Exploring the limitations of CORINE Land Cover for monitoring urban land-use dynamics in metropolitan areas

Jaime Diaz-Pacheco* and Javier Gutiérrez

Human Geography, Complutense University, Madrid, Spain

(Received 29 May 2012; final version received 19 December 2012)

One of the sources of data most utilized in Europe by planners, policy makers and researchers is CORINE Land Cover (CLC) because of its wide territorial coverage. However, the users of this spatial database do not have information on its quality for analysing urban land and its evolution. This study analyses the limitations of CLC for monitoring urban land use dynamics, using Madrid Land Use as a mapping reference. This is a local geographical database with higher resolution for the Madrid region that coincides with CLC at the years of reference (2000, 2006). The results show that with respect to the total urban land area in 2000 and 2006 and the area converted to urban land between 2000 and 2006, CLC is relatively reliable. However, the confusion matrices show very high error rates, especially in dynamic analyses (changes between 2000 and 2006) and when there is differentiation between different types of land use. Researchers, planners and decision-makers should therefore exercise considerable caution when using this source for analysing urban land use and its dynamics for local approaches.

Keywords: accuracy assessment; urbanization; spatial data; land-use change; CORINE Land Cover; Europe

Introduction

The current scientific interest in understanding urban land-use dynamics is undeniable. Although only a small percentage of the earth’s surface is occupied by cities (2% according to Glüber, 1994, cited in Lambin et al., 2001), the repercussions of social, economic and environmental processes that take place on urban land are felt beyond their limits. These not only affect the transformation of nearby urban-rural areas (Lambin et al., 2001) but also more distant areas. For example, the conversion of a certain number of hectares of woodland into industrial land has a direct effect locally relating to the loss of land for agroforestry in the municipality. At the same time, the so-called ecological footprint in the area is being changed (Wackernagel, 1996), which affects overall the environmental balance.

Monitoring of urban land-use dynamics is carried out using LULCC (land-use/land-cover change) geodatabases. The purpose of these databases is to provide spatially explicit information on LULCC. ‘It is not sufficient to know the human-environment dynamics that will expand, shrink, or intensify a particular land use; the global change community needs to know where these changes will take place’ (Turner, 2001, p. 271).

*Corresponding author. Email: jdiazpac@ghis.ucm.es

© 2013 Taylor & Francis
In the European Union, the most important initiative with respect to the building of an LULCC database is CORINE Land Cover (CLC). CLC has been largely used by planners, policy makers and researchers. However, there is a dearth of papers on determining the quality of CLC data for analysing urban land and its evolution. The present study systematically examines the accuracy of CLC with respect to the analysis and dynamics of urban land use. This entails carrying out comparative analyses of CLC with another more precise geographical database on a larger scale developed for the Madrid region, known as Madrid Land Use (MLU). The quality of the CLC data is measured by means of error matrices, using MLU as reference mapping.

This article is structured in six sections. This introduction is followed by the second section, which presents a literature review on LULCC geodatabases and their usefulness for monitoring urban land-use dynamics. The third section gives a brief description of the area of study, justifying its suitability with respect to the objectives of the study. The fourth section describes the sources and methodology used. The fifth section shows the results and the sixth section describes the main conclusions.

**LULCC geodatabases and their use for observing changes in urban land use**

Approaches to land transformation research have been based on two elements of analysis that are interrelated but different, namely, land use and land cover. The term land use denotes the human employment of the land, including settlement, cultivation, pasture, rangeland, recreation and so on, while land cover denotes the physical state of the land that involves, for example, the quantity and type of surface vegetation, water and earth materials (Turner & Meyer, 1994). Land use and land cover are strongly interconnected. In fact, physiognomic characteristics of landscape objects cannot be quite separated from their functional properties. For instance, urbanized features (artificial surfaces) or intensively used agricultural features (arable land, permanent crops) also indicate the social function of the land use (Feranec, Hazeu, Christensen, & Jaffrain, 2006).

The observation of LULCC dynamics has been carried out on different spatio-temporal scales. Global evaluations of aggregate information based on assumptions and simple models are not spatially explicit, while the specificity of local or regional evaluations is difficult to extrapolate to global areas (Turner et al., 1995). Geographers and natural scientists in particular have led the way to developing spatially explicit models of land-use change at highly disaggregate scales (Irwin & Geoghegan, 2001).

**LULCC geodatabases for monitoring land-use change dynamics on a global scale**

The most prominent global initiatives with a broad temporal retrospective are the ‘Land-Use and Land-Cover Change (LULCC)’ project, jointly developed by the International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions Programme on Global Environmental change (IHDP), aimed at achieving a ‘better understanding of land-use and land-cover changes (e.g. degradation, desertification, biodiversity loss)’ (Turner et al., 1990, cited in Goldewijk, 2001, p. 417), and the History Database of the Global Environment (HYDE), aimed at obtaining patterns of population distribution and land use for the last 300 years (Goldewijk, 2001). Like HYDE, LULCC also sets out to examine how land cover has been changed by human land use over the last 300 years, with a view to being able to model future changes over the next 50–100 years.

