anomalies assemblage), assessment of the spatial distribution of anomalies in zone FB2 was required to decide whether it was ultimately applicable. To do this, the adequacy and validity of using systematic cluster design to sample LPIS anomalies was verified *a posteriori*.

The spatial distribution of anomalies within pre-test zone FB2 was determined by statistical and metric assessment.

##### **14.1. Statistical assessment of the spatial distribution of anomalies in zone FB2**

Statistical analysis of the spatial distribution of LPIS anomalies in zone FB2 is simple. Over the whole of zone FB2, sampled using a 1 km<sup>2</sup> quadrat size, the observed frequency distribution of the number of anomalies per quadrat was counted. Then, this observed frequency distribution was fitted against the expected frequency distribution. If a random spatial pattern prevailed, the Poisson distribution would be the appropriate descriptor of the data. If not, the spatial pattern could be aggregative, in which case negative binomial distribution should fit the data distribution (Patil *et al.*, 1971). By default, any observed frequency distribution not fitted by either Poisson or negative binomial distributions would suggest a uniform spatial pattern (Krebs, 1999).

The Poisson and negative binomial anomalies frequency distributions were tested and the goodness of fit verified.

##### **14.2. Metric assessment of the spatial distribution of anomalies in zone FB2**

To confirm the spatial pattern obtained from frequency distribution comparisons, several other indices of dispersion for quadrat counts were calculated:

- The “variance-to-mean” ratio; because it is a function of the sample size, application of this ratio to clumped populations is a problem. Nevertheless, Myers (1978) showed that the variance-to-mean ratio is only slightly affected by population density.
- Green’s coefficient of dispersion (Green, 1966) is based on the variance-to-mean ratio:

$$G_{Id} = \frac{\left(\frac{s^2}{\bar{x}}\right) - 1}{\sum(x) - 1} \quad \text{(equation 11)}$$

where  $\sum(x)$  is the total number of anomalies within the quadrats.

A negative value of  $G_{Id}$  indicates a uniform pattern, a positive value indicates a clumped pattern and 0 equals a random distribution. This was considered one of the indices almost independent of sample size and population density (Myers, 1978).

- Morisita’s index of dispersion (1962) (not to be confused with Morisita’s index of similarity, see Section C.14) is relatively independent of population density but is affected by sample size; consequently, it is not as good as Green’s coefficient of dispersion. For that reason, Smith-Gill (1975) set out to improve Morisita’s index of dispersion by standardising it on an absolute scale from -1 (uniform) to +1 (clumped).

From Morisita’s initial index of dispersion:

$$I_d = n * \left[ \frac{\sum x^2 - \sum x}{(\sum x)^2 - \sum x} \right] \quad \text{(equation 12)}$$

where  $I_d$  = Morisita’s index of dispersion,

$n$  = sample size,

$\sum x$  = the total number of anomalies within the quadrats,

$\sum x^2$  = the square of the total number of anomalies within the quadrats,

the uniform index ( $M_U$ ) and the clumped index ( $M_C$ ) are calculated from the following formulae:

$$M_U = \frac{\chi_{.975}^2 - n + \sum x_i}{(\sum x_i) - 1} \quad \text{(equation 13)}$$

and

$$M_C = \frac{\chi_{.025}^2 - n + \sum x_i}{(\sum x_i) - 1} \quad \text{(equation 14)}$$

where  $\chi_{.975}^2$  and  $\chi_{.025}^2$  are chi-squared values from the table with  $(n-1)$  degree of freedom that have respectively 97.5% and 2.5% of the area to the right,

$x_i$  = the total number of LPIS anomalies (all categories combined) within the  $i^{\text{th}}$  quadrat,

$n$  = the number of quadrats.



Then, the standardised Morisita's index of dispersion is calculated by one of the following formulae:

$$\text{when } I_d \geq M_C > 1.0 \quad I_p = 0.5 + 0.5 \left( \frac{I_d - M_C}{n - M_C} \right) \quad (\text{equation 15}),$$

$$\text{when } M_C > I_d \geq 1.0 \quad I_p = 0.5 \left( \frac{I_d - 1}{M_U - 1} \right) \quad (\text{equation 16}),$$

$$\text{when } 1.0 > I_d > M_U \quad I_p = -0.5 \left( \frac{I_d - 1}{M_U - 1} \right) \quad (\text{equation 17}),$$

$$\text{when } 1.0 > M_U > I_d \quad I_p = -0.5 + 0.5 \left( \frac{I_d - M_U}{M_U} \right) \quad (\text{equation 18}).$$

The standardised Morisita's index of dispersion is independent of population density and sample size. Green (1966) recommended that the minimum sample size should be 50 quadrats and that, in case of a highly clumped pattern, at least 200 quadrats are required. In the case of zone FB2, 1 023 non-empty quadrats of 1 km<sup>2</sup> were available.

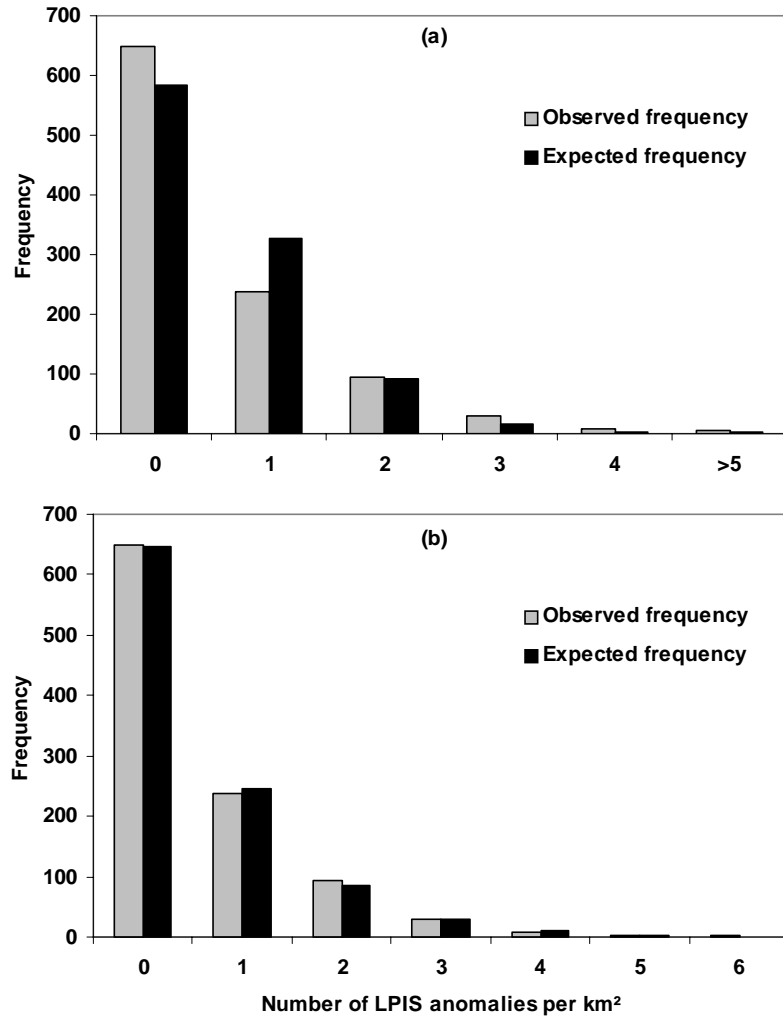
Table 3 shows the limits of each of the three indices of dispersion used.

**Table 3: Expected (maximum) uniformity, (maximum) aggregation and randomness values calculated for the three indices of dispersion used for LPIS anomalies spatial pattern determination ( $\sum x$  = sum of quadrat counts)**

	Value expected under		
	Maximum uniformity	Randomness	Maximum aggregation
Variance-to-mean ratio	0	1	$\sum x$
Green's index of dispersion ( $G_{Id}$ )	$\frac{-1}{\sum (x) - 1}$	0	+1
Standardised Morisita's index of dispersion ( $I_p$ )	-1	0	+1

### 14.3. Results of the statistical estimation of the spatial distribution of anomalies in zone FB2

The LPIS spatial pattern was tested on the whole of zone FB2. From the 1 023 quadrats, both the observed and the expected frequency from Poisson and negative binomial distributions were calculated. Figure 12 sets out the diagrams obtained.



**Figure 11: Observed and expected frequency distributions of the number of LPIS anomalies per quadrat counted over 1 023 quadrats (quadrat size = 1 km<sup>2</sup>) in zone FB2 ( $\bar{x} = 0.559$  anomalies.km<sup>-2</sup>;  $s^2 = 0.818$ ; (a) expected frequency from Poisson distribution and (b) expected frequency from negative binomial distribution)**

In the case of the Poisson distribution, the null hypothesis that the Poisson distribution fits the observed data was tested by the chi-squared goodness-of-fit test:

$$\chi^2 = \sum \left[ \frac{(\text{obs. freq} - \text{exp. freq})^2}{\text{exp. freq}} \right] = 54.21 \quad (\text{equation 19})$$

As the minimum number of individuals expected by class was 3, quadrat counts for 5, 6 and 7 anomalies per quadrat were added together to satisfy this condition. Then, the tabular value of  $\chi^2$  for  $\alpha = .05$  and  $\nu = 4$  was equal to 9.49. As the observed chi-squared value (54.21) was larger than the tabular value, the null hypothesis was rejected.

Consequently, contrary to the hypothesis of a random distribution of LPIS anomalies within zone FB2, the test showed that the distribution was not random. Furthermore, the index of dispersion from the variance-to-mean ratio (equal to 1.43) suggested that the distribution was clumped.

In the case of the negative binomial distribution, the null hypothesis was that the negative binomial distribution fitted the observed frequency if the LPIS anomalies were aggregated. The negative binomial distribution is a discrete probability distribution governed by two parameters: the negative binomial  $k$  and  $p$  ( $p = q - 1$ ) related to the mean such as  $\mu = kp$ . For each class of frequency, a probability of a quadrat containing  $i$  anomalies is given by:

$$P_i = \left[ \frac{\Gamma(k+i)}{i! \Gamma(k)} \right] \left( \frac{\mu}{\mu+k} \right)^i \left( \frac{k}{\mu+k} \right)^k \quad (\text{equation 20})$$

where  $\mu$  = the mean of the distribution,

$\Gamma$  = a gamma function (see Krebs, 1999),

$k$  = the negative binomial exponent for which an initial approximate

$\hat{k} = \frac{\bar{x}^2}{s^2 - \bar{x}} = 1.2069$  can be calculated and used to calculate a maximum likelihood

estimate for  $k$  by trial and error from equation 4.14 proposed by Krebs (1999). Applied here, the maximum likelihood estimate for  $k$  was 1.2034. Applying equation 20, the expected probabilities for each counter of anomaly by quadrat was calculated (see Figure 11(b)).

Then, the null hypothesis that the negative binomial distribution fits the observed data was tested by: (i) the chi-squared goodness-of-fit test and (ii) the U-statistic goodness-of-fit test:

- Chi-squared goodness-of-fit test: for  $\alpha = 0.05$  and  $\nu = n - 3 = 4$ , the tabular value of  $\chi^2$  was equal to 9.49, whereas the observed value of  $\chi^2$  was equal to 2.673; because the observed value was smaller than the tabular one it was possible to accept the null hypothesis;
- U-statistic goodness-of-fit test: the corresponding formulae can be found in Krebs (1999). For  $\alpha = 0.05$ , if the observed value of U exceeds two standard errors of U ( $2SE_U$ ), the null hypothesis that the spatial pattern of LPIS anomalies is aggregative is rejected. In this case, an observed value of  $U = -7.6 \cdot 10^{-4}$  and  $2SE_U = 5.86 \cdot 10^{-2}$  was obtained; the observed value of U was smaller than  $2SE_U$ .

Based on these two tests, the null hypothesis that the negative binomial distribution was an adequate fit to the observed data was accepted and, in accordance with the previous results, the LPIS anomalies distribution pattern was clumped for zone FB2 in 2005.

#### 14.4. Results of the metric estimation of the spatial distribution of anomalies in zone FB2

Even though statistical evaluation of the spatial dispersion of LPIS anomalies within zone FB2 can be easily performed by specific software (this study used Ecological Methodology Software ©; Krebs, 1999), it remains complex. For this reason, the results set out above concerning the spatial distribution of anomalies were confirmed by using spatial distribution metrics.

The different indices of dispersion used here produced a variance-to-mean ratio and Green's and standardised Morisita's indices of dispersion of 1.46,  $8.11 \cdot 10^{-4}$  and 0.50 respectively. Positive and non-null randomness values in Table 3 suggested an aggregative pattern of LPIS anomalies. Moreover, the standardised Morisita's index of dispersion was recognised as the best estimator of spatial distribution (Krebs, 1999) and the value of 0.50 confirmed results obtained from previous

statistical tests: the LPIS anomalies spatial pattern in zone FB2 was aggregative (in 2005). Also the systematic cluster sampling design could be robustly used.

### **15. Preliminary conclusions and decisions concerning the final sampling method**

To sum up this part, concerning pre-test zone FB2 in 2005, the study confirmed that: (i) the LPIS anomalies pattern was clumped and, consequently, systematic cluster design was relevant to sampling LPIS anomalies from orthophotos, (ii) a minimum sample size of 45 km<sup>2</sup> should allow sampling of LPIS anomalies in most European LPIS situations and (iii) 1 km<sup>2</sup> was the most accurate quadrat size to use. It also showed that (iv) although systematic cluster design was less precise than other designs it provided an excellent estimate of mean anomaly occurrence and was more cost-effective.

For these reasons, the calibration of the sampling method can be considered validated and will be applied in the rest of this study.

The sampling design characteristics specified above have been calibrated for the sole case of zone FB2 in 2005. However, this sampling method might not be suitable for application to every Member State. Consequently, adequacy assessment would be of prime importance for validation of systematic cluster sampling, as discussed in Part E.

## ***PART E. Analysis of diversity in LPIS anomalies***

### **16. Validation of sampling method**

As indicated in the preliminary conclusions (Section 15), confirmation of the adequacy of the sampling method was an important milestone on the way to interpreting the diversity in LPIS anomalies. This took the form of assessing the adequacy of the sampling method calibrated from zone FB2 for eliciting LPIS anomalies assemblages and, especially, anomalies richness for all the zones.

As explained by Coleman *et al.* (1982), to estimate the anomalies richness of areas larger than those sampled, the “anomalies richness/sampled area” relationship is important. Indeed, as suggested by Arrhenius (1921), all other conditions being equal, the number of species is related to the area sampled. Then, the observed number of anomalies ( $A_{\text{obs}}$ ) sampled is used as a surrogate for the true number of species ( $A_{\text{true}}$ ). If  $A_{\text{obs}}$  is expected precisely to match  $A_{\text{true}}$ , the bias of the method leads to the exclusion of many rare species and to underestimation of  $A_{\text{true}}$  when  $A_{\text{true}}$  is high or when species detectability is low (Palmer, 1990). A large number of extrapolation methods have been developed to reduce this bias (Colwell, 2004). They are classified as either species accumulation curves or as non-parametric estimators. The latter are generally less biased but seldom used, as the former is preferred because of its simplicity.

