Construction of a large scale (1:50k) land cover database in Hungary

G. Büttner, M. Bíró, B. Kosztra, G. Maucha, R. Pataki & O. Petrik
FÖMI Remote Sensing Centre, Budapest, Hungary

ABSTRACT: The standard CORINE Land Cover (CLC) database (scale 1:100,000) covers most of the European Member States and accession countries and is used to support policy making at the pan-European level. National applications however require a more detailed database on Land Cover. As part of Hungary's preparation to join the European Union, a CLC mapping programme (scale 1:50,000) is financed by the Ministry of Agriculture and Regional Development and Ministry of Environment. The standard (level-3) CORINE Land Cover nomenclature has been enhanced to include nearly 80 level-4/5 classes. Orthorectified SPOT-4 satellite images taken in 1998-99 and visual photointerpretation on computer screen have been applied to provide high positional accuracy. The 0.04 km^2 minimum mapping unit (0.01 km^2 for water) provides enhanced geometric detail. Internal and external quality control procedures are other key elements to yield a high quality database. The paper highlights the European co-operation aspects, as well as introduces the technical novelties of the CLC50 project. Completed results of CLC50 are described.

1 THE EU'S CORINE LAND COVER PROJECT

The idea to produce a uniform pan-European land cover database dates back to the early 80's. It has been recognised that land cover is a basic information requirement for the management of the environment and natural resources. Land cover mapping has become an integral part of the CORINE (Co-ordination of Information on the Environment) Programme, started in 1985 by the European Commission Directorate General XI (now known as DG Environment) with the main aim to compile consistent and compatible information on the environment for Member States of the European Union.

Information provided by earth observation satellites provides the base data to support the production of a land cover inventory. Following a feasibility study, the basic methodological questions addressed (nomenclature, scale, guidelines for visual photo-interpretation, etc) to facilitate the production of the CORINE Land Cover (CLC) database. A pilot project was initiated in Portugal (1986-1990) to test and refine the methodology and resulted in a CLC for the country. Following this the CORINE Land Cover methodology was finalised and a Technical Guide produced (European Commission 1993). In 1994 the European Environment Agency based in Copenhagen undertook the maintenance and use of the CORINE Land Cover database.

The CORINE Land Cover project has been implemented in most of the EU countries as well as in the 13 Phare partner countries in Central and Eastern Europe and in two countries of North Africa (Morocco and Tunisia). In each country local team(s) have implemented the project with supervision of the CORINE Land Cover Technical Unit (LCTU). North European countries (Sweden, Finland and Great Britain) have developed specific geographic information system (GIS) based procedures to assist in the derivation of the CORINE Land Cover classes, starting with the automatic classification of satellite imagery (Jaakkola 1994, Swedish Space Corporation 1994). Up to the year 2000, 28 European countries have been involved in the CORINE Land Cover mapping project resulting in a total area of 4.4 Mkm^2 being mapped (Büttner et al. 1998).

The basic aim of the CORINE Land Cover project is to provide an inventory of the Earth surface features for managing the environment. Computer assisted visual interpretation of satellite images has been chosen as the preferred mapping technology. The choice of scale (1:100,000), minimum mapping unit (0.25 km^2) and minimum width of linear elements (100 metres) represents a trade-off between the costs of production and the details of land cover information derived. The main technical parameters of CLC100 are summarised in Table 1.

The acquisition years of the satellite imagery used to derive the CLC100 database for the different participating countries vary. A new project, IMAGE&CLC2000, is undertaking to update the standard CLC database based on satellite imagery giving a "snap shot" of Europe for the year 2000. The complete update of CLC, known as CLC2000, is expected to be completed during 2003 (Steenmans 2000).

