Agriculture and fisheries

Authors: Laura MARTINO, Marco FRITZ

eurostat
Statistics in focus
33/2008

New insight into land cover and land use in Europe

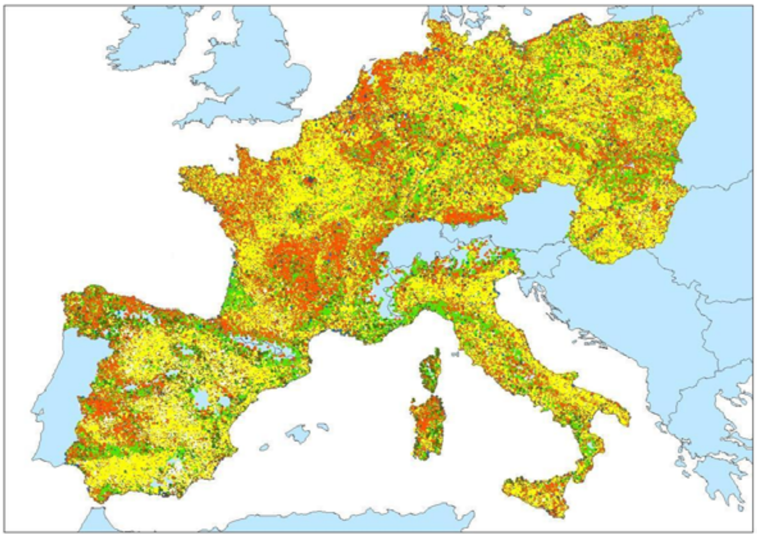
Land Use/Cover Area frame statistical Survey: Methodology and Tools

The "Land Use/Cover Area frame statistical Survey" (LUCAS[1]) aims to inform decision makers and the general public about changes in management and coverage of the European territory. This SiF provides an overview of the methodologies applied in LUCAS from 2006 onwards, resulting from the experience gathered in the previous years (2001-2005). The current paper is the first of a series of publications on LUCAS. They will give a deeper insight into the results of the survey in the most recent exercises and their contribution to agro-environmental and landscape quality topics.

Background

Area frame surveys represent a common approach to gather land cover and land use data. In contrast to mapping techniques (e.g. the CORINE Land Cover project), they provide quantitative statistical results with precision indicators attached to them. Based on the visual observation of a sample of geo-referenced points, estimates of the extent of the main land cover/use classes are computed applying nomenclatures, sampling procedures, data collection methods and statistical estimators, which are harmonized at European level. This approach has the important advantage of creating only minimum disturbance to land owners and farmers, since data are directly collected in the field by surveyors.

Figure 1: Visualization of LUCAS results. Distribution of app. 170.000 surveyed points in eleven EU Member States according to main land cover categories (2006): dark blue = artificial land; yellow = cropland; green = woodland; brown = shrubland; orange = grassland; white = bare land; light blue = water.



[1] *The LUCAS project has been implemented following the Decision N°1445/2000/EC of the European Parliament and of the Council of 22 May .2000 "*On the application of area-frame survey and remote-sensing techniques to the agricultural statistics for 1999 to 2003*". It continued until 2007 by Decision 2066/2003/EC of 10 November 2003, extended to EU-N10 by Decision 786/2004/EC of 21 April 2004.*

eurostat
EUROPEAN COMMISSION

Purpose of LUCAS

The overall objectives of the LUCAS project are:

- developing a standard survey methodology in terms of the sampling plan, nomenclature, data collection process and statistical estimators to obtain harmonised and unbiased estimates at EU level of the main land use/cover areas and changes;
- collecting information on agricultural land cover and land use within an acceptable time-lag and providing estimates of crop areas;
- offering a common sampling base (frame, drawing procedure, scheme) that Member States can use to obtain representative data at national/regional level by increasing the sampling rate while respecting the general LUCAS approach;
- testing the suitability of the survey for collecting data on environment, multi-purpose land use, landscape and sustainable development.