One of the LULCC geodatabases used in the projects mentioned is the IGBP DISCover (International Geosphere-Biosphere Programme, Data and Information Systems Initiative),
developed by researchers from the US Geological Survey (USGS) at the University of Nebraska-Lincoln and the European Commission Joint Research Centre in Ispra (Italy). This geodatabase has a global land-cover resolution of 1 km. It ‘provides a unique view of the broad patterns of the biogeographical and ecoclimatic diversity of the global land surface, and presents a detailed interpretation of the extent of human development’ (Loveland et al., 2010, p. 1303).

Both the global environmental monitoring programmes and the LULCC geodatabases mentioned above work very well for analysing phenomena to the detail that the accuracy of their scales allows. This means that when it comes to monitoring changes in urban land use, only regions with high urban dynamics can be detected. With respect to the IGBP DISCover database, attention has already been drawn to the fact that due to the heterogeneous nature of urban land use it is ‘practically impossible to map urban land cover using computer-assisted image classification methods’ (Loveland et al., 2010, p. 1315).

**LULCC geodatabases for monitoring land-use change dynamics at European level**

In the European Union, the most important initiative regarding the building of a database on land occupation is the CLC. This geodatabase has been widely used for analysing and monitoring urban land use. For example, Thinh, Arlt, Heber, Hennersdorf, and Lehmann (2002) used CLC for measuring the degree of compactness of cities; Fuller and Gaston (2009) focused their analysis on the evolution of urban green space, comparing 386 European cities; Guéros and Pumain (2008) tried to extract models of urban occupation from a comparative study of 40 cities in different European regions by measuring urban gradients from a central core and Hewitt and Escobar (2011) used CLC to study the growth of urban use on areas of agricultural land near Madrid.

This database has been used for observing changes in land use, in addition numerous studies have confirmed its usefulness for supporting environmental and socioeconomical analyses of very diverse characteristics. CLC has been used for observing pollution in the ‘Statistical air pollution interpolation model, RIO’ project (Janssen, Dumont, Fierens, & Mensink, 2008, p. 4884), for drawing up dasymetric maps of population density to exclude non-populated areas from analysis and obtain accurate geodemographic data (Gallego & Peedell, 2001; Bielecka, 2005), for supporting econometric calculations on environmental costs, relating mobility with urban morphology (Camagni, Gibelli, & Rigamonti, 2002) and for developing agro-environmental indicators (EEA, 2001).

As can be seen from the studies mentioned, CLC has covered the need to add spatial information to the study of dynamic processes of change in land use and related phenomena. However, the users of these data do not generally have information regarding the quality of CLC. Very few studies have been done on trying to determine the degree of accuracy or quality of results based on CLC and none of these has specifically dealt with the uses of urban land. Siedentop and Meinel (2004) compared CLC 2000 with another land-use database with higher spatial resolution and concluded that CLC could be used as a complementary database in urban land-use monitoring programmes. However, the ‘masking’ of small-scale development runs a risk of underestimation of sprawl dynamics. Thus, for local approaches, CLC 2000 cannot provide a meaningful database (Siedentop & Meinel, 2004). Bach et al. (2006) compared CLC with a field inventory of two districts in Germany and with another two land-use layers to evaluate the accuracy and congruency of these digital land-use maps. Catalá, Bosque, and Plata (2008) tried to detect possible errors in CLC using a simple cross tabulation procedure between 1990 and 2000, in which land-use changes considered unusual were located and analysed. The study concluded that
68% of the ‘unusual’ changes detected are related to assignation errors in land cover/use mapped. Our paper extends the previous work on the quality of CLC for monitoring urban land-use dynamics. Like in Siedentop and Meinel (2004) and Bach et al. (2006), the quality of CLC data is determined by comparison with a local and more reliable data source. However, unlike these studies, analysis is carried out by means of error matrices, which is both static and dynamic and differentiates different classes of urban land.

**Area of study**
The area selected for study is the Madrid region, chosen not only because of the availability of spatial data and statistics but, above all, for the diversity of its urban spaces and the great dynamism that this region experienced during the period of change analysed in this study (2000–2006). The Madrid region covers 8004 km², divided into 179 municipalities (Figure 1). It has a total of six million inhabitants, half of whom live in the Madrid municipality (Table 1). This region has experienced an extraordinary dynamism in recent years. Between 2000 and 2006, its population rose by almost one million from 5.2 to 6 million.