Table 1. Main parameters of CLC100 and CLC50

Parameter	CLC100 Hungary	CLC50
Nomenclature	standard EU level-3	extended level-4/5
Methodology	hardcopy PI	softcopy PI
Area resolution	0.25 km^2 for all categories	0.04 km^2 0.01 km^2 for lakes
Linear resolution	100 m	50 m
No. of classes	27 (out of 44)	78
No. of polygons	24.000	>150.000 (estimation)
Positional accuracy	<100 m (RMS)	<20 m (RMS)
Thematic reliability	>80%	>90%
Supervision	not documented: direct corrections on plastic overlays	documented: remarks on polygon level (instructions for corrections)
External quality control	no	yes
Final product	topologically structured database	

2 HUNGARY'S CLC50 PROJECT

To support Hungary's accession to the European Union there is strong need for detailed and up-to-date land cover information. CLC100 is already outdated for Hungary (as it was produced using satellite images taken in 1990-92), and, in addition does not provide sufficient detail. A resolution of the Hungarian government on the "Development of environmental information systems" (2339/1996(XII.6.)) declared the "Setting up a CORINE Land Cover database at a scale of 1:50,000". In 1999, as part of the preparation to the EU accession – in the frames of the Acquis National Programme - the Ministry of Agriculture and Regional Development initiated the CORINE Land Cover mapping project at a scale of 1:50,000 (CLC50). The results of the project will be used in several national programmes principally in the fields of agri-environment and nature protection, e.g. the development of:

- sustainable land use (e.g. converting arable land to grassland and forest land),
- integrated land use management for landscape, soil and hydrological conservation areas,
- network of Environmentally Sensitive Areas,
- habitats directive for nature protection, and
- rural development.

The CLC50 project has direct links to the standard European CORINE Land Cover. Thematic classes have been developed from the standard (level-3) nomenclature, using recommendations of Phare countries for 1:50.000 scale mapping, prepared for the European Environmental Agency (Feranec 1995). This enhanced nomenclature (level-4 and level-5 classes) includes nearly 80 categories, adapted for Hungarian conditions (only 27 level-3 CLC classes have been applied in the CLC100 Hungary project). Orthorectified SPOT-4 satellite images taken in 1998-99 and the visual photointerpretation on computer screen allow for high positional accuracy of polygon delineation. The 0.04 / 0.01 km^2 size of minimum mapping unit (0.25 km^2 in the standard CLC) provides

enhanced geometric detail. A rigorous internal supervision and an independent external quality control by experts of the Nature Protection Directorates and Plant Health and Soil Protection Services are the other key elements of this high quality database. Figure 1 shows the flowchart of CLC50 data processing.

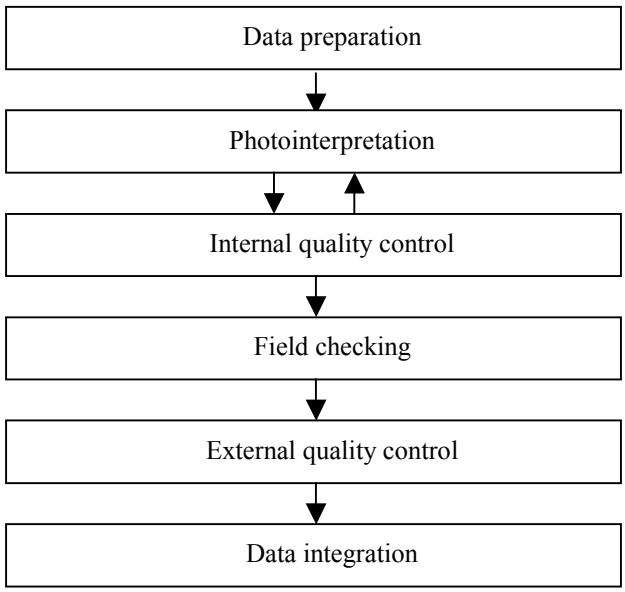

Figure 1. Flowchart of CLC50 project

2.1 *Data preparation*

Soon after its launch in 1998, SPOT-4 satellite data was recognised as fulfilling the geometric as well as thematic specifications of CLC50 (Table 1). This is because of the 20/10 meter pixel size in multispectral (XI) and monospectral (M) modes, and the availability of short-wave infrared band in the XI mode. Therefore this satellite was programmed to acquire imagery for the entirety of Hungary with simultaneous XI and M images during the summer period. About 85% of the country was acquired cloud-free during 1998, and the remainder during 1999. In total 49 XI and M image pairs have been acquired with <1% cloud cover.