The 2006 sample has been designed to provide area estimates for land cover and land use at European level. However, results may be obtained at a more detailed geographical level if land use/ cover items are properly aggregated and if the basic characteristics of the survey plan are taken into account during the interpretation of results.

Sampling scheme

The LUCAS survey is based on an area-frame sampling scheme. Advantages of this sampling method compared to the traditional one, where the frame is based on a list of farms, are the complete coverage of the population (territory), the objectivity of data collection (land cover/use are directly observed by surveyors in the field) and the reduced burden on farmers. Drawbacks of the method are the limited precision of estimates provided for small areas or highly concentrated land cover/use classes and the high technical expertise required for conducting the survey.

Starting from the 2006 exercise onwards, a *two-phase sampling method with stratification* was adopted in the LUCAS survey, after being tested in Greece and Italy in 2004 and in Latvia, Lithuania and Poland in 2005.

First phase (base and master sample)

The **base sample** was obtained using a 1 km-squared grid resulting from the INSPIRE (INfrastructure for SPatial InfoRmation in Europe) recommendations; it included around 4,000,000 points in the entire European Union territory.

In addition to the ease of data management, the insensitivity to national borders and the ease of extension of the system, many other reasons speak in favour of selecting a common grid covering the Member States. Those include the possibility of better harmonisation and the exchange of cross-sector information.

The LUCAS **master sample** is a subset of the base sample corresponding to a 2 km grid covering the EU-territory. After excluding points located on small islands, the sample included 958,325 points spread over 21 countries (Bulgaria, Cyprus, Latvia, Lithuania, Malta and Romania were not covered). Each point of the master sample was photo-interpreted and assigned to one of the pre-defined land cover/use strata. The photo-interpretation in Latvia and Lithuania was conducted separately in the frame of a pilot study.

The projection used is the Lambert Azimuthal Equal Area coordinate reference system (LAEA 89). The grid is squared, with origin: 4,321,000 m West of centre point of the projection (52N, 10 E), and 3,210,000 m South of the projection centre point (52N 10E) and orientation: South – North, West – East. Each point has been given a unique numeric code going sequentially from South-West to North-East direction.

Infobox: Why this area frame approach?

LUCAS is an area frame survey. Such a data collection approach is used for collecting information on land cover and land use and establishing time series on their changes mainly on European scale. Its implementation has different steps: first a hypothetical grid is laid over the EU territory. The grid nodes are super-imposed over aerial photos and satellite images, with the land cover on these points photo-interpreted and pre-classified ("stratification"). For receiving the necessary detailed classification and avoiding errors due to photo-interpretation, a sample of these points is physically surveyed on the ground. The results, which the surveyors report to the office, are combined with the results of the stratification, for calculating area estimates on the land cover and land use classes all over Europe.