- On the one hand, species accumulation curves extrapolate species richness vs. sample size (i.e. area) data to an asymptote of total richness (Colwell and Coddington, 1994). The most commonly used models are the exponential (Holdridge *et al.*, 1971), the Michaelis-Menten (Michaelis and Menten, 1913) or the sigmoid models (Tjorve, 2004). As reported by Brose *et al.* (2003), the performances of these equations are related to the species abundance distribution, the spatial species aggregation and/or the habitat heterogeneity.
- On the other, a number of non-parametric methods have been developed specifically for estimating species richness from samples. All are non-parametric in the statistical sense, although their performance clearly depends on the underlying empirical distribution. Chao (1984) derived simple estimator *Chao1* of the true number of species in an assemblage by taking into account singletons (species represented by only one individual within the assemblage) and doubletons (species with exactly two individuals in the assemblage). She also proposed *Chao2* as an estimator requiring only presence/absence data. In parallel, other estimators such as *Jackknife1* and *Jackknife2* (Burnham and Overton, 1979) estimate species richness on the basis of the number of species that occur in only one and in exactly two samples respectively.

Colwell and Coddington (1984) provided a complete review of these parametric and non-parametric methods to estimate maximum species richness in samples and Tjorve (2003) tested a wide range of exponential and sigmoid asymptotic models for species accumulation curve methods.

In this survey, sampling method adequacy was assessed by assuming that a constant sampling effort of 45 km<sup>2</sup> was sufficient to estimate the maximum anomalies richness reliably. This assumption was tested by applying anomalies accumulation curves and non-parametric estimators to evaluate the completeness of the inventories of LPIS anomalies in each of the twelve zones.

#### **16.1. Anomalies accumulation curves theory**

Concerning accumulation curves, the order in which samples are added to the total is relevant to the shape of the curve; to eliminate this effect, the order of entry of the samples was randomised

500 times using Estimate-S software (Colwell, 2004). Then, as advocated by Keating and Quinn (1998), least squares non-linear regressions with the quasi-Newton algorithm in JMP 6.0 (SAS Institute) were used to fit the three accumulation curve models selected to the mean species richness obtained from Estimate-S.

Next, the number of quadrats required to reach 90% and 95% of the estimated maximum richness and the percentage of completeness achieved when applying a sampling effort of 45 quadrats ( $E_{45}$ ) were calculated as proof of the completeness of the sampling method (Shiu and Lee, 2003; Hernandez-Stefanoni and Ponce-Hernandez, 2004).

The three anomalies accumulation curve models were:

- the Clench model (Soberon and Llorente, 1993):

$$A = \frac{bq}{1 + cq} \quad (\text{equation 21})$$

where  $A$  = the estimated number of anomalies or richness,  
 $q$  = the number of quadrats,  
 $b, c$  = the parameters of the model given by the regression analysis,  
with  $\frac{b}{c}$  = the asymptote or the maximum number of anomalies.

From this, the number of quadrats required to reach 90% and 95% of the estimated anomalies richness was given by:

$$q_{90} = \frac{0.9}{c(1-0.9)} \quad \text{and} \quad q_{95} = \frac{0.95}{c(1-0.95)} \quad (\text{equation 22})$$

- the Linear Dependence model (Soberon and Llorente, 1993):

$$S_A = \frac{b}{c} * (1 - \exp(-cq)) \quad (\text{equation 23})$$

where  $S_A$  = the estimated number of anomalies or richness,  
 $q$  = the number of quadrats,  
 $b, c$  = the parameters of the model given by the regression analysis,  
with  $\frac{b}{c}$  = the asymptote or the maximum number of anomalies.

From this, the number of quadrats required to reach 90% and 95% of the estimated anomalies richness was given by:

$$q_{90} = \frac{-1}{c * \text{Ln}(1-0.9)} \quad \text{and} \quad q_{95} = \frac{-1}{c * \text{Ln}(1-0.95)} \quad (\text{equation 24})$$

- the Gompertz model (Ratkowsky, 1990; Tjorve, 2003):

$$S_A = b * e^{(-c * e^{(-d * q)})} \quad (\text{equation 25})$$

where  $S_A$  = the estimated number of anomalies or richness,  
 $q$  = the number of quadrats,

$b, c, d$  = the parameters of the model given by the regression analysis,  
with  $b$  = the asymptote or the maximum number of anomalies.

From this, the number of quadrats required to reach 90% and 95% of the estimated anomalies richness was given by:

$$q_{90} = \frac{\ln\left(\ln\left(\frac{0.9 * b}{b}\right)\right)}{\frac{-c}{-d}} \quad \text{and} \quad q_{95} = \frac{\ln\left(\ln\left(\frac{0.95 * b}{b}\right)\right)}{\frac{-c}{-d}} \quad (\text{equation 26})$$

### 16.2. Non-parametric estimators

Non-parametric estimators were *Chao1*, *Chao2* and first- and second-order *Jackknife* (*Jack1*, *Jack2*). They were calculated using Estimate-S software (Colwell, 2004).

### 16.3. Anomalies accumulation curves for completeness of the sampling method

The anomalies accumulation curves obtained from the Linear Dependent (LD) and Clench models and the observed data are shown in Appendix B for each of the twelve zones selected. The anomalies accumulation curves from the Gompertz model are not included to avoid overloading the diagrams. The estimated parameters and goodness-of-fit ( $r^2$ ) and non-parametric estimators are indicated in Table 4.

Based on the traditional  $r^2$  rule of thumb ( $r^2$  values higher than or equal to 0.95 confirm a good fit), the LD model does not fit the observed data well (mean  $r^2 = 0.89 \pm 0.15\text{SD}$ ). This was especially true for zone CP1, for which the LD model did not fit the observed data ( $r^2 = 0.44$ ). On the other hand, the anomalies accumulation curves fitted very well with the Clench and Gompertz models (mean  $r^2 = 0.98 \pm 0.03\text{SD}$  and  $99.22 \pm 1.43\text{SD}$  respectively) (see Table 4). Because the goodness of fit (i.e.  $r^2$ ) was sometimes poor or even bad when using the LD model, it is advisable to choose the Clench, or even better, Gompertz model to achieve a higher level of significance.

For each of the twelve zones, the LD model predicted a lower asymptote than the Clench model (see Figure 13 and Appendix B), always providing an estimated maximum richness ( $S_{\text{est}}$ ) lower than the observed values ( $S_{\text{obs}}$ ). By contrast, the maximum estimated richness from the Clench model was generally higher than the observed richness, sometimes by more than 10% (for zones FB2 and PB2) (see Figure 13). This confirmed previous results about the tendency of the Clench and LD models respectively to overestimate and underestimate species richness (Moreno and Halffer, 2001; Hernandez-Stefanoni and Ponce-Hernandez, 2004). In the case of the Gompertz model,  $S_{\text{est}}$  is usually lower than  $S_{\text{obs}}$ , but the Gompertz model gave the most accurate estimate of  $S_{\text{obs}}$  with a mean difference of -2.11% ( $\pm 1.20\text{SD}$ ) expressed as a percentage of  $S_{\text{obs}}$ . The mean difference between  $S_{\text{obs}}$  and  $S_{\text{est}}$  was equal to 4.69% ( $\pm 4.28\text{SD}$ ) and to -5.55% ( $\pm 3.30\text{SD}$ ) for the Clench and LD models respectively.

Altogether, use of asymptotic models to fit the anomalies accumulation curves mainly confirmed the adequacy of a 45 km<sup>2</sup> sampling effort to sample  $S_{\text{obs}}$ . Apart from zone FB2 (with the Clench model), the rule of thumb  $E_{45} \geq 90\%$  was respected in every zone with all the models. It allowed a fair representation of the number of anomalies in the zones selected. On average,  $S_{\text{obs}}$  obtained from 45 km<sup>2</sup> equalled 93.70% ( $\pm 3.04\text{SD}$ ), 99.92% ( $\pm 0.14\text{SD}$ ) and 99.22 ( $\pm 1.43\text{SD}$ ) of  $S_{\text{est}}$  from the Clench, LD and Gompertz models respectively (see Table 4). Under these conditions, detection of one extra anomaly within the assemblages became increasingly expensive in terms of sampling effort and the observed richness values were therefore considered satisfactory and the inventories of LPIS anomalies in the different zones comparable.

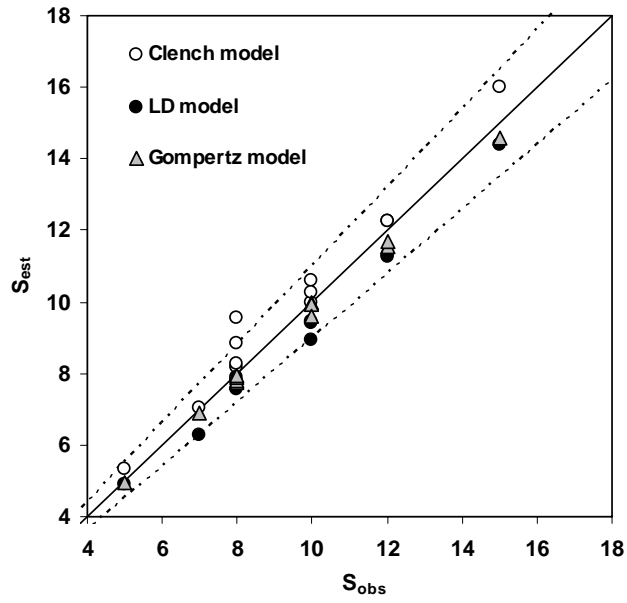


Figure 12: Observed ( $S_{obs}$ ) against estimated ( $S_{est}$ ) values of anomalies richness for each of the twelve LPIS zones studied from the three asymptotic models used (solid line = the x/y relationship; dashed lines =  $\pm 10\%$  of the x/y relationship)



**Table 4: Predicted parameters from two anomalies accumulation curve models (Clench and linear dependence) fitted to the twelve LPIS zones studied ( $S_{obs}$  = observed richness; a, b = predicted parameters of the asymptotic models;  $S_{est}$  = estimated richness (= a/b for the Clench and LD models and = a for the Gompertz model);  $E_{90\%}$  ( $E_{95\%}$ ) = minimum number of quadrats to sample 90% (95%) of  $S_{est}$ ;  $E_{45}$  =  $S_{obs}/S_{est}$  ratio expressed as a percentage when applying a sampling effort of 45 quadrats (= 45 km<sup>2</sup>); underlined values =  $r^2$  values below the 0.95 rule of thumb; bold values =  $E_{45}$  values lower than 90% of the expected completeness of the estimated richness  $S_{est}$ )**

			Anomalies accumulation curves													
			Clench model						LD model							
LPIS	Zones	$S_{obs}$	a	b	$S_{est}$	$R^2$	$E_{90\%}$ (km <sup>2</sup> )	$E_{95\%}$ (km <sup>2</sup> )	$E_{45}$ (%)	a	b	$S_{est}$	$R^2$	$E_{90\%}$ (km <sup>2</sup> )	$E_{95\%}$ (km <sup>2</sup> )	$E_{45}$ (%)
	AP1	15	5.15	0.32	15.99	0.99	27.9	59.0	93.5	3.14	0.22	14.39	0.96	10.6	13.7	99.9
AP	AP2	7	1.85	0.26	7.05	0.95	34.2	72.3	92.2	0.99	0.16	6.28	<u>0.92</u>	14.6	20.9	99.8
	AP3	10	2.82	0.27	10.59	0.96	33.7	71.2	92.3	1.70	0.18	9.44	<u>0.90</u>	12.8	16.7	99.9
CP	CP1	10	5.56	0.56	9.99	<u>0.91</u>	16.2	34.1	96.2	9.19	1.03	8.93	<u>0.41</u>	2.2	2.9	100.0
	CP2	8	4.05	0.49	8.19	0.99	18.2	38.5	95.7	2.29	0.30	7.56	<u>0.91</u>	7.6	9.9	100.0
	CP3	5	1.87	0.35	5.31	0.99	25.5	53.9	94.1	1.06	0.22	4.90	0.97	10.6	13.8	99.99
FB	FB1	12	4.99	0.41	12.26	0.98	22.1	46.6	94.8	2.73	0.24	11.24	<u>0.91</u>	9.5	12.3	100.0
	FB2	8	1.20	0.13	9.53	0.99	71.3	150.5	<b>85.0</b>	0.92	0.12	7.82	0.99	19.5	25.4	99.5
	FB3	12	6.03	0.49	12.27	0.98	18.3	38.7	95.7	3.42	0.30	11.33	<u>0.87</u>	7.6	9.9	100.0
PB	PB1	8	6.69	0.81	8.26	0.99	11.1	23.5	97.3	3.59	0.46	7.80	<u>0.90</u>	5.0	6.5	100.0
	PB2	8	2.50	0.28	8.84	0.99	31.9	67.3	92.7	1.54	0.20	7.90	0.98	11.8	15.4	99.9
	PB3	10	4.08	0.40	10.28	0.99	22.7	46.1	94.9	2.40	0.26	9.41	<u>0.92</u>	9.0	11.0	100.0

(to be continued)

			Anomalies accumulation curves								Non-parametric estimators							
			Gompertz model								$E_{45}^{Chao1}$		$E_{45}^{Chao2}$		$E_{45}^{Jack1}$		$E_{45}^{Jack2}$	
LPIS	Zones	S <sub>obs</sub>	a	b	c	S <sub>est</sub>	R <sup>2</sup>	E <sub>90%</sub> (km <sup>2</sup> )	E <sub>95%</sub> (km <sup>2</sup> )	E <sub>45</sub> (%)	Chao 1	(%)	Chao 2	(%)	Jack 1	(%)	Jack 2	(%)
AP	AP1	15	14.58	1.16	0.19	14.58	0.97	12.89	16.75	99.97	15.00	100.00	15.00	100.00	15.98	93.87	16.00	93.75
	AP2	7	6.88	0.88	0.08	6.88	0.96	25.43	34.06	97.96	7.00	100.00	7.98	<b>87.72</b>	8.96	<b>78.13</b>	10.87	<b>64.40</b>
	AP3	10	9.96	0.91	0.10	9.96	0.99	21.03	28.05	99.10	10.00	100.00	10.00	100.00	10.98	91.07	11.00	90.91
CP	CP1	10	9.91	0.6	0.10	9.91	0.98	17.83	25.15	99.28	10.00	100.00	10.00	100.00	10.98	91.07	11.00	90.91
	CP2	8	7.73	0.86	0.18	7.73	0.97	12.37	16.01	99.98	5.00	100.00	5.00	100.00	5.00	100.00	5.00	100.00
	CP3	5	4.93	1.21	0.2	4.93	0.98	11.53	15.48	99.98	8.00	100.00	8.00	100.00	8.98	<b>89.09</b>	9.93	<b>80.56</b>
FB	FB1	12	11.57	0.85	0.15	11.56	0.96	14.40	19.38	99.87	13.00	92.31	12.98	92.45	13.96	<b>85.96</b>	15.87	<b>75.61</b>
	FB2	8	7.82	1.69	0.14	7.82	0.99	19.71	24.83	99.70	8.00	100.00	8.00	100.00	8.98	<b>89.09</b>	9.93	<b>80.56</b>
	FB3	12	11.70	0.77	0.15	11.70	0.97	13.16	17.91	99.92	12.00	100.00	12.00	100.00	12.98	92.45	13.00	92.31
PB	PB1	8	7.91	0.78	0.26	7.91	0.96	7.66	10.41	100.00	8.00	100.00	8.00	100.00	8.00	100.00	8.00	100.00
	PB2	8	7.95	1.32	0.19	7.95	<u>0.89</u>	13.18	16.94	99.98	8.00	100.00	8.00	100.00	8.00	100.00	8.00	100.00
	PB3	10	9.61	0.98	0.17	9.61	0.94	21.86	46.14	94.88	10.00	100.00	10.00	100.00	10.98	91.07	11.00	90.91

#### *16.4. Non-parametric estimators for completeness of the sampling method*

Concerning non-parametric methods, the different estimators used provided very variable estimation of the maximum richness for all the zones. The *Chao1* and *Chao2* estimates were close to the observed richness (mean  $E_{45}$  values of  $99.36 \pm 2.13SD$  and  $98.35 \pm 3.82SD$  respectively), but both the first- and, even more so, second-order *Jackknife* estimators indicated more limited completeness of the sampling effort ( $91.82 \pm 6.08SD$  and  $88.33 \pm 10.52SD$  respectively). The higher variance of  $E_{45}$  values obtained from *Jackknife* estimators revealed that their completeness was not always good, depending on the zone considered. Four zones were considered poorly sampled by a 45 km<sup>2</sup> effort with *Jackknife*, whereas only one was incomplete when estimated with *Chao* (see Table 4). The lowest  $E_{45}$  value was observed for zone AP2 from the *Chao2*, *Jack1* and *Jack2* estimators alike, whereas no zones were considered under-sampled with *Chao1*.