Orthocorrection of level-1A imagery was undertaken using the RSG package (Almer et al. 1991) and ERDAS Orthobase. These packages transform the raw imagery into the required cartographic projection in a single process by using a rigorous mathematical model of imaging geometry and the actual Earth surface. Ground control points (GCP) were selected from 1:10,000 scale topographic maps. A digital elevation model (DEM) was generated from contour lines of 1:100,000 scale topographic maps and refined by using about 50,000 measured elevations of the 4th order geodetic network. For the flat regions of the country, high quality GCPs were used from the Hungarian Crop Monitoring Programme (Csornai et al. 1998), and rectification was performed on a single image basis. For hilly regions of the country block adjustment was used with fewer GCPs, but with several tie points selected from overlapping areas between images. Positional accuracy of the rectified images was <10 meter (RMSE) for the entirety of the country. Orthocorrected XI and M images were finally merged using the HPF method (Chavez 1991) to provide the basic corrected satellite data for photointerpreters to use to derive the CLC50 database.

In order to increase the thematic accuracy of interpretation, photointerpreters were provided with other, additional satellite images as well. Images taken in spring helped interpretation of coniferous forests or grasslands. Images of the CLC100 project taken in 1990-92 and recent (1997-99) summer time series provided additional data to improve interpretation of those classes, where time evolution yielded important information (e.g. arable land or fallow land, transient or permanent water body). The number of additional images (Landsat-5 TM and IRS-1C/D data) for a map-sheet varied between 2 and 5.

2.2 *Nomenclature*

Several initiatives exist to extend the CORINE Land Cover methodology to a working scale of 1:50,000 and 1:20,000 (ETC/LC 1997). The Phare programme funded a research project that resulted in the successful development of a methodology and production of 120 CLC map sheets at 1:50,000 scale for the Czech Republic, Hungary, Poland and Slovak Republic. Using integrated SPOT PAN and Landsat TM data it was possible to use 0.04 km^2 as the minimum mapping unit. An international team of experts extended the standard nomenclature with level-4 classes to represent the landscape conditions of the above four countries (Feranec et al. 1995). The number of level-4 classes was about twice that of level-3 ones. This research project demonstrated that a CORINE Land Cover database at larger scale could be produced. Later the nomenclature was further extended as part of the Phare Topic Link/Land Cover activities, to represent all Phare countries. This nomenclature includes 104 land cover classes and could be a base of an all-European level-4 nomenclature. As preparation to the CLC50, it has been adapted to Hungarian conditions as follows:

- some of the definitions were refined,
- categories not recommended for use in Hungary were excluded,
- two new level-4 categories were introduced,
- some level-5 categories were introduced.

The resulting CLC50 nomenclature (see Annex 1) is fully compatible with the standard European CLC100.

2.2.1 *Artificial surfaces*

This includes mainly land use categories. It provides enhanced detail for settlements, industrial areas, transport infrastructures, mines, dumpsites and green urban areas.

2.2.2 *Agricultural areas*

Arable land (a dominating category in Hungary) is distinguished according to field size (limiting parcel size is 0.1 km^2). This group includes classes for greenhouses, permanently irrigated arable land, rice-fields, vineyards, orchards and three other types of permanent cultures. The definition of 231 ("pastures") has been refined to allow for better separation from 321 ("natural grassland"). This refinement is expressed in the name of the category: 231="intensive pastures, degraded grassland". It is subdivided into two level-4 classes, representing pastures without trees and shrubs and pastures with trees and shrubs. The agricultural mosaics ("complex cultivation pattern") are separated whether they include scattered buildings or not. Farmsteads are characteristic features of the Great Hungarian Plain are included in level-5. There are five different level-4 classes to detail class 243 ("agriculture with significant representation of natural vegetation") according to the dominating cover type.

2.2.3 *Forests and semi-natural areas*

This group includes more thematic details for all level-3 forest classes according to habitat (dry or wet) and the canopy closure (continuous or discontinuous). Natural (semi-natural) forests and plantation forests are also separated. There are two level-4 classes for "natural grassland" whether it includes trees and shrubs or not. Category 324 "transitional woodland shrub" includes four level-4 classes to represent clear-cuts and new plantations, natural regeneration areas, forest nurseries and damaged forests. Areas covered by sparse vegetation are discriminated according to the characteristic soil type: sands / loess, rock or saline. There is a separate category for burnt natural areas.