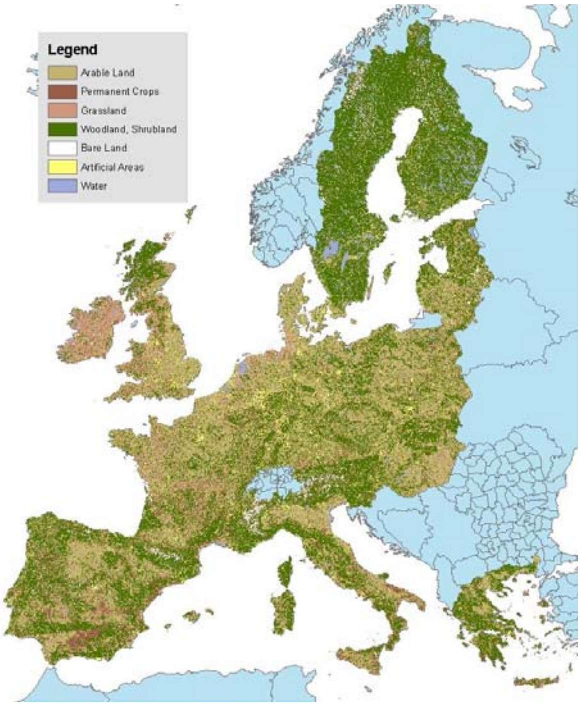
Stratification

Each point of the master sample is photo-interpreted in order to classify the sample into seven strata (“arable land”, “permanent crops”, “permanent grassland”, “wooded areas, shrubland”, “Bare land, low or rare vegetation”, “artificial land” and “water"). This photo-interpretation is based on the most recent ortho-photos or, where ortho-photos are not available, on satellite imagery (CORINE Image 2000 Landsat Images).
Results of the stratification on 21 countries are reported in Table 1 and Fig. 2 (data source: LUCAS).

Table 1: Stratification results

First phase sample		
	Area in %	**Variation Coefficient (%)**
Arable land	25.18	0.008
Permanent crops	2.94	0.010
Grassland	16.65	0.009
Woodland and shrubland	45.87	0.006
Bare land	2.06	0.010
Artificial land	4.09	0.010
Water	3.21	0.010

Figure 2: Land cover distribution in 21 EU Member States according to ortho-photo interpretation (2005)



Second phase (field survey sample)

From the stratified master sample, a sub-sample of points (**field sample**) was extracted to be classified by field visit according to the full land nomenclature[2]. The 2006 survey was carried out in 11 Member States (Luxembourg, Belgium, Czech Republic, Germany, Spain, Poland, Italy, France, the Netherlands, Hungary and Slovakia) to test the methodology at EU level with a restricted budget. The focus of the survey was agricultural land with a sampling rate of 50% for arable land and permanent crops, and of 40% for grassland, (non-agricultural strata are covered with a sampling rate of 10% each).

Table 2: Field sampling size per country (2006)

Country	Number of points
BE	2372
CZ	5626
DE	27771
ES	34629
FR	39304
HU	8468
IT	20528
LU	198
NL	2925
PL	24130
SK	3392
Total	169343

All points above 1.200 m were discarded from the sample (with the help of a digital terrain model), thus limiting the cost per point. After the ground survey, when the points were allocated to the appropriate land cover class, the sample density (per 100 km^2) resulted between 3 and 12 points (see Table 3). Each sampling unit represented around 10 km^2 for cropland, grassland and bare land; more than 22 km^2 for artificial land, shrubland and water; and 33 km^2 for woodland, being the least represented land cover. The unexpected high density of points in the bare land (selected with a sampling rate of 10% in the second phase) was due to the frequent confusion of bare land as arable land in photo-interpretation.

Table 3: Sample density and point representa-tivity by land cover classes (2nd phase sample)

LC	Est. Area	S-Size	Dens.	Repr.
Artificial Land	122,863	5,592	4.55	22
Cropland	667,369	78,467	11.76	9
Woodland	635,851	19,247	3.03	33
Shrubland	112,934	5,029	4.45	22
Grassland	521,555	50,914	9.76	10
Bare land	78,801	7,702	9.77	10
Water	36,745	1,453	3.95	25
Total	2,176,117	168,404	7.74	13

LC = Land cover class; Est. Area = Estimated area (in km²); S-Size = 2nd phase sample size; Dens. = 2nd phase sample density (per 100 km²); Repr. = Representativity of each 2nd phase sample point (in km²)