![Figure 1. Region of Madrid, municipalities, urban land and road network.](image1.png)

**Table 1. Population growth in the Madrid region by rings.**

| Madrid region                | Rings  | 2000     | 2006     | Difference (Abs.) | Difference (%) |
|------------------------------|--------|----------|----------|-------------------|----------------|
| (1) Municipality of Madrid   |        | 2,882,860| 3,128,600| 245,740           | 8.5            |
| (2) Suburban municipalities  |        | 1,806,396| 2,136,315| 329,919           | 18.3           |
| (3) Hinterland               |        | 516,152  | 743,268  | 227,116           | 44.0           |
| Total                        |        | 5,205,408| 6,008,183| 802,775           | 15.4           |

Source: Statistics Institute of the region of Madrid.
inhabitants, an increase of 15.4%. The main areas of absolute growth were found in the ring of suburban municipalities. In relative terms, growth was greater in the more peripheral rings. This dispersion of the population was also accompanied by an intense process of job decentralization.

This population dynamics has been accompanied by a considerable expansion of urban land. Between 2000 and 2006, almost 15,000 ha was converted to urban land. This meant a relative growth of 17.4%, which was higher than the population growth (15.4%). New developments were therefore not only more peripheral but also less dense.

The expansion of urban land has been highly influenced by distance from the city centre and by the road network, especially high-capacity roads around which most new development has taken place. The motorway network has a radio-concentric pattern: 12 radial motorways converge on the Madrid municipality, and this is encircled by three orbital motorways which link the radial roads to each other (Figure 1).

Sources and methodology

Sources

In order to analyse the limitations of CLC for studying urban land-use dynamics, the MLU geodatabase, with its higher resolution and greater accuracy, has been taken as mapping reference.

CLC is a European project managed by the EEA (European Environmental Agency). Its function is to collect, coordinate and homogenize information on the state of the environment and natural resources in the European Union (resolution CE/338/85, 27 June 1985). The main objective of the project is to collect land-use/land-cover data and homogenize the data-gathering processes for the whole European territory. This allows comparable data to be used to apply policies and environmental protection criteria to all countries taking part in the project.

Until now, CLC has been carried out for three years: 1990, 2000 and 2006. The CLC scale is 1:100,000. The building of this geodatabase is based on two processes: on the one hand, satelite and image processing; on the other hand, interpretation and mapping through the integration of automatic classification methods and remote sensing with subsequent visual interpretation, generally known as computer-assisted photo-interpretation (EEA, 2000).

In its addendum 2000, the CLC technical guide (EEA, 2000) warns that there are certain limits to this semiautomatic process of generating spatial information. First, one should be aware that the image quality and contents can vary due to factors such as atmospheric conditions, technical image quality and relief. Second, the differences between land cover, sensu stricto, and functional intrusion from land use make it difficult to set up such an automated conversion. Third, there is an essential difference between pixel-based classification and human holistic interpretation capacity. The latter can map spatial organization of pixels to higher level spatial patterns or objects and allows pixels to be grouped into heterogeneous or composite CLC classes. Bearing the above in mind, it is clear that, with close monitoring of the conversion process, one can obtain valuable CLC results (EEA, 2000).

For its part, the MLU contains information on urban land use in the Madrid region for 2000, 2006 and 2009. Data were collected on a highly detailed scale, with the basic reference scale for mapping established as 1: 10,000 and the minimum mapping unit being 1 ha. As the technical processes did not include automatic or computer-assisted classification
tasks, these were carried out entirely by photo-interpretation of high-resolution (0.5 m) aerial orthophotographs, supported by large-scale cartographic and cadastral information (1:5000 and 1:1000, respectively). MLU has been repeatedly checked and validated by experts in land-use change in Madrid. For these reasons, MLU clearly represents a better available cartographic data set for Madrid urban land use. It is also true that MLU focuses only on land use and does not mix land use and land cover like CLC.

Methodology
In order to compare CLC and MLU, the artificial land classification of each of the two sources was analysed. In level 3 of class 1 (artificial land), CLC has 11 categories (continuous urban fabric, discontinuous urban fabric, industrial or commercial units, road and rail networks and associated land, port areas, airports, mineral-extraction sites, dump sites, construction sites, green urban areas and sport and leisure facilities). MLU, however, classifies urban land use into 12 categories (residential multi-household, residential single-household, industrial, educational facilities, public health facilities, other facilities, office and retail, services, infrastructures, urban green, sport and leisure and airports). The creation of new classes containing the categories of both data bases was found to be possible and this led to the categories being reclassified (Table 2). By means of this process of harmonization, a new classification was obtained that enabled both data bases to be compared. The new classification consisted of seven classes of urban land use (urban residential, industrial/commercial/services, infrastructure, airports, sport and leisure, urban green and non-urban land) (Table 2).