2.2.4 *Wetlands*

There are four categories under this group: fresh-water marshes, saline marshes, explored and un-explored peat bogs.

2.2.5 *Waters*

This group includes rivers, channels, natural and artificial lakes. Natural lakes are further subdivided on level-5 as water bodies with permanent water supply, and temporary, salt affected lakes. (The latter is a speciality in Europe, and "ex-lege" protected). "Fish ponds" are a separate

level-5 category, because they represent an important economic factor in the country, and the regular land use inventory of the Hungarian Central Statistical Office includes such an item.

2.3 *Computer assisted photointerpretation*

The standard CLC100 project is based on traditional photointerpretation. A transparent overlay is fixed on top of the satellite image hardcopy, and the photointerpreter draws polygons on it which are coded by a three digit code (European Commission 1993). Although the display and interpretation of satellite imagery using an image processing system was recommended so as to compensate for drawbacks of using hardcopy images, these facilities were not always available. After completing the interpretation and supervision, the polygons were digitised, a topology was created and the CORINE codes were entered. This procedure often gave several different types of error:

- geometrical errors given by an imprecise hardcopy image size, distortions of the hardcopy image, improper alignment of overlay and image;
- geometrical errors of digitization;
- thematic mistakes introduced during the database coding phase;
- thematic errors because of the limitations of hard copy image and single-image interpretation.

Computer assisted photointerpretation completely eliminates the first three of the errors described above, and reduces the last one by optimal combination of capabilities of the human expert and that of the computer. The main benefits over the traditional approach include:

- magnification of the imagery using the zoom functionality;
- creation of multitemporal imagery;
- more precise delineation of polygons;
- easy corrections (lines, codes);
- automatic checking of validity of codes;
- automatic checking of polygon geometry (area, average width);
- on-line nomenclature;
- the possibility to use comments and remarks at a polygon level (i.e. a tool for „discussion" between photointerpreter and quality supervisor);
- data exchange via e-mail.

The above functions were implemented under ESRI ArcView V3.1/3.2 as a macro package, called InterView (Taracsák 2001). Special care was taken to support error-free conversion of ArcView shape files into polygon topology under workstation ArcInfo V8.0.1.

2.4 *Internal quality control*

The role of internal quality control is to harmonise the work of photointerpreter team to ensure similar understanding and application of the nomenclature in order to provide a homogeneous database. In the frames of the standard CLC100 project quality control was accomplished by the LCTU. This meant visual evaluation of drawings on the plastic overlay and corrections by red pencil.

In CLC50 this task was also modernised. Members of the photointerpreter team (located in different parts of the country) submitted their intermediate results (ArcView files) for supervision usually by e-mail. For the expert controlling the work the same images and ancillary data are available as for the photointerpreter. An ArcView macro package called InterTest (Taracsák 2000) is used for supervision. This allows for entering remarks on polygon level to instruct the interpreter about the problem and help the correction, e.g.:

- "incorrect code, proposed code is: xxxx";
- "a separate polygon with a code xxxx should be delineated at location: X= uuu, Y=vvv");
- "field checking is necessary";
- "polygon needs re-interpretation".

InterTest includes several standard messages to speed up supervision, but intentionally has no functionality to change neither lines nor codes. The remarks are "read only" for the InterView user (the photointerpreter). They can enter their comments in a separate field (e.g. "corrected", or explain why they do not agree with the remark). This way the evolution of the work is fully documented and the functions of interpretation and quality control are totally separated. The

interpreters get back their work with these remarks in the database. Also a protocol is sent with the list of remarks and a summary report (including technical remarks, e.g. warnings for un-coded areas, and a summary evaluation of the thematic work). Usually two iterations were needed to achieve an acceptable quality. After field checking (1 day per sheet) another, final evaluation is done, usually revealing only a few additional errors.

2.5 *External quality control*

After the final internal quality control the photointerpretation is subjected to an external quality control. Experts from two networks, the Nature Conservation Directorates and Plant Health and Soil Protection Services are contracted to control the results of photointerpretation. These two networks operated by the Ministry of Environment and Ministry of Agriculture and Regional Development have lots of field knowledge and access to field data that is compared to the photointerpretation. As handling of digital geographic data is not yet common in these offices, so external experts are equipped with SPOT-4 satellite image map sheets as hard copy, and the corresponding CLC maps as line drawing, printed on transparency. Experts have to provide the central team with their written, numbered remarks. The numbers are indicated also on the transparency for easy identification and pin-pointing of queries.