#### *16.5. Final statement on the comparability of anomalies assemblages*

Regarding the results obtained from anomaly accumulation curves and from non-parametric estimation, anomalies' assemblages obtained from 45 km<sup>2</sup> sampling effort were relatively complete. If maximum anomalies' richness could be considered as non-reached for FB2 when assessing by the anomalies accumulation curve method (depending of the model used), AP2 would be the one from the non-parametric method (common to all the non parametric estimators). Because zones considered as potentially under sampled are not common among methods or even more among models or estimators, we assumed here that no method and even more, no precise models or estimators constantly suggested a specific zone as under sampled. Then, underestimation situation could be more related to the method than to the sampling effort and we considered that 45 km<sup>2</sup> sampling size provided an accurate enough estimation of the maximum richness i.e. the number of anomalies existing in each zone for all the zones selected. Concerning PB3 for which a sampling effort of 120 km<sup>2</sup> was initially calculated (Table 2), all the parametric and non-parametric methods showed here that maximum anomalies' richness was reached when performing 45 km<sup>2</sup> sampling effort. This indicated that the sampling effort was adequate and that its determination from a pre-test zone where quality is known as good or very good is a solution. All together, this authorized to further perform analysis of the different components of the LPIS anomaly diversity because of the adequacy of our 45km<sup>2</sup> sampling size to sample at least 90% of the anomaly richness existing in all zones.

### **17. Description of samples**

Reference parcels in the quadrats sampled were counted. Then, the absolute occurrence and abundance by quadrat (or by parcel when necessary) were determined for each of the seventeen anomaly types listed. The preliminary results by LPIS zone are summarised in Table 5. Because they were mean occurrence values by zone, they have been corrected to integrate the  $\sqrt{MSE}$  relative to a systematic cluster sampling design of 1 km<sup>2</sup> quadrats, as described in Section 12.5 (i.e.  $\pm 23.16\%$  of the mean).

#### *17.1. Areas sampled*

On the one hand, as mentioned in the introduction, LPIS types displayed high variability in the mean reference parcel area, from 0.64 ha ( $\pm 0.93SD$ ) for zone CP3 to 54.76 ha ( $\pm 61.85SD$ ) for PB3. As expected, CP and AP had lower reference parcel areas and FB and PB the highest. Because of that, at a constant sampling effort of 45 km<sup>2</sup>, an inversely proportional number of parcels were sampled by zone (from 154 in zone PB3 to 6 243 in CP3). If it were decided to sample reference parcels instead of quadrats, at a constant number of parcels the total area

sampled would vary by a factor of 50. This would have introduced a considerable source of bias based on the parcel area-related assumption made.

Use of a constant sample size (45 km<sup>2</sup>) made it possible to free the result from the possible reference parcel area effect. The parcel area sampled was between 16.85 km<sup>2</sup> (AP3) and 56.77 km<sup>2</sup> (CP2). Most of the zones offered a sample area around or lower than 45 km<sup>2</sup>. However, three zones (PB2, CP2 and, especially, PB3) showed an over-sampled area; at the opposite end, under-sampling was observed for AP3 (< 20 km<sup>2</sup>), mainly due to (i) exclusion of areas with non-agricultural land uses (forests, urban zones, etc.) and (ii) agricultural parcels not registered in the LPIS as eligible parcels (i.e. not receiving CAP aid) but sampled within quadrats. In this typical case, a relatively correct agricultural area close to 45km<sup>2</sup> was sampled in quadrats, but a large proportion of the parcels contained within the quadrats were not registered within the LPIS and, consequently, not counted. Concerning over-sampling, the reference parcels-to-quadrat rules (see Part C) induced over-sampling of the number of parcels and, hence, of the area when the reference parcels area was large (especially PB).

To solve this problem, two parcels-to-quadrat rules could be envisaged. First, by deciding to select solely quadrats in which at least 75% of the area contained reference parcel(s), under-sampling should be limited. Second, to address over-sampling, only quadrats for which the maximum sampled parcel area outside the quadrat does not exceed 20% could be suitable. Although easy to implement within the GIS, these rules remain controversial: in highly segmented LPIS, such as in mountainous zones or fragmented landscapes (farmland criss-crossed by hedges and trees), the number of quadrats complying with these rules could be limited and the sample difficult to obtain. For that reason, and because the results were observed *a posteriori*, these new parcel-to-quadrat rules were not applied in this study. Although the authors suggest that all future users should take them into consideration.

**Table 5: Primary results concerning the LPIS zones selected during the 2005 survey: sample size - 45 km<sup>2</sup> for all zones**

LPIS types	LPIS zones	Orthophoto area (km <sup>2</sup> )	Total area sampled (km <sup>2</sup> )	Mean sampled reference parcel area (ha)	Total No of reference parcels	Total No of reference parcels sampled	Sampled reference parcel (%)	No of anomalous reference parcels	Anomalous reference parcels percentage (%) ( $\sqrt{MSE}$ corrected)	Total No of anomalies
AP	AP1	1 213.4	48.77	5.11 ( $\pm 7.28SD$ )	18 307	866	4.73	237	27.37 (21.03 – 32.24)	460
	AP2	595.8	32.32	1.28 ( $\pm 1.18SD$ )	6 507	1 321	20.30	167	12.64 (9.71 – 14.89)	223
	AP3	506.5	16.85	3.70 ( $\pm 4.91SD$ )	5 647	879	15.57	232	16.39 (12.59-19.31)	614
CP	CP1	1 938.9	33.79	4.44 ( $\pm 0.88SD$ )	172 540	6 233	3.61	1 762	28.27 (21.72 – 33.30)	3 061
	CP2	2 003.4	56.77	2.86 ( $\pm 5.40SD$ )	44 784	2 547	5.69	121	4.75 (3.65 – 5.60)	170
	CP3	278.7	39.73	0.64 ( $\pm 0.93SD$ )	69 170	6 243	9.02	504	8.07 (6.20 – 9.51)	623
FB	FB1	784.5	46.27	7.48 ( $\pm 8.70SD$ )	5 365	619	11.54	237	38.29 (29.42 – 45.10)	525
	FB2	930.2	42.94	6.41 ( $\pm 12.33SD$ )	9 936	573	5.77	48	8.38 (6.44 – 9.87)	60
	FB3	311.3	45.88	5.82 ( $\pm 7.38SD$ )	3 302	788	23.86	434	55.08 (42.32 – 64.88)	1 143
PB	PB1	376.8	35.58	6.05 ( $\pm 10.77SD$ )	3 068	547	17.83	271	49.54 (38.07 – 58.36)	578
	PB2	1 105.9	50.79	5.51 ( $\pm 8.92SD$ )	39 118	922	2.36	101	10.95 (8.41 – 12.90)	186
	PB3	928.6	84.32	54.76 ( $\pm 61.85SD$ )	2 144	154	7.18	103	66.88 (51.39 – 78.78)	1 662

### 17.2. Anomalies occurrence and abundance

The main result obtained on LPIS anomaly occurrence was that a very variable number of anomalous reference parcels were detected in the samples (see Table 4). Taking all zones together, around 8 anomalous parcels and 17 anomalies were detected per square kilometre.

For half of the zones, the percentage of anomalous parcels either did not exceed or slightly exceeded 15% of the total number of reference parcels checked. In the other six zones, the observed anomalies occurrence was higher than 25% and in PB1 and PB2 it reached 50%. Although high proportions of anomalous reference parcels were observed, especially for FB and PB, the type of LPIS had no significant effect ( $F(3.8) = 1.33$ ,  $p\text{-value} > .05$ ). The mean anomaly occurrences, expressed as percentages of the anomalous reference parcels sampled, were 18.80 ( $\pm 7.66\text{SD}$ ), 13.70 ( $\pm 12.73\text{SD}$ ), 33.92 ( $\pm 23.66\text{SD}$ ) and 42.46 ( $\pm 28.63\text{SD}$ ) for AP, CP, FB and PB respectively.

$\text{MRA}_i$  and  $\text{TRA}_i$  were the two prevalent classes of anomalies detected during the survey. They were significantly more frequent ( $F(3.44) = 13.16$ ,  $p\text{-value} < .001$ ) and abundant ( $F(3.44) = 14.56$ ,  $p\text{-value} < .001$ ) than  $\text{BRA}_i$  and  $\text{WRA}_i$ . The mean occurrences (%) were 45.12 ( $\pm 21.36\text{SD}$ ), 31.40 ( $\pm 14.33\text{SD}$ ), 18.34 ( $\pm 16.54\text{SD}$ ) and 5.14 ( $\pm 6.04\text{SD}$ ) and the mean abundance (%) was 48.33 ( $\pm 25.04\text{SD}$ ), 28.43 ( $\pm 14.93\text{SD}$ ), 18.69 ( $\pm 18.71\text{SD}$ ) and 4.55 ( $\pm 5.78\text{SD}$ ) for  $\text{MRA}_i$ ,  $\text{TRA}_i$ ,  $\text{BRA}_i$  and  $\text{WRA}_i$  respectively (see Figure 14). Whatever the zone considered, both occurrence and abundance displayed a relatively constant distribution between classes of anomaly. Only zone PB3 showed some variation in the proportions:  $\text{TRA}_i$  and  $\text{MRA}_i$  each accounted for approximately 45% of the occurrences, but  $\text{MRA}_i$  for 70% of the abundance of anomalies. In this case, in a high proportion of the anomalous reference parcels sampled, several buildings belonging to the parcels were often detected, increasing the relative abundance of  $\text{MRA}_i$  in this zone. The main reason was that the reference parcels boundaries were drawn up to constant landscape features, such as asphalt roads, including buildings and facilities on both sides of the road.

The fact that  $\text{MRA}_i$  and  $\text{TRA}_i$  were the two major anomalies suggests that the regulatory area discrepancy tolerance of 1 ha could have been frequently exceeded. Indeed, by nature, the objects at the root of these anomalies (patches of trees, parts of a forest, new buildings and facilities, roads, etc.) often occupy large areas. However, since the area discrepancy was not measured, further work should be carried out to verify this assumption. By contrast,  $\text{BRA}_i$  and, even more so,  $\text{WRA}_i$  appeared rarer anomalies with relatively low abundance. Most of the zones selected presented all four classes of anomalies listed. Only CP2 and CP3 showed no  $\text{BRA}_i$  anomalies and CP2 no  $\text{WRA}_i$  anomalies.

Finally, whatever the category of the anomalies detected, 25.3% ( $\pm 22.6\text{SD}$ ) of the anomalous reference parcels presented a single anomaly, whereas 21.6% ( $\pm 9.26\text{SD}$ ) of the anomalous parcels presented exactly two, 15.7% ( $\pm 10.5\text{SD}$ ) between three and five and 6.46% ( $\pm 12.1\text{SD}$ ) more than five anomalies (see Figure 15). Only zone CP2 presented anomalous reference parcels with just a single anomaly detected. PB3 was the zone with the highest proportion of anomalous parcels, with more than five anomalies per parcel. Because of the large mean reference parcel area (see Table 2) and the high  $\text{MRA}_i$  abundance (see Figure 14), an area discrepancy per reference parcel higher than the 1 ha regulatory tolerance could be expected.

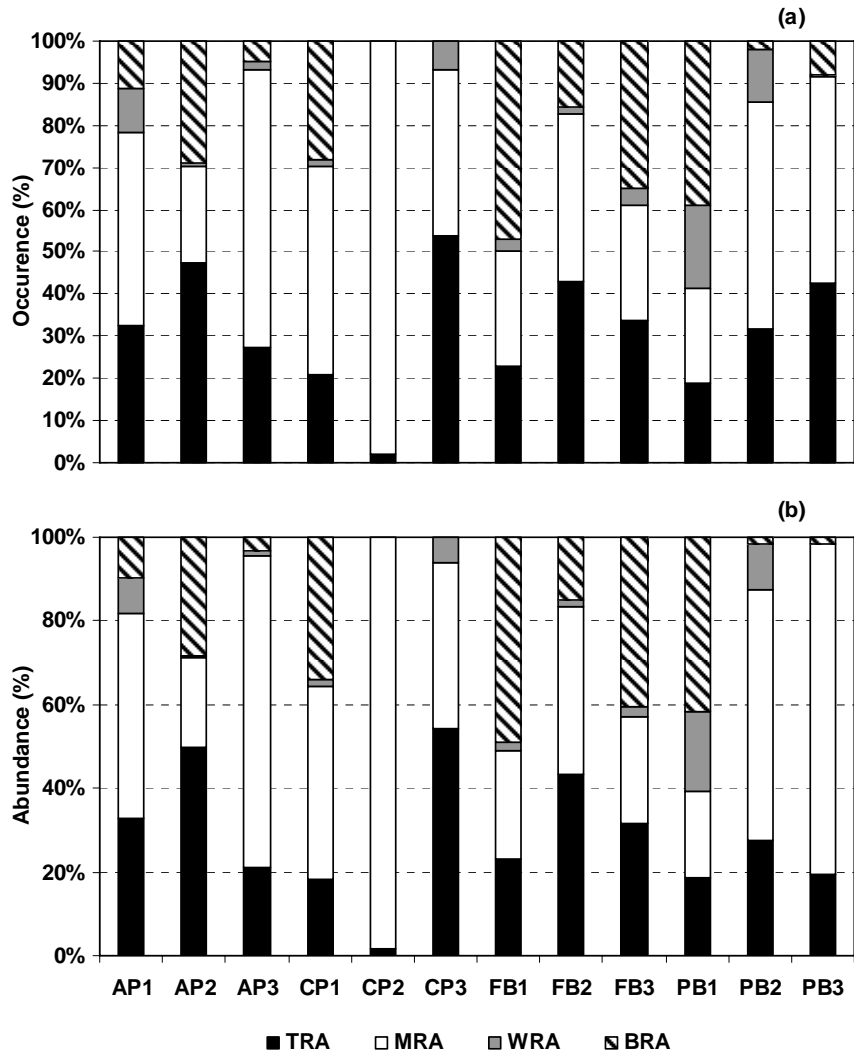
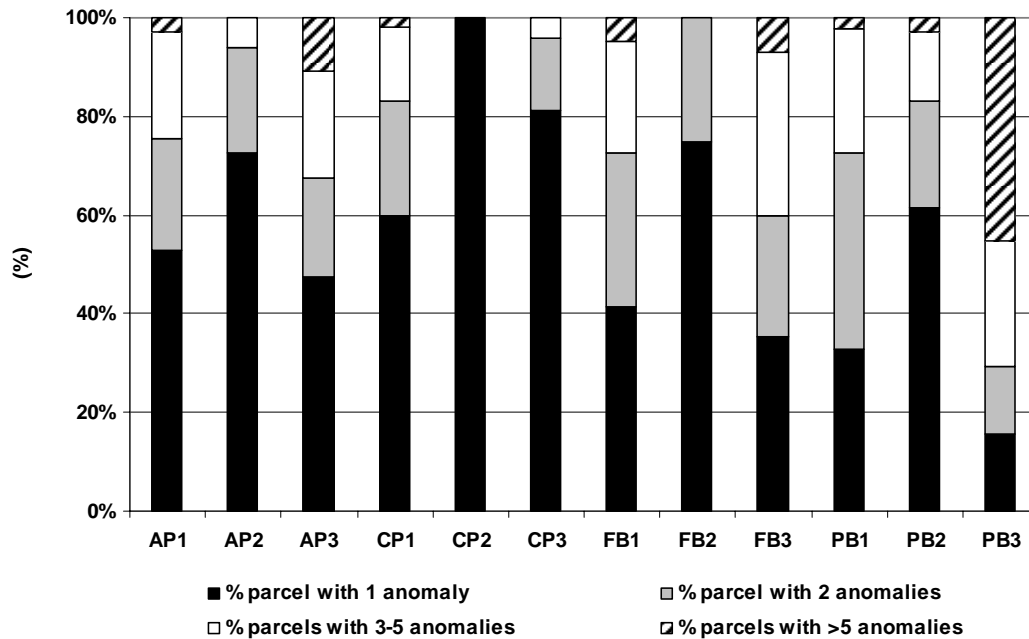