The effects of the use of different scales (CLC, 1:100,000; MLU, 1:10,000) on data quality can quickly be deduced from a simple visual analysis of a detail of the information on land use from each of the sources in accordance with the new classification (Figure 2). The limits of the mapped elements are less precise with CLC (see, for example, the airport). The CLC polygons classed as urban residential mask small elements of other uses (for example, urban green or industrial/commercial/services) because the minimum mapping unit for CLC is 25 ha. On the other hand, transport infrastructures are under-represented

| Level 3 CLC classes (Class 1) | New classes (Harmonization) | MLU classes               |
|-------------------------------|-----------------------------|---------------------------|
| Continuous urban fabric       | Urban Residential           | Residential multi-household|
| Discontinuous urban fabric    |                             | Residential single-household|
| Industrial and commercial     | Industrial/Commercial/      | Industrial                 |
| units                         | Services                    | Office and retail facilities|
| Road and rail network and     | Infrastructure              | Services                   |
| associated land               |                             | Infrastructure             |
| Airports                      | Airports                    | Airports                  |
| Sport and leisure facilities  | Sport & Leisure             | Sport and leisure facilities|
| Green urban areas             | Urban Green                 | Urban green                |
| Dump sites                    | Non-Urban Land              | Non-urban land             |
| Construction sites            |                             |                           |
| Mineral-extraction sites      |                             |                           |
| Other non-urban land          |                             |                           |

Source: CLC, MLU.
with CLC because of specific digitization rules for transport infrastructure (>100 m wide). The effects of scale and cartographic generalization are also clearly seen when changes to urban land between 2000 and 2006 according to CLC and MLU are compared (Figure 3).

In order to assess CLC data quality during the monitoring of urban land-use dynamics, MLU is used as a mapping reference. The first step is to calculate the area of urban land (in total and by class) according to each of the databases in 2000 and 2006. To evaluate the accuracy of CLC, the layers are overlain: CLC2000 and MLU2000 layers on the one hand, and CLC2006 and MLU2006 layers on the other.

A similar analysis is then carried out on changes in land use between 2000 and 2006. The data layer of CLC2000/2006 changes (land-cover changes between the years 2000 and 2006) is generated by overlaying the CLC2000 and CLC2006 data layers. Similarly, the data layer of MLU2000/2006 changes is generated by overlaying the MLU2000 and MLU2006 data layers. The area that changes to urban between 2000 and 2006 (in total and by class) is calculated according to each source (CLC2000–2006 and MLU2000–2006). The accuracy of the CLC2000/2006 changes is measured in relation to the mapping reference (MLU2000/2006). To do this, both data layers of changes are overlain.

Confusion matrices are used to evaluate the spatial coincidence of the elements classified in CLC with respect to those in MLU. This comparison is generally made with raster layers, so counts are done by pixels. In this case, as the two databases were in vectorial format, it was decided to maintain this format to avoid introducing another source of error (vector-raster conversion). The comparison was therefore done directly in hectares. In these confusion matrices, the rows correspond to the mapping to be evaluated and the columns to the reference mapping (in our case, CLC and MLU, respectively).

The values of the main diagonal of the matrices represent the area that coincides spatially in both layers. For example, in Table 5, there was a spatial coincidence of

Figure 2. Detail of urban land use according to CLC and MLU, 2000. Sources: CLC, MLU.

![Detail of Corine Land Cover, 2000.](image)
![Detail of Madrid Land Use, 2000.](image)
Figure 3. Detail of changes to urban land between 2000 and 2006 according to CLC and MLU. Sources: CLC, MLU.

2489 ha for ‘Airport’ land use between the two databases in 2000. Values outside this diagonal represent the area that does not coincide spatially for each land use. For example, 361 ha classified by CLC in 2000 as ‘Airport’ correspond in reality to the ‘Industrial/Commercial/Services’ class according to MLU (commission error); inversely, 165 ha that CLC classifies as ‘Industrial/Commercial/Services’ in 2000 really correspond to the ‘Airport’ class according to MLU (omission error). The commission error column shows these errors as percentages of the total area classified by CLC in that class, whereas in the omission error row they are shown as a percentage of the real area of that class according to MLU (mapping reference). Thus, for example, 25.7% of the total area classified by CLC as ‘Airport’ really corresponds to other classes (commission errors), and 12.9% of the area actually occupied by the ‘Airport’ class has not been classified as such by CLC (omission errors).

The overall classification accuracy for each year is computed by dividing the sum of all the correctly classified hectares (diagonal of the confusion matrix) by the total number of hectares in the confusion matrix. This value tends to overestimate classification accuracy, since it overlooks the proportion of random agreement between data sets (Congalton & Green, 1999). Because of this, the kappa coefficient was also computed. This is basically the proportion of agreement after chance agreement has been excluded (Kiage, Liu, Walker, Lam, & Huh, 2007).