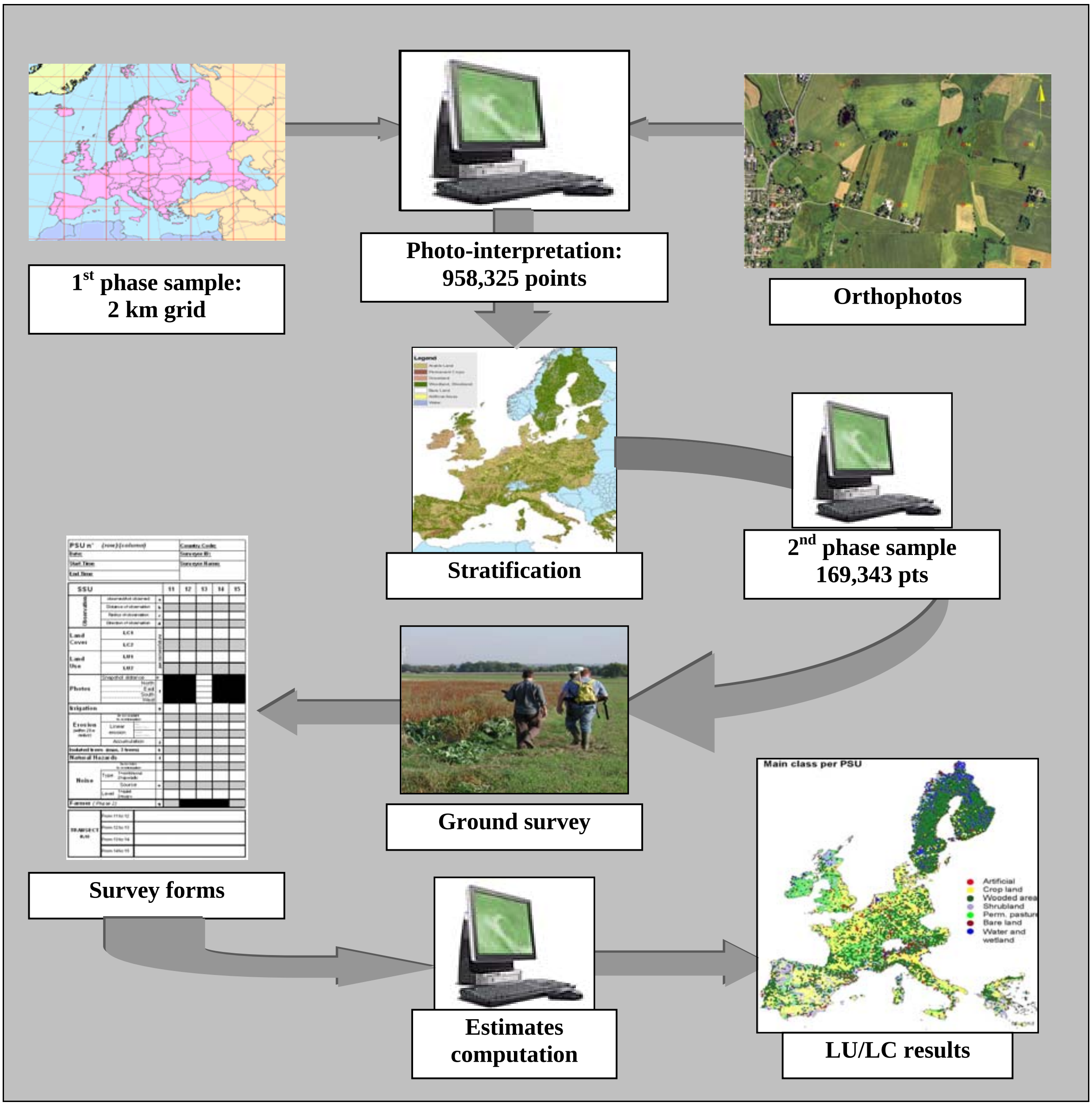
[2] *Two levels nomenclature for Land Cover – 54 classes, and three levels nomenclature for Land Use – 33 classes.*

Overview of the process

The entire 2006 survey process can be summarized in the following steps:

- Selection of the master sample. The LUCAS master sample consists in the intersection points of a 2 km-grid covering the territory of 23 European countries. It includes 958,325 geo-referenced points;
- Photo-interpretation and stratification. Each point of the master sample is photo-interpreted, and classified into 7 strata;
- Selection of the field sample. From the stratified master sample, a sub-sample of around 170,000 points (spread over 11 countries) is extracted in order to be classified by field visits according to the full land cover and land use nomenclature;
- Field work. In the period March – June 2006, surveyors collect the information on the field sample in order to provide estimates of "southern" countries by mid-June and of other countries by mid-July;
- Double blind survey. In parallel with this data collection, an additional survey is conducted, repeating the ground visit on a subset of points to detect and reduce the measurement errors;
- After validation of the data, final estimates of land cover/use extend are produced.