Figure 13: Type-by-type relative percentage of anomalies occurrence and abundance for each of the twelve selected zones

Type-by-type occurrence means the total number of parcels containing TRA anomalies (for instance) over the total number of anomalous parcels; Type-by-type abundance means the total number of TRA anomalies (for instance) divided by the total number of anomalies detected.

TRA = tree-related anomalies (4 categories); MRA = man-made related anomalies (5 categories); WRA = water-related anomalies (4 categories); BRA = boundary-related anomalies (4 categories).

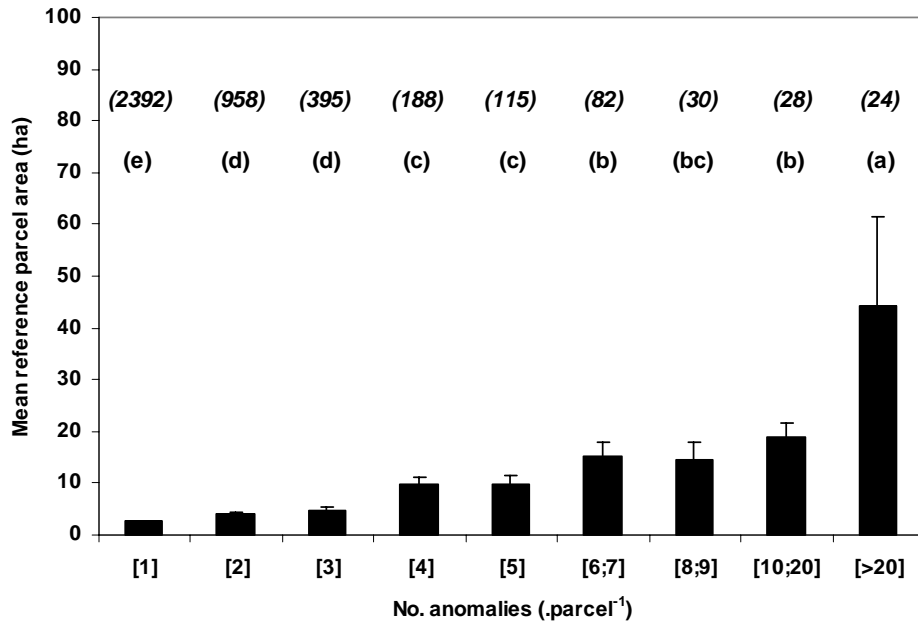


**Figure 14: Distribution of the mean number of anomalies per reference parcel for all twelve selected zones**

### 17.3. Relationship between anomalies abundance and reference parcel area

The last result also suggested a possible relationship between reference parcel area and number of anomalies per reference parcel. To build on the “parcel area” assumption made initially, anomalous reference parcels from all zones were grouped together. Then, eight classes of anomaly abundance by parcel were decided and the individual reference parcels were attributed to the corresponding classes, namely [1], [2], [3], [4], [5], [6-7], [8-9], [10-20] or [ $>20$ ] anomalies.parcel<sup>-1</sup>. This classification made it possible to obtain a sufficient number of reference parcels per class and to avoid empty classes of anomaly abundance. Then, an analysis of variance was performed on the reference parcel area between classes of anomaly abundance. Significant differences were observed between classes of abundance ( $F(1.4211) = 258.90$ ,  $p\text{-value} < 0.001$ ). The number of anomalies per parcel increased in parallel to the mean reference parcel area (see Figure 16).





**Figure 15: Mean reference parcel area for the eight different classes of anomaly abundance observed during the survey**  
(ANOVA: classes of anomalies abundance not connected by the same letter had significantly different mean areas ( $\alpha = 0.05$ ); the values in italics were the number of reference parcels by class of anomaly; the vertical bars are one mean standard error.)

The initial “parcel area” assumption seemed justified. The number of anomalies could partly be correlated with reference parcel area. Because  $TRA_i$  and  $MRA_i$  were the most frequent classes of anomalies detected, it was assumed plausible that landscape elements could be more numerous around and so within reference parcels with a higher area.

However, this result could be questionable. The zone with the highest mean reference parcel area had the highest anomaly occurrence (PB3: 66%). Moreover, four of the six PB and FB zones showed anomaly occurrences higher than 25%. It could therefore be argued that the number of anomalies is indirectly related to the reference parcel area and more certainly a consequence of organisation of land use and land ownership (i.e. LPIS type) within each reference parcel. On the other hand, the survey showed that the number of anomalous parcels and of anomalies detected in FB and PB were no different than the values observed for AP and CP. In fact, at constant sampling size ( $\approx 45\text{km}^2$ ), this study just suggested that no difference existed between LPIS types but, from the point of view of the field inspector, confirmed the necessary differential effort to check the area of reference parcels.

Whatever the origin of this relationship between reference parcel area and anomaly abundance, any future work on LPIS quality should consider this effect. Time-span analysis of LPIS updating efficiency could be elaborated by considering results for distinct classes of reference parcel area.

## **18. Description of anomalies assemblages**

### *18.1. Single anomaly detectability and $\gamma$ -diversity*

Out of the seventeen LPIS anomalies proposed in the anomalies typology, only TRA4 (high density of trees within the reference parcel) was not detected (see Appendix C), despite the fact that TRA<sub>i</sub> was the second highest class of anomalies represented (see Figure 14). This suggested that TRA4 was not an appropriate anomaly category. Most of the tree-related infringements observed were due to continuous tree features such as hedges, woods or part of a forest within the reference parcel. “High number of individual trees relatively constantly distributed over the reference parcel”, as in agro-forestry or in poplar-dominated patches, was never detected as an anomaly. Consequently, deletion of the undetected anomaly category TRA4 could be envisaged.

On the one hand, it is difficult to define the tree density threshold below which a clump of trees is considered TRA4 and above which it becomes TRA1 (patches of trees within the reference parcel). The risk of misinterpretation increases when spring/summer orthophotos with full foliage make it difficult to assess the density. On the other, enforcing animal welfare requirements and maintaining biodiversity as part of CAP cross-compliance could encourage planting of high-density single trees within permanent pastures to provide shade for animals or alongside arable land for biomass production. In these two cases, detection of TRA4 could become suitable to allow agriculture and conservation authorities to ease controls and to decide on the eligibility of agroforestry and biomass production. Consequently, even if TRA4 were not observed, supplementary results should be obtained from specific areas (agroforestry) before deciding whether to delete it. Considering every zone, all sixteen remaining anomalies were detected at least once, but the presence/absence and abundance of each type of anomaly varied highly between zones.

Judging from the last two results, the anomaly typology proposed *ex ante* to address the 2005 LPIS quality assessment was suitable. The  $\gamma$ -diversity of LPIS anomalies observed in 2005 was equal to 16 distinct anomaly types. Finally, the fact that none of the area discrepancies detected was due to any cause other than those proposed in the typology strongly confirmed the representativeness of the anomalies typology, for a wide range of LPIS types and national situations.

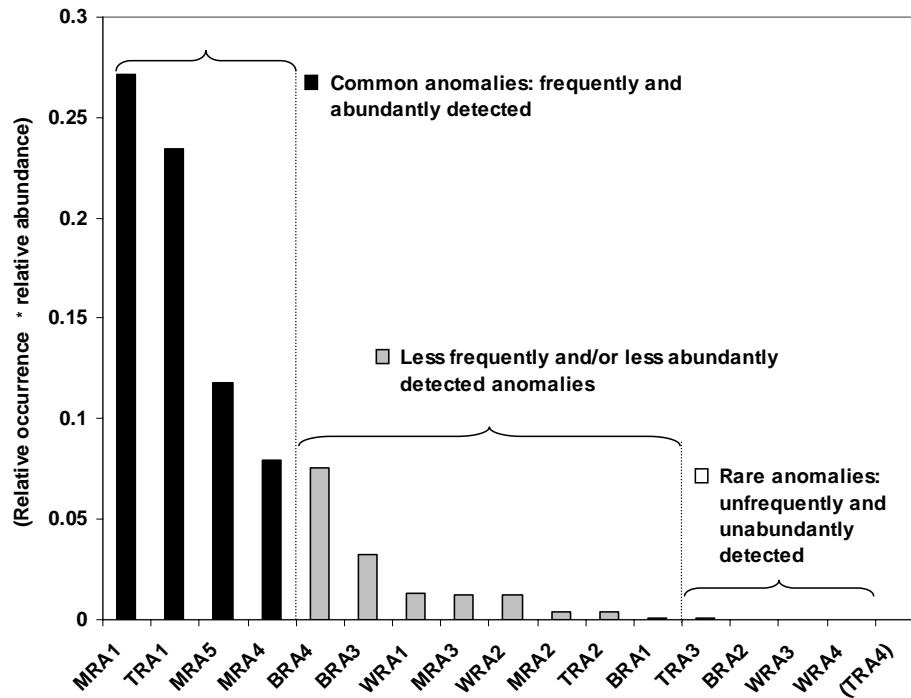
### *18.2. Single anomaly occurrence and abundance*

Out of the 16 different anomalies detected, four were commonly detected in all the zones: TRA1, MRA1, MRA3 and MRA4. With more than 2 000 occurrences each, together TRA1 and MRA1 accounted for more than 50% of all anomalies detected. Together, all four of these “common” anomalies made up approximately 70% of the total abundance.

By contrast, four individual anomalies were detected infrequently and were not counted more than twenty times each: TRA3, WRA3, WRA4 and BRA2 (see Appendix C). These anomalies are therefore called “rare” in the rest of this document.

Between these two groups of “common” and “rare” anomalies, the eight remaining anomalies were (i) abundantly detected in a few zones (BRA4 = 1 685 occurrences in 5 zones) or (ii) non abundantly detected but present in almost all zones (WRA1 = 130 in 11 zones).

Thus, classification of individual anomalies by considering abundance and occurrence separately was not easy. We then decided to calculate the product of relative occurrence (the proportion of the twelve zones displaying a given single anomaly) – relative abundance (the proportion of the abundance of a single anomaly in the total abundance) (see Figure 15).



**Figure 16: Classification of the single anomalies detected during the LPIS survey within the twelve zones studied as product of (relative occurrence \* relative abundance)**  
**(TRA4 is between parentheses as a reminder that this anomaly was not detected during the 2005 survey)**

This classification clearly shows the weight that should be lent to each type of anomaly when seeking to improve LPIS quality and object detection within the LPIS, considering all LPIS types and Member States. Attention could be paid, *inter alia*, to elevation-related segregation of objects appearing as 3D events on the one hand and to reflectance-related segregation of biological versus anthropogenic objects on the other. Together, these two possibilities could help to detect and classify objects within reference parcels and to decide on necessary on-the-spot verification.

However, other anomalies were present, such as BRA4 and BRA3 in 66% and 42% of the zones respectively. Altogether, they accounted for approximately 20% of the total anomalies abundance. The majority of the cases of BRA3 (boundaries not following the reference parcel perimeter) were detected in zones PB1 and FB1, whereas BRA4 (reference parcels overlap resulting in allocation of the same area twice to two contiguous parcels) was detected in CP1, FB1 and FB3 (the last two in the same Member State) (see Appendix C). In this case, because these anomalies were not intrinsically connected to any identified landscape object, it could be assumed that LPIS creation and updating processes were the cause of the anomalies. The delineation of the reference parcel was false and the photointerpreter had failed to establish the correct reference area for a given reference parcel. This could be a consequence of use of an unsuitable imagery sensor or imagery resolution (Pluto-Kossakowska *et al.*, 2007) or even of inexperience on the part of the photointerpreter (Pluto-Kossakowska *et al.*, 2008). The similarities between the anomalies assemblage between zones FB1 and FB3 (richness: 12 anomalies detected and relatively identical abundance for each type of anomaly) reinforced the idea that (i) landscape and/or (ii) LPIS creation and updating influences the final community of anomalies. For these two zones, the landscapes were very similar (almost open fields containing a residual part of a

forest/hedges and a low proportion of urban features). By contrast, zone CP2 (very large open fields without residual forests or hedges and large presence of permanent olive/nut trees) showed a limited number of anomalies ( $n = 170$ ), mainly in the form of MRA1 (agricultural buildings to stock olive tree production and material) with TRA1 virtually absent.

From all these results, the anomalies assemblage appeared to differ highly between Member States. With each zone representing a single Member State (except FB1 and FB3), it was assumed that a specific landscape composition or complexity could be responsible for part of the anomalies assemblage (around 75%) and LPIS creation and updating processes for 16%. Further experiments addressing, first, the relationship between landscape composition and complexity and, second, the efficiency of the LPIS creation and updating workflow should make it possible to determine the cause of LPIS anomalies and, consequently, possible action to take to improve LPIS quality significantly.

## **19. Diversity of LPIS anomalies**

Considering anomalies assemblages zone by zone, different indicators of the diversity of LPIS anomalies were calculated, namely the  $\alpha$ -diversity for each of the zones and the  $\beta$ -diversity (or spatial pattern) between zones, when occurrence- and abundance-based similarity measurements were performed.