Results

Urban land-use analysis for 2000 and 2006: CLC vs. MLU

First, the urban land area was calculated for both sources of data. As Table 3 shows, CLC registers 1424 ha less urban land than MLU in 2000 and 2946 ha less than MLU
Table 3. Urban land area by use (ha) according to CLC and MLU (ha).

| Land use (2000)                             | CLC  | MLU  | Differences | Differences (%) |
|--------------------------------------------|------|------|-------------|-----------------|
| Airport                                    | 3350 | 2858 | 492         | 17.2            |
| Sport & Leisure                            | 2490 | 3325 | −835        | −25.1           |
| Industrial/Commercial/Services             | 14,189 | 14,343 | −154       | −1.1            |
| Infrastructure                             | 5348 | 7761 | −2413       | −31.1           |
| Urban Residential                          | 56,087 | 51,842 | 4245       | 8.2             |
| Urban Green                                | 3607 | 6366 | −2759       | −43.3           |
| Total                                      | 85,071 | 86,495 | −1424      | −1.6            |
| Land use (2006)                            |      |      |             |                 |
| Airport                                    | 5201 | 5650 | −449        | −7.9            |
| Sport & Leisure                            | 3186 | 3900 | −714        | −18.3           |
| Industrial/Commercial/Services             | 14,578 | 16,947 | −2369      | −14.0           |
| Infrastructure                             | 8459 | 10,197 | −1738      | −17.0           |
| Urban Residential                          | 63,894 | 57,766 | 6128       | 10.6            |
| Urban Green                                | 3906 | 7710 | −3804       | −49.3           |
| Total                                      | 99,224 | 102,170 | −2946      | −2.9            |

Source: MLU, CLC.

in 2006. This means that CLC is relatively reliable when it comes to finding the total amount of urban land, as this amount is underestimated by only 1.6% for 2000 and 2.9% for 2006.

The differences between CLC and MLU in calculations of the area for each urban use reflect what is expected with relation to CLC digitization rules; it is because of these rules (minimum mapping unit and mapping specific digitization rules for transport infrastructure) that CLC overestimates areas of residential use and tends to underestimate areas with other urban uses (Table 3 and Figure 2).

Table 3 details the amount of area occupied by urban land and its breakdown by class, but no information is given on the thematic accuracy of the elements mapped by CLC. To obtain this information for each reference year (2000 and 2006), the two geodatabases (CLC and MLU) were overlain, and the results obtained were analysed using confusion matrices, which revealed the extent to which CLC coincides spatially with the reality expressed by MLU as mapping reference.

Table 4 shows the 2000 and 2006 confusion matrices, with urban land and non-urban land being the only classes considered. Overall classification accuracy was relatively high (95.4% and 95.6%, respectively). However, these values are greatly influenced by the high number of hectares in the non-urban class. When adjusted for chance agreement, kappa statistic values are considerably lower (0.76 and 0.80, respectively). There is quite a high rate of commission and omission errors. Of the hectares classified by CLC as urban in 2000, 20.7% (17,640 ha) were actually non-urban (commission error), while 22.0% (19,067 ha) of the hectares that were really urban were classified as non-urban by CLC (omission error). In 2006, there were rather fewer commission and omission errors (16.2% and 18.6%, respectively). The difference between commission and omission errors explains the underestimation of CLC urban land shown in Table 3 (1424 ha in 2000 and 2946 ha in 2006).

Table 5 shows the confusion matrices differentiating urban land uses. The kappa values (0.69 in 2000 and 0.73 in 2006) are lower than those registered in Table 4 (0.76 and 0.80, respectively) because of the breakdown of urban land by class. The
Table 4. Confusion matrix of CLC and MLU (reference mapping): urban land vs. non-urban land 2000 and 2006 (ha).

|                  | Urban  | Non-Urban | Row total (ha) | Commission error (%) |
|------------------|--------|-----------|----------------|----------------------|
| **CLC (2000)**   |        |           |                |                      |
| Urban            | 67,431 | 17,640    | 85,071         | 20.7                 |
| Non-Urban       | 19,067 | 696,235   | 715,299        | 2.7                  |
| Column total (ha)| 86,495 | 713,875   | 800,370        |                      |
| Omission error (%)| 22.0   | 2.5       |                |                      |
| Overall accuracy |        |           |                | 95.4                 |
| Overall kappa statistics | | | | 0.76 |
| **CLC (2006)**   |        |           |                |                      |
| Urban            | 83,132 | 16,092    | 99,224         | 16.2                 |
| Non-Urban       | 19037  | 682,108   | 701,146        | 2.7                  |
| Column total (ha)| 102,169| 698,201   | 800,370        |                      |
| Omission error (%)| 18.6   | 2.3       |                |                      |
| Overall accuracy |        |           |                | 95.6                 |
| Overall kappa statistics | | | | 0.80 |

Source: CLC, MLU.