External quality control further increased the quality of the database, especially in the following ways:

- better discrimination of "intensive pastures, degraded grasslands" (231x) and "natural grassland" (321x),
- better separation of semi-natural forests and plantations,
- inclusion of very new plantations (vineyards, orchards, forests).

2.6 *Data integration*

This step includes the following procedures:

- converting ArcView shape files into a topologically structured vector database (ArcInfo);
- checking that neighbour polygons having the same code (whether intentional or a mistake), since such polygons are dissolved during further processing;
- merging adjacent sheets;
- controlled integration of the results of external quality control phase.

3 RESULTS

The project was started at the end of 1999. During the three years since, about 7/8 of the country's area has been mapped. Areas mapped include almost 100% of territories under nature conservation (Figure 2). 16 photointerpreters and 56 external experts have contributed to the work.

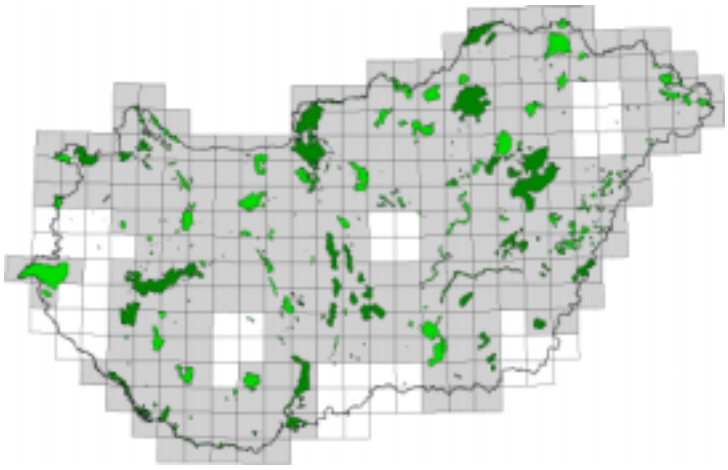

Figure 2: Status of CLC50 (end of 2002) with areas under Nature Conservation.

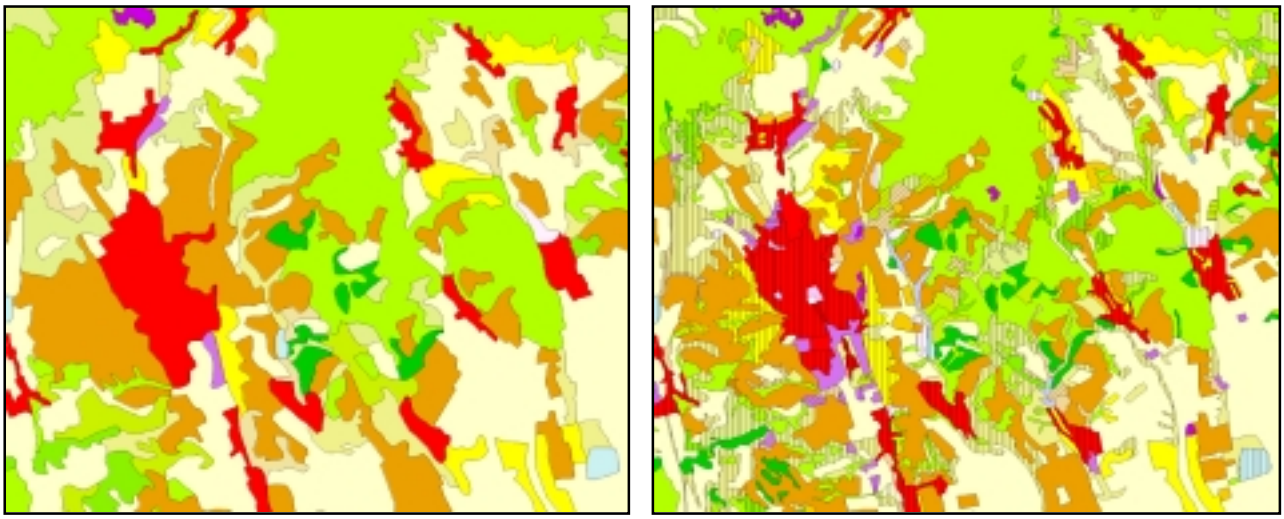

Figure 3. Comparison of CLC100 and CLC50 databases for the region of Eger (NE Hungary)

There has not yet been time for validation of geometric and thematic accuracy. Naturally, results of CLC50 include more details than CLC100. The number of linear elements (rivers, channels, highways, railway sections) has increased significantly compared to CLC100. The ratio of the number of polygons in CLC50 to that of CLC100 is between 5 and 8.