### *19.1. $\alpha$ -diversity*

$\alpha$ -diversity involves two distinct concepts: anomalies richness or the maximum number of anomaly types in an assemblage ( $S$ ) and evenness ( $E$ ) or quantification of the unequal dominance of certain types of anomaly in two assemblages with the same richness. Evenness can be seen as a measure of the extent to which the occurrences observed are equitably distributed across anomalies to express any dominance of a few common anomalies within the assemblage.

- *Richness*

Earlier in this study, the estimated anomalies richness ( $S_{est}$ ) was determined when samples were tested for completeness (see Section 16). However, even if close to the observed richness ( $S_{obs}$ ), the estimated maximum richness ( $S_{est}$ ) obtained varied, depending on the method used. The results were consequently not accurate enough to be considered when addressing  $\alpha$ -diversity. For this reason, it was decided to calculate new measurements of  $\alpha$ -diversity (using the Ecological Methodology – Diversity Measures module; Krebs, 1999).

Richness was estimated by calculating the Simpson's reciprocal index (Mac Arthur, 1972):

$$\frac{1}{D} = \frac{1}{\sum p_i^2} \quad (\text{equation 27})$$

where  $1/D =$  the Simpson's reciprocal index of diversity,  
 $p_i =$  the proportion of anomaly  $i$  in the assemblage.

- *Evenness*

The metric index of evenness was decided in accordance with the initial assumption of considering rare and common anomalies equally (Routledge, 1983). Because it is relatively unaffected by rare anomalies, giving the same weight to all anomalies, Camargo's index of evenness ( $E'$ ) (Camargo, 1993) was used (Smith and Wilson, 1996). It is also relatively independent of the richness of the sample:

$$E' = 1 - \left( \sum_{i=1}^S \sum_{j=i+1}^S \left[ \frac{p_i - p_j}{S} \right] \right) \quad (\text{equation 28})$$

where  $E'$  = the Camargo's index of evenness,  
 $p_i$  = the proportion of anomaly  $i$  in the assemblage,  
 $p_j$  = the proportion of anomaly  $j$  in the assemblage,  
 $S$  = the total number of anomalies in the assemblage.

Because several quadrats per zone had anomaly abundances equal to zero, the calculation of  $\alpha$ -diversity indexes for each quadrat sampled was not possible individually. Mean values of Simpson's reciprocal index of richness (heterogeneity) and Camargo's index of evenness were calculated from anomalies assemblages obtained from each of the five blocks per zone (one block containing nine square kilometre quadrats). By doing this, the mean value and variance of the diversity indices from five replicates in a zone were obtained.

Heterogeneity and evenness measurements for each of the zones selected are shown in Table 6.

**Table 6: Non-parametric diversity indices zone by zone ( $S_{\text{obs}}$  = observed richness in samples;  $1/D$  = Simpson's reciprocal index of richness;  $E'$  = Camargo's index of evenness; ANOVA = diversity indices not connected by the same letter had significantly different mean values ( $\alpha = 0.05$ ))**

	$\alpha$ -diversity		
	Richness		Evenness
	$S_{\text{obs}}$	$1/D$	$E'$
<b>AP1</b>	15	4.40 ( $\pm 1.99$ ) (a)	0.50 ( $\pm 0.11$ ) (bcd)
<b>AP2</b>	7	2.83 ( $\pm 0.69$ ) (cd)	0.57 ( $\pm 0.09$ ) (b)
<b>AP3</b>	10	3.31 ( $\pm 0.70$ ) (bcd)	0.48 ( $\pm 0.03$ ) (bcde)
<b>CP1</b>	10	4.00 ( $\pm 0.28$ ) (ab)	0.55 ( $\pm 0.11$ ) (bc)
<b>CP2</b>	5	2.33 ( $\pm 0.51$ ) (d)	0.58 ( $\pm 0.14$ ) (b)
<b>CP3</b>	8	2.51 ( $\pm 0.89$ ) (d)	0.43 ( $\pm 0.08$ ) (cde)
<b>FB1</b>	12	4.24 ( $\pm 0.51$ ) (ab)	0.48 ( $\pm 0.05$ ) (bcde)
<b>FB2</b>	8	3.36 ( $\pm 1.00$ ) (abcd)	0.71 ( $\pm 0.19$ ) (a)
<b>FB3</b>	12	3.72 ( $\pm 0.81$ ) (abc)	0.39 ( $\pm 0.06$ ) (de)
<b>PB1</b>	8	4.03 ( $\pm 0.70$ ) (ab)	0.54 ( $\pm 0.03$ ) (bc)
<b>PB2</b>	8	3.21 ( $\pm 0.29$ ) (bcd)	0.56 ( $\pm 0.05$ ) (b)
<b>PB3</b>	10	2.84 ( $\pm 0.36$ ) (cd)	0.38 ( $\pm 0.08$ ) (de)

Even though the observed richness ( $S_{\text{obs}}$ ) varied between zones, no significant difference was observed ( $F(3.11) = 0.921$  p-value = 0.474) between the four LPIS types. This was confirmed from Simpson's and Camargo's indices; no significant effect of LPIS type was observed for Simpson's and Camargo's indices of diversity ( $(F(3.59) = 1.820$  p-value = 0.154 and  $F(3.59) = 0.179$  p-value < 0.910 respectively). However, significant differences in Simpson's and Camargo's indices of diversity were observed between zones ( $F(11.59) = 3.275$  p-value = 0.0021 and  $F(11.59) = 4.559$  p-value < 0.0001 respectively) (see Table 6).

Some zones (AP1, CP1, FB1 and PB1) were found to be richer (high Simpson's index) than others, whereas AP2, CP2, CP3 and PB3 presented low richness (heterogeneity) values. In these cases, low heterogeneity, did not signify a high dominance of one of the anomalies. On the contrary, zones CP3 and PB3 showed a small number of anomalies with a relatively well

balanced abundance between them, i.e. a small evenness value. On the other hand, heterogeneous zones such as AP1, CP1, FB1 and PB1 showed medium values of evenness. Only zone FB2 was strongly dominated by one category of anomaly (TRA1, see Appendix C), despite the fact that it was a medium heterogeneous zone.

The previous result showed no clear trend between zones and that use of  $\alpha$ -diversity metrics was relatively difficult to interpret because of the variability of the measurement obtained. Zones with medium heterogeneity could have low (FB3) or high (FB2) evenness and homogeneous zones showed independently low (CP3 and PB3) or medium (CP2 and AP2) evenness. Although informative, use of single  $\alpha$ -diversity metrics to distinguish LPIS zones appeared unsuitable.

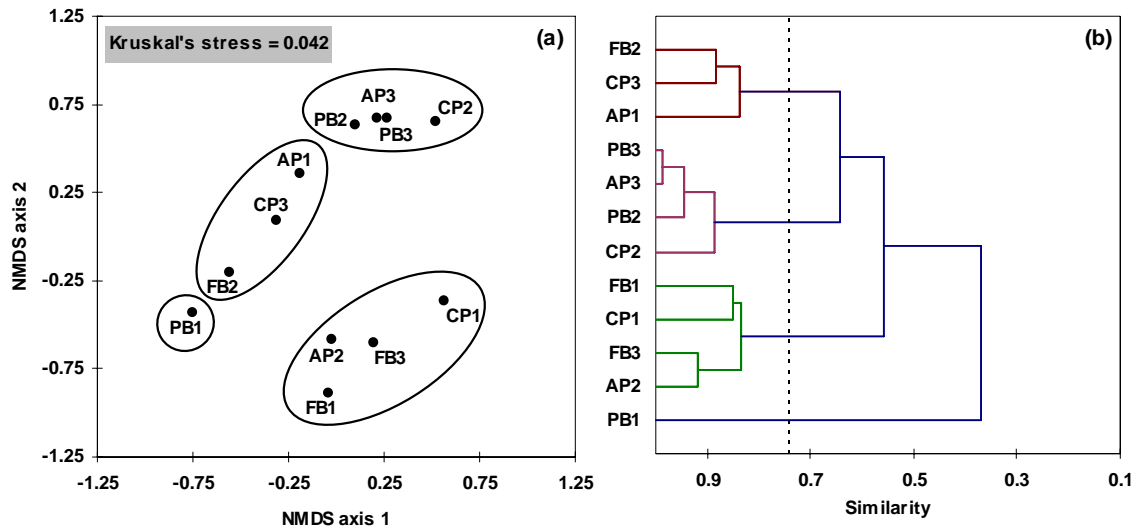
These results also confirmed previous conclusions suggesting that classification of zones by LPIS type was no longer suitable when addressing LPIS anomalies assemblages. Driving forces other than LPIS type should be considered to explain the anomalies assemblage observed in each zone.

### 19.2. *$\beta$ -diversity*

To allow comparison and classification of LPIS zones, an analysis of  $\beta$ -diversity was performed, by considering the similarity between pairs of zones. This took the form of ordering the twelve zones by two-stage NMDS into occurrence- and abundance-related similarity measures. Jaccard's (Romesburg, 1984) and Morisita's (Krebs, 1999) measures of similarity were calculated between each pair of zones with Estimate-S software (Colwell, 2004). In parallel, HAC (Spearman's correlation coefficient of similarity; average linkage) was performed to facilitate visualisation of the clusters obtained from NMDS.

Because the robustness of the ordination of Jaccard's index of similarity was insufficient (three NMDS axes and Kruskal's stress = 0.246), the clusters were not considered. Unlike Jaccard's result, ordination from Morisita's index of similarity provided a very good ranking of the zones between four different clusters (see Figure 17).  $\beta$ -diversity was therefore interpreted only from Morisita's index of similarity (three NMDS axes and Kruskal's stress = 0.042).

Considering all pairs of zones, the mean similarity value was 0.62 ( $\pm$  0.21SD). However, the Morisita's similarity index observed was between 0.15 (CP2/PB1), indicating wide dissimilarity between the two zones and almost total similarity between AP3 and PB3 (similarity = 0.99). These extreme similarity values corresponded to the longest and shortest distances between NMDS coordinates in respectively (see Figure 17-a).



**Figure 17: (a) NMDS ordination and (b) hierarchic ascendant classification of the anomalies assemblages obtained by zone from Morisita's index of similarity (HAC: Spearman's correlation coefficient/average linkage/4 classes expected; NMDS using SMACOF (Scaling by MAjorizing a CONVex Function) algorithm from Leuw, 1977)**

From the NMDS ordination, without the help of the HAC result, three different clusters could be considered: AP2 and the three neighbouring zones at the bottom of the chart; PB1 and FB2; and a single cluster with all the remaining zones. HAC, on the other hand, proposed four different clusters. Verification of the NMDS clusters displayed in a 3D chart clearly separated PB1 from FB2, then FB2, CP3 and AP1 from the rest of the four zones in the upper part of the chart; consequently, four clusters were obtained, corresponding to the four ranked by HAC. To facilitate visualisation on the 2D chart, clusters were then circled in the NMDS chart.

Together with the HAC results, NMDS made it possible to classify the zones into four clusters for which the cause of the dissimilarity between clusters was analysed.

### 19.3. *Anomalies clusters and spatial pattern*

Considering  $\alpha$ -diversity metrics (see Table 6), none of the clusters classified from NMDS presented any particular organisation of the  $\alpha$ -diversity metrics. All the clusters presented very similar mean heterogeneity and evenness values. Consequently, the cluster origin was ascertained for the anomalies assemblages of the zones in each cluster.

Anomalies assemblages (see Appendix C) were grouped on the basis of clusters obtained from NMDS. For each zone embedded within a cluster, proportion for each of the sixteen anomaly categories was calculated; then, a mean proportion (and standard deviation) for each of the sixteen anomaly categories was calculated per cluster and expressed as a percentage. The resulting mean percentages of the sixteen anomaly categories per cluster are set out in Table 7.

**Table 7: Mean anomaly abundances (%) for each anomaly categories per cluster**

(Numbers between parenthesis are standard deviation – Cluster 4 being composed of one single zone, standard deviation was not calculable) (Dark cells correspond to the first-order anomalies in a cluster ( $\geq 30\%$ ); dark grey cells to second-order anomalies ( $15\% \leq < 30\%$ ) and light grey cells to third-order anomalies ( $10\% \leq < 15\%$ ). Values below 10% are shown in smaller font size.)

Single anomalies	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	AP2, FB1, FB3, CP1	AP3, PB2, PB3, CP2	AP1, CP3, FB2	PB1
TRA1	41.60 ( $\pm 10.20$ )	29.55 ( $\pm 12.58$ )	16.85 ( $\pm 9.46$ )	18.69
TRA2	1.61 ( $\pm 1.55$ )	1.00 ( $\pm 1.24$ )	0.41 ( $\pm 0.70$ )	0
TRA3	0.29 ( $\pm 0.41$ )	0.02 ( $\pm 0.04$ )	0.18 ( $\pm 0.31$ )	0
MRA1	17.50 ( $\pm 7.70$ )	15.28 ( $\pm 4.37$ )	49.85 ( $\pm 5.98$ )	4.33
MRA2	4.35 ( $\pm 4.19$ )	0.63 ( $\pm 0.69$ )	1.35 ( $\pm 2.02$ )	0
MRA3	5.52 ( $\pm 1.88$ )	0.82 ( $\pm 0.75$ )	3.26 ( $\pm 4.60$ )	0.87
MRA4	2.64 ( $\pm 1.91$ )	5.39 ( $\pm 5.62$ )	6.91 ( $\pm 3.64$ )	11.42
MRA5	12.83 ( $\pm 0.94$ )	7.69 ( $\pm 0.75$ )	16.53 ( $\pm 4.75$ )	3.98
WRA1	3.99 ( $\pm 1.69$ )	1.13 ( $\pm 0.65$ )	1.21 ( $\pm 1.79$ )	1.90
WRA2	0.84 ( $\pm 0.61$ )	0.57 ( $\pm 0.35$ )	1.82 ( $\pm 2.70$ )	17.30
WRA3	0.58 ( $\pm 0.82$ )	0	0	0
WRA4	0.07 ( $\pm 0.10$ )	0	0	0
BRA1	1.45 ( $\pm 2.05$ )	0.09 ( $\pm 0.09$ )	0	0
BRA2	0.58 ( $\pm 0.82$ )	0.05 ( $\pm 0.08$ )	0	0
BRA3	6.16 ( $\pm 6.41$ )	5.55 ( $\pm 7.99$ )	1.60 ( $\pm 1.10$ )	41.52
BRA4	0	32.23 ( $\pm 3.66$ )	0.04 ( $\pm 0.07$ )	0

Each of the clusters showed a particular anomalies assemblage, in which at least one single anomaly prevailed. In every case, these first-order anomalies accounted for more than 30% of the cluster's total abundance; first-order anomalies were never the same from one cluster to another. These first-order anomalies were TRA1, BRA4, MRA1 and BRA3 respectively for clusters 1 to 4. For cluster 1, only one second-order anomaly was identified: MRA1 accounted for more than 20% of the total abundance. For the remaining clusters, there were two second-order anomalies: TRA1 and MRA1 for cluster 2, TRA1 and MRA5 for cluster 3 and, finally, TRA1 and WRA2 for cluster 4.

Some of the anomalies were classified as first- or second-order anomalies for only one cluster each: BRA4 for cluster 2, BRA3 for cluster 4, MRA5 for cluster 3 and WRA2 for cluster 4. Because they represented landscape features at the root of area discrepancies commonly detected within each cluster, this table strongly suggests that LPIS anomalies assemblage could be inextricably linked to landscape organisation and fragmentation around the reference parcels.