Commission and omission error rates are very high. Commission errors are higher than 40% in the Sport & Leisure, Industrial/Commercial/Services and Infrastructure classes; omission errors are higher than 50% in these same classes and in the Urban Green class.

As expected, there are more omission errors than commission errors in most of the classes (Sport & Leisure, Industrial/Commercial/Services, Infrastructure, Urban Green), but the opposite is true for Urban Residential and Non-Urban classes. This is because areas belonging to the above land uses are more likely to be mapped by CLC as Urban Residential or Non-Urban than for the opposite to occur. For example, of the 56,089 ha that CLC maps as Urban Residential in 2000, 1179 ha are really Sport & Leisure, 2529 Industrial/Commercial/Services, 2110 Urban Green and 1138 Infrastructure (commission errors), whereas in the opposite direction errors are only 295, 1600, 287 and 385 ha, respectively (omission errors).

There are basically two sources of error that explain the lower accuracy of CLC with respect to MLU. The first is the 1:100,000 digitization scale: the limits of the CLC polygons do not coincide exactly with those of MLU (taking the latter as the real limits of the mapped elements). This affects both commission and omission errors. The second source of error derives from the digitization rules and affects omission errors more than commission ones. Small polygons (<25 ha) of Sport & Leisure, Urban Green, Industrial/Commercial/Services and Infrastructure (<100 m wide) classes are not explicitly mapped by CLC, which includes them in the Urban Residential and Non-Urban classes. CLC does not give sufficient detail on urban areas because it includes small polygons of different uses in the Urban Residential category and does not register small-scale developments (sprawl) in non-urban areas. The first does not affect the total calculation of urban land, but the second leads to urban land being underestimated in CLC, something already detected in Table 3.
Table 5. Confusion matrix of CLC and MLU (reference mapping): urban land classes 2000 and 2006 (ha).

| MLU (reference mapping) | Airport | Sport & Leisure | Industrial/Commercial/Services | Infra-structure | Urban Residential | Urban Green | Non-Urban | Row total (ha) | Commission error (%) |
|------------------------|---------|----------------|-------------------------------|----------------|------------------|------------|-----------|----------------|----------------------|
| CLC (2000)             |         |                |                               |                |                  |            |           |                |                      |
| Airport                | 2489    | 0              | 361                           | 24             | 28               | 0          | 449       | 3350           | 25.7                 |
| Sport & Leisure        | 2       | 1275           | 70                            | 9              | 295              | 151        | 688       | 2490           | 48.8                 |
| Industrial/Commercial/ | 165     | 177            | 6997                          | 655            | 1600             | 270        | 4325      | 14,189         | 50.7                 |
| Services               |         |                |                               |                |                  |            |           |                |                      |
| Infrastructure         | 2       | 37             | 289                           | 2915           | 385              | 154        | 1566      | 5348           | 45.5                 |
| Urban Residential      | 72      | 1179           | 2529                          | 1138           | 39,074           | 2110       | 9986      | 56,087         | 30.3                 |
| Urban Green            | 0       | 52             | 54                            | 71             | 287              | 2516       | 627       | 3607           | 30.2                 |
| Non-Urban              | 129     | 604            | 4044                          | 2949           | 10,174           | 1165       | 696,235   | 715,299        | 2.7                  |
| Column total (ha)      | 2858    | 3325           | 14,343                        | 7761           | 51,842           | 6366       | 713,875   | 800,370        |                      |
| Omission error (%)     | 12.9    | 61.6           | 51.2                          | 62.4           | 24.6             | 60.5       | 2.5       |                |                      |
| Overall accuracy       | 93.9    |                |                               |                |                  |            |           |                |                      |
| Overall kappa statistics| 0.69    |                |                               |                |                  |            |           |                |                      |
| CLC (2006)             |         |                |                               |                |                  |            |           |                |                      |
| Airport                | 4594    | 0              | 360                           | 22             | 19               | 0          | 206       | 5201           | 11.7                 |
| Sport & Leisure        | 2       | 1636           | 122                           | 29             | 493              | 204        | 701       | 3186           | 48.7                 |
| Industrial/Commercial/ | 128     | 218            | 8700                          | 718            | 1672             | 323        | 2819      | 14,578         | 40.3                 |
| Services               |         |                |                               |                |                  |            |           |                |                      |
| Infrastructure         | 94      | 38             | 419                           | 4528           | 498              | 210        | 2671      | 8459           | 46.5                 |
| Urban Residential      | 72      | 1417           | 2811                          | 1374           | 46,393           | 2744       | 9082      | 63,894         | 27.4                 |
| Urban Green            | 0       | 67             | 69                            | 117            | 243              | 2797       | 613       | 3906           | 28.4                 |
| Non-Urban              | 760     | 524            | 4465                          | 3410           | 8446             | 1432       | 682,108   | 70,1146        | 2.7                  |
| Column total (ha)      | 5650    | 3900           | 16,947                        | 10197          | 57,766           | 7710       | 69,8201   | 80,0370        |                      |
| Omission error (%)     | 18.7    | 58.1           | 48.7                          | 55.6           | 19.7             | 63.7       | 2.3       |                |                      |
| Overall accuracy       | 93.8    |                |                               |                |                  |            |           |                |                      |
| Overall kappa statistics| 0.73    |                |                               |                |                  |            |           |                |                      |