Figure 3: Steps of the survey



Infobox: What is the use of such a ground survey?

LUCAS is currently the only in-situ land cover and land use collection with a harmonised nomenclature over the whole EU territory. It is not a mapping exercise but is based on statistical calculations. It gives reliable results on area entities (with the currently applied sample density on regional level and higher). Due to the direct observation of the ground, it can deliver a detailed land cover classification (e.g. distinction between different cereals like wheat and rye). Such a detailed classification is technically not yet possible with photo-interpretation approaches. Satellite interpretation mapping exercises like CORINE land cover use however the ground-surveyed LUCAS data and the photos of the sample point, crop and landscape taken for the verification of their image interpretation results. As all LUCAS points are geo-referenced and re-sampled during the time, the survey also allows monitoring the condition of the environment in Europe.

Double-blind survey

A second survey was organized in 2006 in parallel with the main LUCAS survey. An independent company, not involved in the main LUCAS exercise, was charged with its implementation. No information was provided on the results of the main survey. The double blind-survey sample size represented 5% of the total survey sample and covered almost 8.200 points.

The outcomes of the double-blind survey were:

- Land cover/land use being the same between the two surveys (correct points observed from the same location);
- Land cover/land use being different between the two surveys but both correct (different rules of observation applied – look North/East – or change in land cover between the two visits (crops harvested, new crop sown, building built, etc...);
- Land cover/land use being different between the two surveys and lack of sufficient information to say which survey is correct;
- Land cover/land use being different between the two surveys and the double-blind survey being correct;
- Land cover/land use being different between the two surveys and the main survey being correct.

At the end of the exercise, the percentage of agreement in the two data collections ranged from 63% to 81% in the various countries, but it improved significantly (always over 90%) after some post-processing of survey data with the help of visual analysis of crop, point and landscape photos.

The pre-eminent result of the double blind survey was that it introduced as much errors as the main survey for many reasons (time-lag between the two surveys, same difficulties faced by surveyors as in the main survey, difficulties in locating the exact point, etc.). The main conclusion of the double-blind survey was that the analysis of pictures taken by surveyors is a very powerful tool for correcting errors and improving the data quality.

Accuracy of stratification

In order to evaluate the accuracy of the stratification, a comparison of the classification of the points according to the ground observation and the photo-interpretation was conducted in a comparable nomenclature of seven classes. Since points belonging to different strata were sub-sampled with a probability, which is, in some cases, five times larger, a weighted proportion of agreement is computed in addition to the unweighted one (Gallego, 2006[3]).

The unweighted proportion of agreement is 70.8%, while the weighted agreement is 74.8%. The main cases with erroneous results were the following:

- permanent grass and bare land are often photo-interpreted as arable areas;
- Woodland or shrubland is interpreted as grass. This may be partly due to problems in applying the definition of forest or woodland, in particular related to the threshold of crown coverage of the terrain.

Two main relevant remarks can be made:

1. The photo-interpretation errors affect the efficiency of the stratification but do not introduce any bias in the estimates, as long as photo-interpretation is used for stratification and not to substitute ground observations. In any case, in the 2006 survey the stratification efficacy still remained at a good level;
2. If the total area of arable land had been estimated, based on the photo-interpretation of the points (generally on ortho-photos with 1 m resolution), the bias would have been around 35%. The field survey still remains the only appropriate method if good quality data are required even with a rough classification nomenclature, such as the one used in the stratification.

[3] *Gallego (2006): Accuracy of the point photo-interpretation for stratification in LUCAS 2006. Internal documentation*

Efficiency of the design

The efficiency of the sampling scheme used in LUCAS 2006 has been compared with other possible single-stage sampling schemes with no clustering of points (within Gallego, 2007[6]).