- Cluster 1 consisted of zones where “patches of trees” (TRA1) prevailed, followed by “buildings” (MRA1). This composition accurately describes composite rural landscapes with a mix of urban (or anthropogenic) and natural elements (forests and woods). In practice, the zones in this cluster were urban zones within residual forests corresponding to medium-moisture continental or Mediterranean climates. Another interesting finding was that zones FB1 and FB3, from two different regions of the same Member State (with two different contractors in charge of the LPIS), were considered as belonging to the same cluster. This reinforced the idea that the LPIS anomalies assemblage and LPIS quality improvements could be more landscape composition and fragmentation driven than LPIS creation and updating driven.
- Cluster 2 consisted mainly of zones where the highest proportion of anomalies was caused by “overlapping boundaries of contiguous parcels” (BRA4). It was the sole cluster presenting a huge proportion of this single anomaly. The second-order anomalies were then TRA1 and MRA1. From this it was pointed out that the landscapes were relatively similar to those described for cluster 1: apart from CP2, which was under the influence of a dry Mediterranean climate, all the zones benefited from the influence of oceanic conditions, with high annual water availability favouring biomass production, and were on the western side of Europe. The fact that zones PB2 and PB3 were grouped in the same cluster also indicated a possible link between BRA4 anomaly category and physical blocks LPIS.

LPIS quality within this cluster could first be addressed by analysing image acquisition, processing and interpretation at the level of the contractors before addressing the landscape structure effect.

- In cluster 3 the segregation was mainly due to the first-order anomaly MRA1 “Buildings”. However, a large share of MRA5 (“farm within reference parcel”) suggested that the main problem with LPIS quality here could be due to the recent development of urban and agricultural anthropogenic elements. Because the Member States concerned were from Eastern Europe, it was assumed that the recent economic expansion and restructuring of Agriculture (and the agricultural landscape) since EU accession were responsible for the construction of numerous new facilities and buildings, to the detriment of agricultural and natural land which, up until now, had been closed to urbanisation. Urban and agricultural extension in these zones has induced a high proportion of man-made related anomalies. Furthermore, as urban extension is continuing, more and more changes of land use could be later responsible for the numerous tree-related anomalies detected in these zones (TRA1 as second-order anomaly).

The situation in this cluster suggested that LPIS creation and especially LPIS updating were crucial to match the continuous land-use changes and to keep pace with the expansion and development of the related Member States. Protection of natural landscape features or of zones of high nature value should also be considered directly by general application of LPIS to identify and trace protected areas or areas of interest precisely.

- Finally, cluster 4 consisted of a single zone. PB1 was the sole cluster presenting BRA3 (“boundary not following the landscape features”) and WRA2 (“ditches within the reference parcels”). This result reflected a highly complex landscape, where aquatic and forest features were widely present. At the same time, the medium proportion of TRA1 (“patch of the trees) and the low proportion of MRA1 (“buildings”) suggested that zone PB1 is in a Member State with a low population density and a limited number of open-field situations. In fact, the zone is in a Scandinavian Member State where continuous

vegetation and water networks are omnipresent. The dispersion index for PB1 (result not shown) was the highest, suggesting the highest clustering of the anomalies around landscape features, in places where human beings have decided to deforest, to drain off and cultivate land.

Concerning cluster 4, landscape structure and fragmentation would be of primary importance if it were to be decided to improve LPIS quality.

## ***PART F. Final discussion and prospects***

### **20. Discussion of the results**

A wealth of information was obtained from the survey addressing the diversity in LPIS anomalies. Some of it just confirmed previously accepted aspects of LPIS quality, while most highlighted the advantages of recourse to ecological methodology to improve regulatory and technical systems such as the LPIS.

- ***The sampling method***

Adaptation of an ecological sampling method for obtaining the anomalies assemblages within the LPIS was guided by the specific constraints of on-the-spot check (remote sensing AND rapid field visit) within CAP as defined by European regulation EC N° 796/2004. This study has shown that, despite the absence of any reference to LPIS anomaly occurrence in the literature, taking all the zones together, the probabilistic approach to sampling size (see Section 12) gave an approximate of the minimum number of reference parcels to sample. However, this paper has demonstrated that this minimum number of reference parcels would not be realistic when applying the mean reference parcel area in a given zone. In the case of zone PB3, the total area to sample corresponding to 367 reference parcels would then be 121 km<sup>2</sup> (see Section 14). At a constant number of reference parcels and time investment, this called for display of each reference parcel on screen, losing visibility in the case of very large reference parcels and increasing the risk of underestimation of anomalies abundance.

Based on the LD and Gompertz models of anomalies accumulation curves, the decision was taken to opt for the second highest sampling area (FB1: 22.87 km<sup>2</sup>), which allowed detection of at least 95% of the richness of each of the twelve zones. By contrast, the Clench model indicated that mean values of 28 km<sup>2</sup> and 59 km<sup>2</sup> were necessary to reach 90% (E<sub>90%</sub>) and 95% (E<sub>95%</sub>) respectively of the estimated richness in all the zones (see Section 18). In that case, a 23 km<sup>2</sup> sample size, as obtained from PB3, would have been inadequate.

However, the decision to double the second highest minimum area to be sampled and to apply a 45 km<sup>2</sup> sample size to all the zones was good. This decision was based on the time and human resources available and allowed sampling, in almost all situations, of 90% of the maximum anomalies richness. Depending on the estimators considered, only zones FB2 and AP2 (E<sub>45</sub>) showed a richness value slightly lower than the 90% initially expected.

*To conclude on sample size, a constant sampling effort of 45 km<sup>2</sup> was necessary and sufficient to obtain good completeness of the LPIS anomalies assemblages, considering all zones (and LPIS types).*

Concerning quadrat size, Wiegert's method clearly identified a medium-sized quadrat of 1 km<sup>2</sup> as efficient for complete anomalies assemblages. It was the best compromise for achieving the lowest variance at the lowest cost. This was also confirmed by the smaller relative bias and imprecision of the mean obtained for anomaly occurrence. Moreover, by using a 1 km<sup>2</sup> quadrat size, the number of quadrats available provided coverage of a sufficient diversity of sub-zones of interest within the image. This result was also valid whatever the sampling designs assessed.

*When applying systematic, systematic cluster or random sampling designs, a quadrat size of 1 km<sup>2</sup> should be used with a sample size of 45 km<sup>2</sup>.*

All three sampling designs tested achieved more than 90% completeness compared with the true anomalies assemblage for the pre-test zone (FB2). Consequently, all three appeared suitable for

sampling anomalies in the LPIS. But, because the choice of the sampling design cannot be considered independent of the occurrences to detect, the clumped distribution of anomalies in the image suggested that the systematic cluster design was the most suitable.

As regards the correlation between the spatial pattern of anomalies and landscape structure, other sampling designs, such as “stratified random” and “adaptive” (also called “response-adaptive”) sampling, could possibly be envisaged in due course. These sampling methods would permit sampling of anomalies from CwRS images by, respectively, (1) applying a higher sampling effort to the sub-parts of a landscape perceived as more complex within the image or (2) using anomalies detection outcomes as they become available to adjust the assignment of future quadrats to zones suspected of containing expected anomalies (Thompson, 1992).

Box 1: Conclusions about the sampling method

- CwRS imagery appeared useful and precise enough to address LPIS quality.
- Cost-effective sampling for LPIS anomalies detection within CwRS imagery requires a square quadrat of 1 km<sup>2</sup>.
- A sampling effort of 45 km<sup>2</sup> is the minimum required to cover at least 90% of the richness of the LPIS anomalies assemblages, whatever the LPIS type and Member State considered.
- If random sampling without replacement is the design which should usually be set up whenever possible, systematic cluster (or stratified and adaptive) sampling design should be envisaged to match the effect of landscape structure on anomalies assemblage characteristics.
- Reference parcels-to-quadrat attribution rules are an important aspect of sampling to take into account before conducting the experiment.

• Validation of the sampling method

Next, the sampling method was calibrated on one particular pre-test zone for which (1) LPIS quality was *a priori* considered good (because of the suspected limited number of anomalous reference parcels) and (2) the landscape structure was considered representative of the diversity of the situations found in the EU.

Use of anomalies accumulation curves and of non-parametric estimators (see Section 18) to verify the adequacy of the sampling corroborated the choice of sampling method. Considered individually, these two methods gave relatively constant and similar estimates of the anomalies richness in the zones selected. However, while the choice of model appeared crucial for the former, the choice of the non-parametric estimators regarding the nature of the data available (occurrence vs. abundance) highly conditioned the estimated anomaly richness within the zone.

*Judging from the results of this study, the linear dependence model was not suitable to fit the observed anomalies-area relationships; by contrast, the Clench and Gompertz models were more precise.*

Concerning non-parametric estimators, because anomalies abundance measurements were available, the *Chao1* and *Chao2* estimators should attract the greatest interest. Colwell and Coddington (1994) suggested that *Chao2* can be perceived as the least biased estimate of species richness for small numbers of samples.

*Consequently, it can be assumed a posteriori that the sampling method was adequate (at least 90% of the maximum estimated richness was sampled) for eleven of the twelve zones. Only AP2*

was sampled at lower completeness but the richness obtained stayed relatively close (87.7%) to the maximum estimated richness.

**Box 2: Conclusions about validation of the sampling method**

- Verification of the completeness of the sampling method requires use of both anomalies accumulation curves and non-parametric estimators.
- Accumulation curves require selection of several models to compare the estimates and finally decide on the adequacy of the sampling method.
- Non-parametric estimators could be used to cross-verify the constancy of the outliers identified by the accumulation curve method.
- If used alone, the choice of non-parametric estimators should be based on the type of data available; with occurrence data *Jack1 & 2* could be favoured, whereas *Chao1 & 2* would be used with abundance data. Other estimators, such as the bootstrap procedure or ICE (Colwell, 2004), could also be envisaged.

**N.B.:** Whatever the validation method, only estimates are obtained, never the true anomalies richness.

• LPIS quality in the EU

All the zones and, consequently, all the Member States surveyed presented a high proportion of anomalous reference parcels (see Table 5). However, because the minimum area set for anomalies was 0.1 ha, not all the zones exceeded the regulatory tolerance (1 ha over-declaration tolerance by reference parcels). Second, the CwRS zones used during this survey had been identified by the dedicated LPIS risk analysis as zones at risk. It would therefore be normal to find a high proportion of anomalous reference parcels in these zones, whereas the rest might be less contaminated, i.e. with LPIS quality matching the regulatory requirements.

However, this study showed that half of the zones presented a proportion of anomalous reference parcels higher than 25% and detected a high proportion of tree- and man-made related anomalies (see Figure 14), suggesting a high probability of some zones exceeding the area discrepancy tolerance set.

To decide finally on the real LPIS quality in Europe, area discrepancies corresponding to the anomalies should be measured in the future to statute onto the proportion of anomalous parcels to be considered as non-compliant.

**Box 3: Overall European LPIS quality**

- The proportion of anomalous reference parcels was higher than initially expected. However, this does not mean that LPIS quality does not comply with the regulatory requirements.
- Although analysis of the diversity in LPIS anomalies was the prime objective, further measurement of the area discrepancies detected during this survey should determine the real frequency and magnitude of irregularities within the LPIS.

- The diversity of LPIS anomalies

The typology of LPIS anomalies decided *a priori* precisely matched the anomalous situations found. No area discrepancy ever called for creation of a new anomaly category.

In all the zones surveyed, tree-related (TRA<sub>i</sub>) and man-related (MRA<sub>i</sub>) anomalies were the major anomalies detected. These two classes generally accounted for more than 50% of the total anomalies abundance detected and were common to all the zones. Other classes of anomalies such as boundary-related (BRA<sub>i</sub>) and, even more so, water-related anomalies (WRA<sub>i</sub>) were rarer and specific to certain zones (see Figure 15).

Moreover, within classes of anomalies, each category was not equally detected. For example, in the case of class TRA<sub>i</sub>, TRA1 (patch of trees in the reference parcel) appeared very common, while MRA1 (buildings), MRA4 (asphalt road) and MRA5 (farm features around the buildings) were very commonly detected within class MRA<sub>i</sub>. This suggested that natural landscape features surrounding the reference parcels (forests, hedges, etc.) and continuous anthropogenic landscape features were the two main sources of anomalies detected. Taking them into account when addressing LPIS quality updating should correct the majority of the potential irregularities existing within the LPIS (Oesterle and Hahn, 2004). This also calls for continuous integration within the LPIS of georeferenced information concerning land-use planning and management.

Box 4: Overall European LPIS quality

- The typology of LPIS anomalies proposed *a priori* to assess LPIS quality in 2005 was validated.
- By focusing on tree-related and man-made related anomalies, every LPIS manager should address and correct more than 50% of the existing anomalies within the LPIS.
- Although initially focusing on solely agricultural concerns, LPIS quality improvements must integrate relevant information on other land uses, such as non-agricultural/natural elements and anthropogenic features.

- Typology of LPIS anomalies and causes of the anomalies assemblage

Throughout this survey, the effect of the well accepted LPIS typology on the composition of anomalies assemblages was analysed. No clear, significant effect was found. At class of anomalies or single anomaly category level, neither anomaly occurrence nor anomaly abundance showed any significant difference between LPIS types. Furthermore, within each LPIS type, this study showed that anomalies diversity was highly variable between zones (see Section 21), limiting the possibility of differentiating between LPIS types.

By contrast, the anomalies assemblages of the zones were clearly and significantly grouped together based on the abundance of the anomaly categories they contained. Since  $\alpha$ -diversity metrics (and occurrence-based  $\beta$ -diversity metrics/Jaccard's similarity index) were difficult to interpret and were not performed significantly, these zones were clustered by using ordination methods based on abundance-based  $\beta$ -diversity metrics (Morisita's index of similarity). This provided a means of depicting the spatial pattern of possible LPIS anomalies.

Whatever the ordination method used, the clusters of zones obtained were constant and correctly differentiated (see Figure 16). From this, construction of the cluster-by-cluster relative proportion (see Table 7) clearly showed that the identity of each cluster was a consequence of the presence (absence) and abundance of the more common and of some of the less frequent (abundant)

anomalies detected. The rare anomalies were never involved in the segregation of the clusters. Distinction between clusters was mainly based on the high proportion of one particular anomaly category: TRA1 for cluster 1, BRA4 for cluster 2, MRA1 for cluster 3 and BRA3 for cluster 4. Together with the clumped spatial distribution of the anomalies (see Section 16), this reinforced the idea of the need to incorporate landscape characteristics in order appropriately to review and update the overall LPIS quality. This also suggested that some of the anomalies for clusters 2 and 4 were strongly related to the GIS-assisted and/or d-GPS delineation of reference parcels during LPIS creation and updating.

Accordingly, the landscape characteristics of a CwRS zone should be considered pertinent indicators of the anomalies composition and of the dominant anomalies which can be expected. The LPIS or at least the CwRS risk analysis could then be updated by considering landscape properties as a surrogate of LPIS quality. By doing this, national LPIS contractors would be able to identify, faster and more accurately, technical and political improvements to decision-making all along the LPIS management chain.