Source: CLC, MLU.
Measuring the CLC capacity and accuracy to detect changes from non-urban to urban land uses

One of the uses of LULCC geodatabases is monitoring of the expansion of urban land. This subsection assesses the capacity of CLC to calculate the total area of land converted from non-urban to urban in the period under analysis (2000–2006) and measures the accuracy of CLC for locating these changes. First, measurements of land-use changes were taken with CLC2000–2006 and MLU2000–2006. Second, the two layers of changes in land use (CLC2000–2006 and MLU2000–2006) were overlain in order to obtain confusion matrices, using MLU as mapping reference.

The total number of hectares converted to urban land between 2000 and 2006 was somewhat higher with CLC (16,401) than with MLU (15,094) (Table 6). However, the differences vary in the different classes. For example, changes to Urban Residential are overestimated by CLC, as they include small polygons that in reality change to other urban uses, such as polygons in the Industrial/Commercial/Services class, which are logically underestimated by CLC. More surprising are changes to the Infrastructure class, which are overestimated by CLC. This is due to the fact that roads built between 2000 and 2006 in the area under study were mainly high-capacity roads and these are represented with excessive width, precisely by CLC, because of the size of the pixels, resulting in an overestimation of the area that converted to this class between the two dates. Changes in the Airport class are due to the expansion of Madrid-Barajas airport and were represented differently by CLC and MLU: CLC demarcated the runway area, whereas MLU also mapped other new tracts of land destined for airport use (Figure 3).

The confusion matrix for CLC and MLU contributes complementary information on changes from non-urban to urban land (Table 7). According to CLC, 16,427 ha was converted from non-urban to urban, while MLU registered 15,809 ha. Greater differences are found in changes in the opposite direction, that is, from urban to non-urban land use; for example, what was an industrial complex in 2000 was in the process of demolition in 2006. By their very nature, changes of this type are uncommon in land-use evolution. CLC calculates 2,275 ha as against the 135 ha for MLU (Table 7). Most of these unusual changes in CLC are therefore attributable to errors. If this type of change is also considered (Table 7), the net increase in urban land would be 14,152 ha according to CLC (16,427 − 2275) and 15,675 ha (15,809 − 135) according to MLU, consistent with the data in Table 3.

There is little divergence between the two sources in the total number of hectares that change to urban land, but this becomes much greater when locational accuracy for changes in urban land use is measured. The commission error rises to 58.0% and the omission error

Table 6. Changes from non-urban to urban land uses between 2000 and 2006 according to CLC and MLU.

| Land use change (2000–2006) | CLC | MLU | Differences | Differences % |
|-----------------------------|-----|-----|-------------|----------------|
| Airport                     | 2059| 2763| −704        | −25.5          |
| Sport & Leisure             | 821 | 549 | 272         | 49.5           |
| Industrial/Commercial/Services | 1834| 2669| −835        | −31.3          |
| Infrastructures              | 3001| 2392| 609         | 25.5           |
| Urban Residential           | 8237| 5467| 2770        | 50.7           |
| Urban Green                 | 449 | 1254| −805        | −64.2          |
| Total                       | 16,401| 15,094| 1306        | 8.7            |

Source: CLC, MLU.
Table 7. Confusion matrix of CLC and MLU (reference mapping): changes from non-urban to urban land use 2000–2006 (ha).

| Changes 2000–2006 | MLU (reference mapping): |
|-------------------|--------------------------|
|                   | Non-urban – Urban | Urban – Urban | Urban – Non-Urban | No change | Row total (ha) | Commission error (%) |
| Non-Urban – Urban | 6.892               | 21            | 6                  | 9.509     | 16.427         | 58.0                 |
| Urban – Urban     | 650                 | 9             | 1                  | 2.396     | 3.055          | 99.7                 |
| Urban – Non-Urban | 64                  | 0             | 15                 | 2.195     | 2.275          | 99.3                 |
| No change         | 8.202               | 112           | 113                | 770.186   | 778.613        | 1.1                  |
| Column total (ha) | 15.809              | 141           | 135                | 784.285   | 800.370        |                      |
| Omission error (%)| 56.4                | 93.8          | 88.8               | 1.8       |               |                      |

Overall accuracy = 97.4
Overall kappa statistics = 0.39

Note: aChanges between classes of urban land use (for example, from Industrial/Commercial/Services to Urban Residential).
Source: CLC, MLU.

to 56.4% (see Table 7). These errors are much higher than those registered in the static analysis (Table 4), due to the fact that the polygons that change to urban land are much smaller than the 2000 and 2006 urban land polygons.