A tool based on ESRI ArcView has been developed to support computer assisted visual photointerpretation. Results have demonstrated that this approach can be successfully operationalised. Orthocorrected SPOT-4 images have provided a high quality basis for photointerpretation. Additional satellite images were very useful to supplement interpretation. The nomenclature, which is compatible with the standard one, can be applied without major difficulties. The rigorous internal quality control was a key element of the methodology, and importantly contributed to the homogeneous application of the interpretation rules. Electronic communication was great advantage to work with experts, located in different parts of the country. Good quality satellite image photomap prints and interpretation overlays helped external experts to control and amend the database by integrating their knowledge into the land cover interpretation.

4 CONCLUSIONS

As part of fulfilment of the government resolution on the "Development of environmental information systems" the implementation of the CORINE Land Cover database at scale 1:50,000 is being completed. The database supports Hungary's accession to the EU in various programmes, such as the planning of sustainable agriculture, rural development, agri-environment and nature protection.

The CLC50 project has direct links to the standard European CORINE Land Cover project, however most elements of the methodology were upgraded according to the present level of technology in geo-data processing. The CLC50 nomenclature has been developed from the standard (level-3) nomenclature and includes nearly 80 level-4 and level-5 classes, which have been adapted for Hungarian conditions. Orthorectified SPOT-4 satellite images taken in 1998-99 and computer assisted photointerpretation allow for high positional accuracy of delineation. The 0.04 km^2 size minimum mapping unit (0.01 km^2 for lakes) provides enhanced geometric detail. A rigorous internal supervision and an external quality control are the other key elements to producing a high quality database.

In three years 7/8 of the country's area has been completed. Results of CLC50 are promising, it has yielded 5-8 times more polygons than the standard CLC100.

REFERENCES

Almer, A. et al. 1991. RSG – State of the art geometrical treatment of remote sensing data. Proceedings of the 11th EARSeL Symposium, Graz, 1991:111-120.

Büttner, G. et al. 1998. The European CORINE Land Cover Database. ISPRS Commission VII Symposium, Budapest, September 1-4, 1998. Proceedings: 633-638.

Chavez, P. et al. 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT Panchromatic. Photogrammetric Engineering and Remote Sensing, 57(3):295-303.

Csornai et al. 1998. Remote sensing based crop monitoring in Hungary. ISPRS Commission VII Symposium, Budapest, September 1-4, 1998. Proceedings: 108-113.

European Commission, 1993. CORINE Land Cover, Technical Guide, EUR12585, Bruxelles, Luxembourg.

ETC/LC 1997. Assessment of the existing experiences of the 4th and 5th level CORINE Land Cover nomenclature. Prepared for the European Environment Agency.

Feranec, J., J. Oťaheľ & J. Pravda 1995. Proposal for a methodology and nomenclature scale 1:50.000 CORINE Land Cover Project. Final Report. Institute of Geography, Slovak Academy of Sciences, Bratislava.

Jaakkola, O. 1994. Finnish CORINE Land Cover – A feasibility study of automatic generalization and data quality assessment. Reports of the Finnish Geodetic Institute, 1994(4).

Steenmans, C. 2000. Update of CORINE Land Cover database. I&CLC2000 Project Document. Discussion Paper, EIONET Workshop Prague, 10-12 April 2000.

Swedish Space Corporation 1994. CORINE landtackning – ett pilotprojekt i Sverige (in Swedish, with English summary).

Taracsák, G. 2001. InterView 3.0: a programme package to support CORINE photointerpretation (in Hungarian). FÖMI.

Taracsák, G. 2000. InterTest macro package to evaluate satellite image interpretation (V1.0). FÖMI internal report.