#### Box 5: Overall European LPIS quality

- Although the official LPIS typology is a useful indication of the administrative elements to be considered in CwRS on-the-spot (OTS) checks, it is never a characteristic conditioning LPIS quality.
- LPIS quality seems highly correlated with the nature and the structure of the landscape in which the reference parcels are established.
- In certain circumstances, LPIS quality could also be a consequence of misuse of GIS/d-GPS technologies when delineating reference parcels.
- Every process for improvement of LPIS quality should (1) integrate landscape structure and fragmentation characteristics and (2) assess the GIS/d-GPS based on procedures for reference parcels delineation in order to be effective.

## **21. Prospects**

### • The LPIS updating process

According to the results outlined above, there could be two main causes for anomalies observed within the LPIS: (1) non-exclusion of landscape features embedded within reference parcels and consequently considered ineligible; (2) errors during delineation of reference parcels from LPIS imagery by national LPIS contractors or farmers. Of course, both could be efficiently addressed during the CAPI (computer-assisted photo-interpretation) process and consideration could be given to reviewing all these anomalies *ex post* during the CAPI step. For logistical and economic reasons, this appears unrealistic. The solution could be to consider them separately:

- at the image processing stage, where correct conversion of the original raster into vector classes of landscape features (to address TRA<sub>i</sub>, MRA<sub>i</sub> and WRA<sub>i</sub>) should help to (i) directly identify potential ineligible objects within reference parcels and (ii) indirectly delineate the zones at risk to search for LPIS anomalies or to conduct CwRS risk analyses;
- then, in the case of BRA<sub>i</sub> anomalies, at the CAPI stage by analysing the critical points within the LPIS updating chain, such as image processing, reference parcels boundary delineation and the procedure for correction of out-of-date boundaries.

The initial statements made before deciding the methodology to apply in this survey commonly acknowledged that the anomalies detection from CwRS images covered only some of the possible aspects of LPIS quality which can be considered. Other critical points in the LPIS creation and updating chain should be analysed at the same time:

- The data acquisition stage: when testing different imagery with different levels of resolution and visibility, Pluto-Kossakowska *et al.* (2007) showed that some of the area measurement discrepancies observed are largely explained by use of imagery which is not accurate enough.
- Depending on the imagery used, the CAPI operators have a non-negligible effect on the area discrepancy: some operators achieved higher precision with delineating reference parcels boundaries using traditional imagery (orthophotos) than when using recent radar imagery. This could be related to their personal level of experience. To describe this potential effect, compliance with any internal and regulatory procedures which exist would have to be assessed.
- Among the possible solutions to update the LPIS, farmers' annual declaration of reference parcels boundaries is the most promising. It should provide 100% up-to-date information every year to the LPIS contractors and greatly limit the risk of incorrect delineation. In this case, consideration should be given to assessment of d-GPS users' procedures and d-GPS device accreditation. A specific effort should be made on the conditions of use and the robustness of the measurements obtained.

- *The landscape ecological approach*

In the near future the authors propose to address the following complementary aspects of the method described above:

- Landscape ecology approach: by analysing the landscape structure and fragmentation characteristics of the zones used during this survey, the authors believe that it will be possible to identify the main relationships between LPIS anomalies assemblages and landscape characteristics. This could be done simply by: (1) establishing the landscape classification from the images in terms of the major anomalies detected (forests, roads, urban zones, water networks, etc.), (2) measuring the landscape pattern and structure metrics and (3) conducting multivariate analyses of landscape factors and anomalies assemblages.

This approach is highly complementary to automatic detection of objects by remote sensing. Whereas automatic detection focuses on detection of anomalous situations by intersecting layers containing each single (2/3D) object existing within the landscape and LPIS reference parcels to flag the anomalous situation (Oesterle and Hahn, 2004; Zielinski, 2009), the landscape ecological approach focuses on identification of the zones at risk (potentially anomalous). Consequently, the former can be seen as a curative system to update the LPIS promptly, while the latter can be perceived as an investigative system designating the zones to cure, as an instrument to perform the CwRS risk analysis.

- In a second stage, descriptors of the LPIS updating procedures performed in each of the twelve zones selected should be incorporated. Documents and databases such as the LPIS census conducted by Milenov in 2007 would be of primary importance. Supplementary inquiries addressed to the Member States should also allow more accurate identification of the principal critical points along the LPIS management chain responsible for part of the anomalies detected.
- Finally, measuring the area discrepancies observed during the survey would make it possible to determine the real LPIS quality by substituting regulatory irregularities to anomalies.



If accepted, these proposals would undoubtedly provide a cost-effective means of adjusting the quality of the LPIS to a level at which its general application to address any agri-environmental problem would be undisputed. Powerful ICT instruments, numerous accurate satellite and airborne images and effective computer-assisted systems are already available to propose rapidly to all land managers and decision-makers the geo-referenced land-use structure needed, with the Land Parcel Identification System at its core.

## BIBLIOGRAPHY

- Arrhenius, O. (1921). Species and area. *Journal of Ecology* 9:95-99
- Bennet, H., Osterburg, B., Nitsch, H., Kristensen, L., Primdhal, J. and Vershuur, G. (2006). Strengths and weaknesses of cross-compliance in the CAP. *Eurochoices* 5(2): 50-57
- Brandt, J., Primdhal, J. and Reenberg, A. (1999). Rural land-use and dynamic forces – analysis of driving forces in space and time. In Krönert, R., Baudry, J., Bowler, I.R. and Reenberg, A. (eds). *Land-use changes and their environmental impact in rural areas in Europe*. UNESCO, Paris, France, pp 81-102
- Brose, U., Martinez, N.D. and Williams, R.J. (2003). Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* 84(9): 2364-2377
- Buckwell, A. (2007). The next steps in CAP reform. *EuroChoices* 6(2): 13-19
- Burnham, K.P. and Overton, W.S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60: 927-936
- Camargo, J.A. (1993). Must dominance increase with the number of subordinate species in competitive interactions? *Journal of Theoretical Biology* 161:537-542
- Chao, A. (1984). Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.*, 11: 265-270
- Chao, A., Chazdon, R.L., Colwell, R.K. and Shen, T.J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, 8: 148–159
- Cheng, C.C. (2004). Statistical approaches on discriminating spatial variation of species diversity. *Bot. Bull. Acad. Sin.* 45: 339-346
- Christman, M.C. (2003). Adaptive two-stage one-per-stratum sampling. *Environ. Ecol. Stat.* 10: 43-60
- Clarke, K.R. and Warwick, R.M. (1994). *Change in marine communities: An approach to statistical analysis and interpretation*. 1st edition: Plymouth Marine Laboratory, Plymouth, UK, 144 pp
- Cochran, W.G. (1977). *Sampling techniques*. 3rd edition: John Wiley and Sons, NY
- Coleman, B.D., Mares, M.A., Wolling, M.R. and Hsieh, Y.H. (1982). Randomness, area and species richness. *Ecology* 63(4): 1121-1133
- Colwell, R.K. (2004). ESTIMATES: Statistical estimation of species richness and shared species from samples, Version 8.0. Available at: <http://viceroy.eeb.uconn.edu/estimates>. Persistent URL <http://purl.oclc.org/estimates>
- Colwell, R.K. and Coddington, J.A. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* 345: 101-118
- Commission Regulation (EC) No 796/2004 of 21 April 2004 laying down detailed rules for the implementation of cross-compliance, modulation and the integrated administration and control system provided for in Council Regulation (EC) No 1782/2003 establishing common rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers:  
(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:141:0018:01:EN:HTML>)

Commission Regulation (EC) No 817/2004 of 29 April 2004 laying down detailed rules for the application of Council Regulation (EC) No 1257/1999 on support for rural development from the European Agricultural Guidance and Guarantee Fund (EAGGF):

(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:153:0030:0081:EN:PDF>)

Commission Regulation (EC) No 972/2007 of 20 August 2007 amending Regulation (EC) No 796/2004 laying down detailed rules for the implementation of cross-compliance, modulation and the integrated administration and control system provided for in Council Regulation (EC) No 1782/2003 establishing common rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers:

(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:216:0003:0009:EN:PDF>)

Council Regulation (EC) No 1257/1999 of 17 May 1999 on support for rural development from the European Agricultural Guidance and Guarantee Fund (EAGGF) and amending and repealing certain Regulations:

(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1999:160:0080:0102:EN:PDF>)

Council Regulation (EC) No 1782/2003 of 29 September 2003 establishing common rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers and amending Regulations (EEC) No 2019/93, (EC) No 1452/2001, (EC) No 1453/2001, (EC) No 1454/2001, (EC) No 1868/94, (EC) No 1251/1999, (EC) No 1254/1999, (EC) No 1673/2000, (EEC) No 2358/71 and (EC) No 2529/2001:

(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:270:0001:0069:EN:PDF>)

Court of Auditors (2005). Official Journal of the European Union – Annual reports on the implementation of the budget – Chapter 5: The common agricultural policy. (2006/C 263/01): 77-115:

(<http://eca.europa.eu/portal/page/portal/publications/auditreportsandopinions/annualreports>)

Cuffney, T.F., Bilger, M.D. and Haigler, A.M. (2007). Ambiguous taxa: effects on the characterization and interpretation of invertebrates assemblages. *Ann. Benthol. Soc.* 26(2): 286-307

De Laroche, E. (2007). LPIS update in France. Annual MARS-PAC Conference, Session T6: LPIS updating, Madrid, 12-14 November 2007:

([http://agrifish.jrc.it/marspac/meetings/Madrid2007/Session\\_T6/3\\_T6\\_Laroche\\_LPIS\\_updating\\_AUP\\_FR.pdf](http://agrifish.jrc.it/marspac/meetings/Madrid2007/Session_T6/3_T6_Laroche_LPIS_updating_AUP_FR.pdf))

De Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. *In* J.R. Barra, F. Brodeau, G. Romier and B. van Cutsem (eds.). *Recent developments in statistics*, Amsterdam, The Netherlands: North-Holland, 133-145

Elliott, J.M. (1977). *In* C.J. Krebs (1999), Chapter 7: 229-265. Some methods for the statistical analysis of samples of benthic invertebrates. *Freshwater Biological Station Association, Scientific Publication No 25*: 1-142

Fahrig, L. (2003). Effects of habitats fragmentation on biodiversity. *Annu Rev Ecol Evol Syst* 34: 487-515

Green, R.H. (1966). Measurement of non-randomness in spatial distributions. *Researches in Population Ecology* 8: 1-7

Hammer, Ø., Harper, D.A.T. and Ryan, P.D. 2001. PAST: Palaeontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1): 9 pp  
([http://palaeo-electronica.org/2001\\_1/past/issue1\\_01.htm](http://palaeo-electronica.org/2001_1/past/issue1_01.htm))

- Heltshe, J.F. and Forrester, N.E. (1983). Statistical evaluation of the Jackknife estimate of diversity when using quadrat samples. *Ecology* 66: 107-111
- Hernandez-Stefanoni, J.L. and Ponce-Hernandez, R. (2004). Mapping the spatial distribution of plant diversity indices in a tropical forest using multi-spectral satellite image classification and field measurements. *Biodiversity and Conservation* 13: 2599-2621
- Holdridge, L.R., Grenke, W.C., Hatheway, W.H., Liang, T. and Tosi, J.A. (1971). *Forest environments in tropical life zones*. Pergamon Press, Oxford, UK
- Johnson, R.A. and Wichern, D.W. (1992). *Applied multivariate statistical analysis*. 3rd edition. Prentice Hall, New Jersey
- Jongmann, R.H.G., Ter Braak, C.J.F. and Van Tongeren, O.F.R. (1995). *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, UK
- Keating, K.A. and Quinn, J.F. (1998). Estimating species richness: the Michaelis-Menten model revisited. *Oikos* 81: 411-416
- Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1-27
- Legendre, L. and Legendre, P. (1983). *Numerical ecology*. Elsevier, New York
- Leteinturier, B., Herman, J.L., de Longueville, F., Quintin, L., and Oger, R. (2006). Adaptation of a crop sequence indicator based on a land parcel management system. *Agriculture, Ecosystems and Environment* 112: 324-334
- Lo, N.C.H., Griffith, D. and Hunter, J.R. (1997). Using a restricted cluster sampling to estimate Pacific hake larval abundance. *CalCOFI Rep.* 38: 103-113
- Mac Arthur, R.H. (1965). Patterns of species diversity. *Biological reviews*, 40: 510-533
- Mac Arthur, R.H. (1972). *Geographical ecology*. Harper and Row, New York
- Michaelis, M. and Menten, M.L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* 49:333-369
- Milčinski, G., Gostenčnik, D., Kadunc, M., Okorn, A., Petkovšek, T., Rotter, A. and Vitas, D. (2007). Extending the use of LPIS infrastructure. Annual MARS-PAC Conference, Session T6: LPIS updating, Madrid, 12-14 November 2007: [http://agrifish.jrc.it/marspac/meetings/Madrid2007/Session\\_T6/2\\_T6\\_Milcinski\\_Extending\\_LPI\\_S\\_Cosylab.pdf](http://agrifish.jrc.it/marspac/meetings/Madrid2007/Session_T6/2_T6_Milcinski_Extending_LPI_S_Cosylab.pdf)
- Milenov, P. and Kay, S. (2006). Status of the implementation of LPIS in EU Member States. Proceedings of the 12th MARS PAC Annual Conference – Geographical information in support of the CAP – Session IV: IACS: sharing data online. Eds. Simon Kay, Aleksandra Sima and Philippe Loudjani. EUR 22160 EN – 2007: 43-47
- Mier, K.L. and Picquelle, S.J. (2008). Estimating abundance of spatially aggregated populations: comparing adaptive sampling with other survey designs. *Can. J. Fish. Aquat. Sci.* 65: 167-197
- Moora, M., Daniell, T., Kalle, H., Liira, J., Püssa, K., Roosalu, E., Öpik, M., Wheatley, R. and Zobel, M. (2007). Spatial pattern and species richness of boreonemoral forest understorey and its determinants – A comparison of differently managed forests. *Forest Ecology and Management* 250: 64-70
- Moreno, C.E. and Halffer, G. (2001). On the measure of sampling effort used in species accumulation curves. *Journal of Applied Ecology* 38: 487-490