The relative efficiency of the sampling approach A versus the sampling approach B has been computed as:

$$Eff\,(A/B) = \frac{Var(B)\times n_B}{Var(A)\times n_A}$$

Where $Var(A)$ is the estimated variance of the approach A and n_A the sample size.

Four sampling schemes have been compared in couples in Table 4:

- Simple random sampling (SRS);
- pure systematic sample (PSS);
- systematic post-stratified sample (SPSS);
- two-phase stratified sampling (TPSS).

Table 4: Relative efficiency of various one stage sampling designs

Efficiency	PSS/ SRS	SPSS/ PSS	TPSS/ SPSS	TPSS/ SRS
Cereals	1.11	1.40	1.26	1.95
Wheat	1.11	1.16	1.42	1.83
Durum wheat	1.43	1.29	1.41	2.60
Barley	1.15	1.17	1.40	1.88
Maize	1.21	1.19	1.43	2.06
Potatoes	1.09	1.06	1.36	1.57
Sugar beet	1.05	1.01	1.59	1.69
Sunflower	1.09	1.07	1.88	2.19
Rapeseed	1.07	1.10	1.50	1.77
Temp. Grass	1.20	1.21	1.28	1.85
Olive groves	1.63	1.82	0.89	2.63
Vineyards	1.43	1.55	1.44	3.19
Forest	1.00	1.74	0.38	0.66
Perm. Grass	1.12	1.38	0.64	1.00

Estimation of the area and its accuracy

The traditional two-phase estimator for stratification was used to derive estimates of land cover and land use rates (Cochran, 1977[4]):

$$\bar{y} = \sum_{h=1}^{H}\left(w_h \sum_{i=1}^{n_h} \frac{y_{ih}}{n_h} \right) = \sum_{h=1}^{H} w_h \bar{y}_h$$

Where h = stratum $h=1, \ldots , H;$

n' , n'_h = first phase sample size (total, by strata);

n, n_h = second phase sample size (total, by strata);

y_{ih} = land use presence in unit i-th stratum h-th.

As for the variance, systematic sampling is in general more efficient than random sampling. Nonetheless, the main drawback of this sampling scheme is the lack of an unbiased efficiency estimator.

The classical variance estimation formulas for two-phase sampling can be used (Cochran 12.24):

$$v(\bar{y}) = \sum_{h=1}^{H} \frac{w_h^2 s_h^2}{n_h} + \frac{1}{n'} \sum_{h=1}^{n_h} w_h \left(\bar{y}_h - \bar{y}\right)^2$$

with s_h^2 sampling estimate of the variance of y and $1/N$ disregarded since N is very large.

This however heavily overestimates the variance in most cases.

For this reason, an adaptation of Matérn's estimator (Matérn 1986[5], Gallego 2007[6]) was adopted, substituting s_h^2 by an estimate of the local variance:

$$\tilde{s}_h^2 = (1 - f_h) \frac{\sum_{i \neq j} \delta_{ij} \left(y_i - y_j\right)^2}{2 \sum_{i \neq j} \delta_{ij}}$$

where $f_h = {n_h}/{N_h}$

with δ_{ij} a decreasing function of the distance between i and j: $\delta_{ij} = \begin{cases} 1/d(i,j) & j \in J_{h,8} \\ 0 & otherwise \end{cases}$

with $J_{h,8} = \left\{ \begin{array}{c} j_l : d(i, j_l) < d(i,k), \quad i, j_l, k \in h \\ j_l \neq k \quad l = 1,..,8 \end{array} \right\}$

4 *Cochran W., 1977, Sampling Techniques. New York: John Wiley and Sons*

5 *Matérn B., 1986, Spatial variation. Springer Verlag lecture notes in statistics, n. 36*

6 *Gallego J. (2007): Sampling efficiency of the EU point survey LUCAS 2006. Proceedings of the 56th ISI session. Lisbon, 22-29 september 2007, in press*

Studies on new methods

Two studies were issued in 2006, aimed at developing new methodologies that could allow a better exploitation of data and/or improving quality of results and the efficiency of the survey organization.