ACKNOWLEDGEMENT

The phases of CLC50 were sponsored by the Ministry of Agriculture and Regional Development (MARD), Ministry of Environment (MoE) KAC Programme, MoE Phare Programme and MoE Nature Conservation Authority. Purchase and preprocessing of SPOT-4 satellite imagery was supported by the MARD and the MoE. The authors express their gratitude to the following photointerpreters: dr. T. Baukó, Cs. Biró, Mn. Bíró, dr. É. Csató, S. Csete, K. Hudák, dr. L. Mari, Zs. Mattányi, dr. I. Márkus, T. Morshhauser, Zs. Pataki, D. Ritter, R. Szász, Cs. Varga, and dr. J. Vámosi.

ANNEX 1: CLC50 NOMENCLATURE (V1.42)

1 ARTIFICIAL SURFACES

1111 Areas of urban centres
1112 Areas of ancient cores
1121 Discontinuous built-up areas with multiflat houses prevailingly without gardens
1122 Discontinuous built-up areas with family houses with gardens
1123 Discontinuous built-up areas with greenery
12111 Industrial and commercial units
12112 Agro-industry
12113 Education and health facilities
1212 Areas of special installations
1221 Road network and associated land
1222 Rail network and associated land
1232 River and lake ports
1233 Shipyards
1234 Sport and recreation ports
1241 Airports with artificial surfaces of runways
1242 Airports with grass surfaces of runways
1311 Open cast mines
1312 Quarries
1321 Solid waste dump sites
1322 Liquid waste dump sites
1331 Construction sites
1411 Parks
1412 Cemeteries
1421 Sport facilities
1422 Leisure areas
1423 Recreation settlements

2 AGRICULTURE AREAS

2111 Arable land with large fields
2112 Arable land with small fields
2113 Greenhouses
2121 Permanently irrigated arable land
2131 Rice fields
22111 Vineyards with large fields
22112 Vineyards with small fields
2221 Orchards
2222 Berry fruit plantations
2223 Hop plantations
2226 Wild willow plantations
2311 Intensive pastures, degraded grassland without trees and shrubs
2312 Intensive pastures, degraded grassland with trees and shrubs
2421 Complex cultivation patterns without scattered houses
24221 Complex cultivation patterns with scattered houses
24222 Farmsteads
2431 Agricultural areas with significant share of natural vegetation and with prevalence of arable land
2432 Agricultural areas with significant share of natural vegetation and with prevalence of grasslands
2433 Agricultural areas with significant share of natural vegetation and with prevalence of scattered natural vegetation
2434 Agricultural areas with significant share of ponds and with prevalence of scattered natural vegetation
2435 Agricultural areas with significant share of permanent crops and with presence of scattered natural vegetation

3 FORESTS AND SEMINATURAL AREAS

3111 Broad-leaved forest with continuous canopy, not on wet area
3112 Broad-leaved forest with continuous canopy, on wet area
3113 Broad-leaved forest with discontinuous canopy, not on wet area
3114 Broad-leaved forest with discontinuous canopy, on wet area
3115 Plantations of broad-leaved forests
3121 Coniferous forests with continuous canopy
3125 Plantations of coniferous forests
3131 Mixed forests created by alternation of single trees with continuous canopy
3135 Mixed forests created by alternation of stands of trees with continuous canopy
3139 Plantations of mixed forests
3211 Natural grassland prevailingly without trees and shrubs
3212 Natural grassland with trees and shrubs
3241 Young stands and clear-cuts
3243 Bushy woodlands, natural regeneration areas
3244 Forest nurseries
3245 Damaged forests
3313 River banks
3321 Bare rocks
3331 Sparse vegetation on sands or loess
3332 Sparse vegetation on rocks
3333 Sparse vegetation on salines
3341 Burnt areas

4 WETLANDS

4111 Fresh-water marshes
4113 Saline (alkali) inland marshes
4121 Explored peat bogs
4122 Natural peat bogs with scattered trees and shrubs

5 WATERS

5111 Rivers
5112 Channels
51211 Natural water bodies with continuous water supply
51212 Natural, temporary, salt affected water bodies
51221 Artificial lakes, reservoirs
51222 Fish ponds