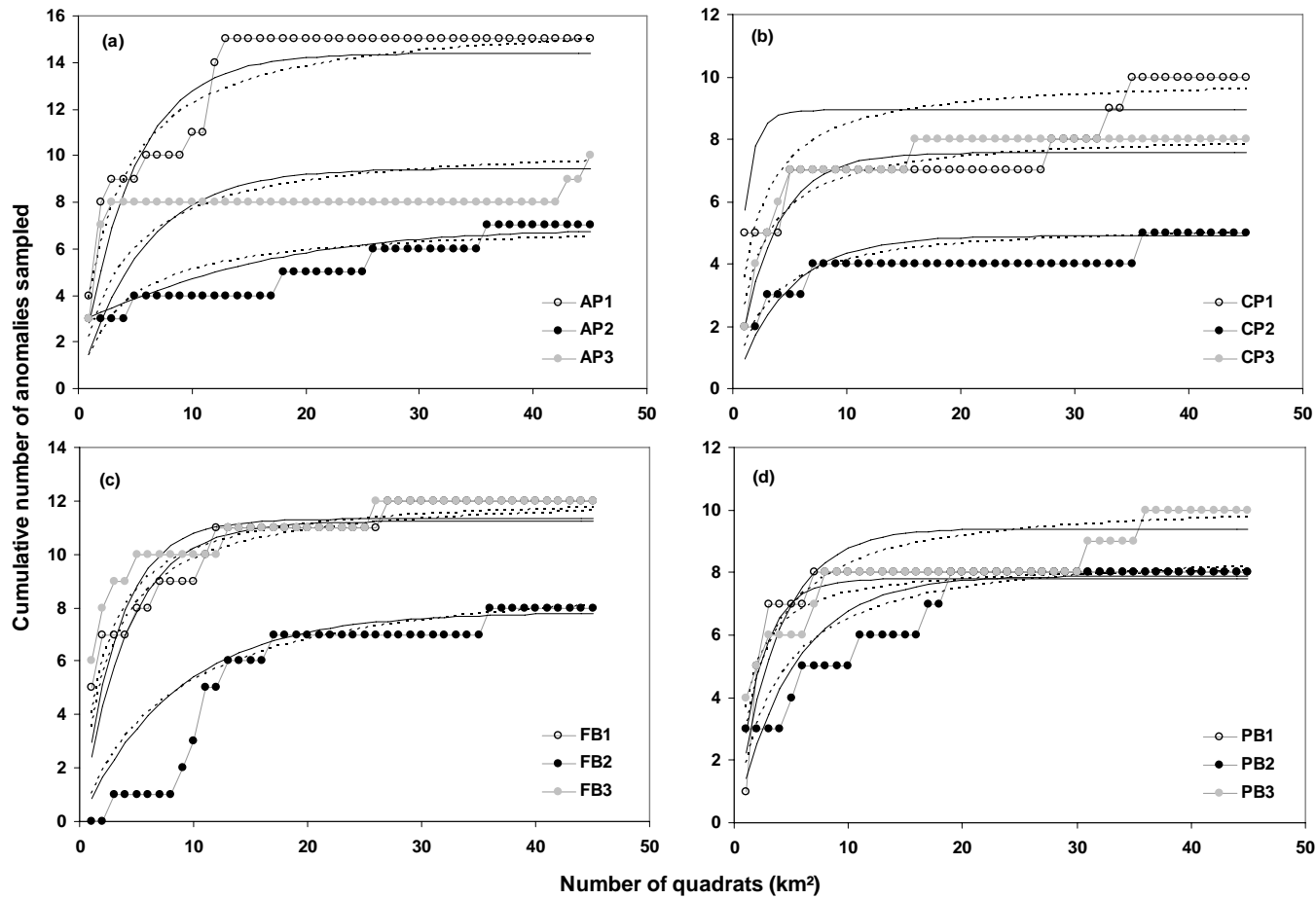
- Morisita, M. (1959). Measuring of interspecific association and similarity between communities. *Memoirs of the Faculty of Science Kyushu University Series E3*: 65-80
- Morisita, M. (1962). Id-index, a measure of dispersion of individuals. *Researches in Population Ecology* 4:1-7
- Myers, J.H. (1978). Selecting a measure of dispersion. *Environmental Entomology* 7: 619-621
- Oesterle, M. and Hahn, M. (2004). A case study for updating land parcel identification systems (IACS) by means of remote sensing. XXth ISPRS (International Society for Photogrammetry and Remote Sensing) Congress; Istanbul, Turkey, 12-23 July 2004
- Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62
- Palmer, M.W. (1990). The estimation of species richness by extrapolation. *Ecology* 71:1195-1198
- Patil, G.P., Pielou, E.C. and Walters, W.E. (1971). *Spatial patterns and statistical distributions*. Pennsylvania State University Press, University Park, Pennsylvania
- Peet, R.K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics*, 5: 285-307
- Pluto-Kossakowska, J., Grandgirard, D. and Kerdiles, H. (2007). Assessment of parcel area measurement based on VHR SAR images. *Proceedings of the 2007 Annual Conference of the Remote Sensing and Photogrammetry Society (RSPSoc2007) – Newcastle upon Tyne, 11-14 September 2007*
- Ratkowsky, D.A. (1990). *Handbook of nonlinear regression models*. Marcel Dekker Eds., New York
- Remenyi, D., Bannister, F. and Money, A. (2007). Designing IT surveys for benefit measurement – Sampling. *In The effective measurement and management of ICT costs and benefits – Third Edition*; Eds. CIMA, Oxford, UK
- Romesburg, H.C. (1984). *Cluster analysis for researchers*. Lifetime Learning Publications, Belmont, California
- Routledge, R.D. (1983). Evenness indices: are any admissible? *Oikos* 40: 149-151
- Sagris, V., Dittmann, C. and Devos, W. (2007). Towards the core conceptual LPIS model. *Proceedings of the JRC LPIS Workshop – LPIS database quality assessment and updating*; Ispra, Italy, 17-18 October 2007
- Sanders, H.L. (1968). Marine benthic diversity: a comparative study. *American Naturalist* 102: 243-282
- Shiu, H.J. and Lee, P.F. (2003). Assessing avian point-count duration and sample size using species accumulation functions. *Zoological Studies* 42(2): 357-367
- Singh, K.N., Lal, B., Singh, R.D., Todaria, N.P. and Ahuja, P.S. (2007). Species richness, distribution pattern and conservation status of higher plants in the Spiti cold desert of trans-Himalaya, India. *International Journal of Biodiversity Science and Management* 3: 223-233
- Sklenicka, P. and Salek, M. (2007). Ownership and soil quality as sources of agricultural land fragmentation in highly fragmented ownership patterns. *Landscape Ecology*, DOI 10.1007/s10980-007-9185-4, to be published
- Smith, B. and Wilson, J.B. (1996). A consumer's guide to evenness indices. *Oikos* 76: 70-82

- Smith-Gill, S.J. (1975). Cytophysiological basis of disruptive pigmentary patterns in the leopard frog *Rana pipiens*. II. Wild type and mutant cell specific patterns. *Journal of Morphology* 146: 35-54
- Soberon, J. and Llorente, J. (1993). The use of species accumulation functions for the prediction of species richness. *Conserv. Biol.* 7: 480-488
- Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Kong. Danish Vidensk. Selsk. Biol. Skr.* (Copenhagen) 5: 1-34
- Su, Z. and Quinn, T.J.II. (2003). Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. *Environ. Ecol. Stat.* 10: 17-41
- Thompson, S.K. (1992). *Sampling*. Ed. John Wiley and Sons, NY
- Tjorve, E. (2003). Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, 30: 827–835
- Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30: 279-338
- Wiegert, R.G. (1962). The selection of an optimum quadrat size for sampling the standing crop of grasses and forbs. *Ecology* 43: 125-129
- Wold, S., Esbensen, K. and Geladi, P. (1987). Principal component analysis. *Chem. Intell. Lab. Syst.* 2: 37-52
- Wolda, H. (1981). Similarity indices, sample size and diversity. *Oecologia* 50: 296-302
- Zielinski, R. (2009). An automated method for selected anomaly detection within LPIS database based on image extraction features. Doctoral thesis, Institute of Photogrammetry and Cartography, Faculty of Geodesy and Cartography, Warsaw University of Technology (WUT) – to be published

Appendix A: LPIS anomalies typology proposed *a priori*

Anomaly classes	Anomaly codes	Description
<b>Tree-related</b> <b>TRA<sub>i</sub></b>	TRA <sub>1</sub>	Patch(es) of trees – relatively homogeneous tree-covered area that differs from its surroundings fully or partly within the reference parcel and with a minimum area greater than 0.1 ha
	TRA <sub>2</sub>	Row(s) of trees – relatively homogeneous tree-covered area that differs from its surroundings in a row shape (length twice its width) fully or partly within the reference parcel and with a minimum area greater than 0.1 ha
	TRA <sub>3</sub>	Forest – an area with a high density of trees fully or partly within the reference parcel with a minimum area greater than 1.0 ha
	TRA <sub>4</sub>	Dense trees – dense coverage of single or small patches (< 0.1ha) of trees within the reference parcel and with a significant total area
<b>Man-made related</b> <b>MRA<sub>i</sub></b>	MRA <sub>1</sub>	Building – man-made structure used for residential or production purposes fully or partly within the reference parcel
	MRA <sub>2</sub>	Other man-made objects not covered by code 5 (“building”) fully or partly within the reference parcel (e.g. telecommunication facilities, electricity facilities, etc.)
	MRA <sub>3</sub>	Ground road – an identifiable ground (land substrate) road, way or path between two or more places with width greater than 2 m
	MRA <sub>4</sub>	Asphalt road – an identifiable asphalt road or way between two or more places with width greater than 2 m
	MRA <sub>5</sub>	Farmland not used for agricultural production, such as structures and facilities, excluding buildings and kitchen gardens
<b>Water-related</b> <b>WRA<sub>i</sub></b>	WRA <sub>1</sub>	Pond – typically a man-made body of water smaller than a lake
	WRA <sub>2</sub>	Ditch – a small to moderate depression created to channel water with width greater than 2 m
	WRA <sub>3</sub>	River – natural waterway that carries water through a landscape from higher to lower elevations called divides
	WRA <sub>4</sub>	Marshland – temporarily or permanently flooded zone where agricultural production is impossible
<b>Boundary-related</b> <b>BRA<sub>i</sub></b>	BRA <sub>1</sub>	Shift
	BRA <sub>2</sub>	Other agricultural land use not covered by another code
	BRA <sub>3</sub>	Boundary not following stable land features
	BRA <sub>4</sub>	The border parcel selected for checking significant overlaps with another parcel. The possible overlap should be noticeable with the working display scale (1:2000-1:25000) on GIS

**Appendix B: Anomalies accumulation curves from Clench and LD asymptotic models**



**Observed and predicted anomalies accumulation curves from Clench (dashed lines) and Linear Dependent (solid lines) models obtained from 45 km² sample size for all 12 LPIS zones selected; the three replicates by LPIS type are presented for (a) agricultural parcels (AP), (b) cadastral parcels (CP), (c) farmer blocks (FB) and (d) physical blocks (PB)**



**Appendix C: LPIS anomalies assemblages from the twelve zones studied**

**A.C.** = Anomaly category; **N<sub>A</sub>**= total number of anomalies. See Appendix A for the list of anomaly categories. See Table 2 for the list of the zones studied. Anomaly category TRA4 is between parentheses because it was undetected.

<b>A.C.</b>	<b>AP1</b>	<b>AP2</b>	<b>AP3</b>	<b>CP1</b>	<b>CP2</b>	<b>CP3</b>	<b>FB1</b>	<b>FB2</b>	<b>FB3</b>	<b>PB1</b>	<b>PB2</b>	<b>PB3</b>	<b>Total</b>
<b>TRA1</b>	130	111	130	553	3	331	105	26	347	108	51	283	<b>2 178</b>
<b>TRA2</b>	17	0	0	0	0	7	16	0	11	0	0	27	<b>78</b>
<b>TRA3</b>	4	0	0	0	0	0	0	0	1	0	0	12	<b>17</b>
<b>(TRA4)</b>	0	0	0	0	0	0	0	0	0	0	0	0	<b>0</b>
<b>MRA1</b>	109	26	280	695	101	138	67	4	160	25	82	836	<b>2 523</b>
<b>MRA2</b>	14	4	2	11	0	0	0	6	4	0	9	4	<b>54</b>
<b>MRA3</b>	37	0	3	5	19	22	9	3	16	5	0	23	<b>142</b>
<b>MRA4</b>	12	3	59	456	19	2	22	3	13	66	4	78	<b>737</b>
<b>MRA5</b>	53	15	112	249	28	85	38	8	99	23	17	370	<b>1 097</b>
<b>WRA1</b>	26	1	3	19	0	29	7	1	24	11	8	1	<b>130</b>
<b>WRA2</b>	5	0	5	28	0	9	4	0	7	100	12	0	<b>170</b>
<b>WRA3</b>	8	0	0	0	0	0	0	0	0	0	0	0	<b>8</b>
<b>WRA4</b>	1	0	0	0	0	0	0	0	0	0	0	0	<b>1</b>
<b>BRA1</b>	20	0	0	5	0	0	1	0	0	0	0	0	<b>26</b>
<b>BRA2</b>	8	0	0	0	0	0	1	0	0	0	0	0	<b>9</b>
<b>BRA3</b>	16	0	19	0	0	0	101	9	34	240	3	28	<b>450</b>
<b>BRA4</b>	0	63	1	1040	0	0	154	0	427	0	0	0	<b>1 685</b>
<b>N<sub>A</sub></b>	<b>460</b>	<b>223</b>	<b>614</b>	<b>3 061</b>	<b>170</b>	<b>623</b>	<b>525</b>	<b>60</b>	<b>1 143</b>	<b>578</b>	<b>186</b>	<b>1 662</b>	<b>9 305</b>

**EUR 23484 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen**

Title: Land Parcel Identification System (LPIS) Anomalies' Sampling and Spatial Pattern: Towards convergence of ecological methodologies and GIS technologies

Author(s): David Grandgirard and Rafal Zielinski

Luxembourg: Office for Official Publications of the European Communities

2008 – 84 pp. – 21 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-09701-0

DOI 10.2788/91912

**Abstract**

To date, the Land Parcel Identification System (LPIS) has often been proposed as the foundation for effective spatial management of agriculture and the environment and many land managers have suggested incorporating it in most of the instruments for sustainable agriculture. The LPIS is originally used for registration of agricultural reference parcels considered eligible for annual payments of European Common Agricultural Policy (CAP) subsidies to farmers. Its intrinsic quality depends on the frequency and magnitude of the discrepancies in area, since some parcels can be under- or over-declared by farmers compared with reference registered within the LPIS. General application of the LPIS therefore depends on our capacity to – first identify and explain the causes of these area discrepancies perceived as anomalies by national CAP payment agencies –second, to propose future improvements in its overall quality.

From a set of images used during the 2005 Control with Remote Sensing (CwRS) campaign, using the geographic information system (GIS) and ecological methodologies we assessed the quality of the LPIS by identifying the diversity of the existing anomalies. To that end, the ecological sampling method was adapted to the specific case of image-based detection of anomalies. The observed anomalies assemblages obtained from a set of European Member States representing the four types of LPIS were analysed to establish the spatial pattern of the anomalies.

We showed that the twelve zones surveyed can be grouped into four different clusters, each individually correlated with the presence of certain categories of LPIS anomaly. Some clusters were more particularly related to the presence of natural and anthropogenic landscape features, whereas others were typified by anomalies which stemmed from the process for creating and updating the LPIS, which accounted for 20% of the anomalies detected. Finally, we also showed that, even if useful for establishing procedures to manage the LPIS, the LPIS typology used in the European Union had no effect on the anomalies assemblage or on the spatial pattern; consequently, the type of LPIS no longer needs to be considered and LPIS anomalies assemblages could be pooled across Europe.

In the light of the results obtained, different proposals are made to improve LPIS quality by:

- identifying the critical points along the LPIS management chain;
- using landscape ecological methodologies to explain the causes of the clusters observed; and
- extrapolating the whole results in the CwRS risk analysis to perform *ex-ante* LPIS anomalies risk map.

**Keywords**

Land Parcel Identification System, Control with Remote Sensing, orthophoto, quality assessment, diversity, spatial pattern, landscape structure

### **How to obtain EU publications**

Our priced publications are available from the EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