Study "Multi-dimensional cross-analysis of LUCAS-data (Landscape classification)"

The objective of this study was to define and elaborate procedures and software for analysing LUCAS results based on the 850.000 landscape photos taken during the surveys and including other data sources. It was aimed at exploiting the potentialities of combining multiple sources of data, including soil type, climate, satellite images, digital elevation models, photos and ortho-photos to provide a well-articulated description of landscape.

Based on innovative techniques (morphological filtering, 3D modelling, virtual reality, segmentation, image, pattern recognition, spatial cognition), landscape photos were decomposed into three categories of objects: (i) Punctual elements like isolated trees, buildings, etc; (ii) linear elements like roads, railways, etc; and (iii) surface elements like forests, plots, lakes, etc. The organisation, quantity, shape, dimensions of those different elements were analysed at different scales to characterise the landscape constituents. A classification system was developed to aggregate the different analysed characteristics to provide useful and informative indicators. As an example, Figure 4 shows how a "landform visual envelope (skyline) index" can be derived, based on LUCAS landscape pictures.

Figure 4: Landscape photo modification

The visual envelope is a measure of the form of the landscape. It includes the geophysical aspects of landscape such as hills and mountains which gives some approximation to the openness of a landscape. It also gives an indication of the biological and manmade features and how they impact on a particular view, helping to define a landscape and the visualisation of the area. The visual envelope measures the length of skyline, attempting to distinguish between net skyline including only hills and mountains and gross skyline including buildings and trees set into the landscape.

In Figure 4, the landscape picture (1st square) is firstly segmented into the main landscape elements (2nd square), secondly sky components are isolated (3rd square), and finally the skyline length extracted as a useful landscape indicator (4th square).

Study "Technological watch"

The objective of this project (running till March 2008) is to perform a technology watch on new technologies that might be suitable for improving the main features of the survey. This includes bench-marking of the different products available on the market, advice on the best price/quality ratio of potential tools, methodological improvements to take into account new innovative products or to improve or replace fieldwork, and proposals for use of new sensors for some specific policy areas. This includes:

- Access to geographic data: orthophotos/high resolution satellite images; new potential sensors with links to policy areas like GMES and INSPIRE;
- Ground positioning techniques/ GALILEO/GPS/ mobile communication to GPS signals;
- Technologies for improved field data gathering;
- Technologies to substitute direct human observation on the field / photo-interpretation.

The improvements expected in the LUCAS survey by the introduction of the new technologies are being evaluated with respect to a pre-specified set of features as summarized below.

Table 5: Potential improvements versus features

Improvements	Character
Simplification of the survey	• Measurement specifications • Additional functions • Field capability • Data entry directly in the device • GPS-Features
Quality of data	• Accuracy /Satellite reception • Additional functions • Quality of photos • Data entry directly in the device (avoiding transcription errors)
Rise in the efficiency of the survey	• Handling • Memory capacity • Paperless survey • Quality of photos
Reduction of cost	• Reduce the time of survey with paperless survey • Using car navigation systems • Shorten the workflow

Further information

Journalists can contact the media support service:

Bech Building Office A4/125 L - 2920 Luxembourg
Tel. (352) 4301 33408 Fax (352) 4301 35349
E-mail: eurostat-mediasupport@ec.europa.eu

European Statistical Data Support:

Eurostat set up with the members of the 'European statistical system' a network of support centres, which will exist in nearly all Member States as well as in some EFTA countries.

Their mission is to provide help and guidance to Internet users of European statistical data.

Contact details for this support network can be found on our Internet site: http://ec.europa.eu/eurostat/

A list of worldwide sales outlets is available at the:

Office for Official Publications of the European Communities.

2, rue Mercier
L - 2985 Luxembourg

URL: http://publications.europa.eu
E-mail: info@publications.europa.eu

Manuscript completed on: 20.03.2008
Data extracted on: 11.11.2007
ISSN 1977-0316
Catalogue number: KS-SF-08-033-EN-N