The confusion matrix considering different uses of urban land (Table 8) offers an overall kappa statistic of only 0.40 and very high rates of commission and omission errors for almost all classes. With the exception of the Airport class, these are all above 50%. As occurs in the static analysis, commission error (74%) is greater than omission error (60.9%) in the Urban Residential class. In addition to digitization errors attributable to the CLC scale, it should also be taken into consideration that in this class CLC includes small polygons that change to other non-urban land classes, such as Industrial/Commercial/Services (commission errors), and that it does not show small developments of less than 25 ha (omission errors). On the other hand, with respect to static analysis, the opposite occurs in the Infrastructure class, that is, commission error (64.8%) is greater than omission error (55.8%). Land bordering high-capacity roads is mapped as Infrastructure by CLC, when, for the most part, it really belongs to the No Change class (commission errors). There are also new infrastructures (like the R2 motorway) that have not been registered by CLC (omission errors). A significant proportion of the changes to Sport and Leisure included in CLC are really changes to Urban Green. Changes to the Urban Green class are not very reliable and here the rate of spatial coincidence between the two sources is very low, not only because of the small size of these polygons but also because of the unreliability of identifying these land areas from satellite images.

Final remarks
The world LULCC geodatabase capacity for monitoring the expansion of urban land is obviously limited. Its scope goes no further than locating areas on a planetary scale where extensive changes towards artificial structures take place. The building of databases on a regional scale, at scales of 1:200,000 or more, as in the case of CLC, has allowed researchers to fill a significant gap in spatial information in studies related to urban land-use dynamics. However, more detailed spatial information on urban land use is required for local studies, with scales even lower than 1:25,000. The mapping of land use on this
Table 8. Confusion matrix of CLC and MLU (reference mapping): changes from non-urban to urban land classes 2000–2006 (ha). a.

| Changes 2000–2006 CLC: | Airport | Sport & Leisure | Industrial/Commercial/Services | Infrastructure | Urban Residential | Urban Green | No Change | Row total (ha) | Column total (ha) | Omission error (%) |
|------------------------|---------|-----------------|--------------------------------|----------------|-------------------|------------|-----------|----------------|-------------------|-------------------|
| Airport                | 1857    | 0               | 4                              | 0              | 0                 | 199        | 2059      | 2763           | 2763              | 32.8              |
| Sport & Leisure        | 0       | 171             | 15                             | 8              | 84                | 102        | 441       | 549            | 821               | 68.8              |
| Industrial/Commercial/| 0       | 10              | 602                            | 28             | 20                | 10         | 1163      | 602            | 1163              | 67.1              |
| Services               | 92      | 0               | 16                             | 1057           | 34                | 1          | 1800      | 92             | 1800              | 97.8              |
| Infrastructure         | 0       | 52              | 157                            | 126            | 2140              | 228        | 5534      | 52             | 5534              | 74.0              |
| Urban Residential      | 0       | 0               | 0                              | 19             | 31                | 27         | 372       | 0              | 372               | 93.9              |
| Urban Green            | 815     | 315             | 1874                           | 1154           | 3159              | 885        | 770,186   | 315            | 770,186           | 1.1               |
| No Change              | 2763    | 549             | 2669                           | 2392           | 5467              | 1254       | 779694    | 2763           | 779694            | 1.2               |

Overall accuracy = 97.6
Overall kappa statistics = 0.40

Note: a This table does not include changes between land use classes (for example, from Industrial/Commercial/Services to Urban Residential).
Source: CLC, MLU.
scale is already being carried out in the European Union, with a minimum unit of detail of 0.25 ha for all the large metropolitan areas in the EU. However, as this mapping currently contains only one temporal reference, it cannot be used to analyse urban growth at present.

Some cities, like Madrid, have geographical databases with high spatial resolution for different years. The fact that the Madrid database (MLU) has temporal moments coinciding with CLC (2000 and 2006) makes it possible to assess the limitations of CLC for monitoring urban land. This analysis may be of interest to cities in which high-resolution mapping on land use for different years is unavailable, obliging them to use sources with lower resolution, like CLC.

The analyses carried out on the Madrid region show that CLC gives relatively accurate results with respect to the total amount of urban land, both in 2000 and 2006, with underestimations in the order of 2–3%. When the different classes of urban land are differentiated, the results are logically less accurate because of CLC digitization rules: CLC overestimates total urban land area by around 10% and underestimates other categories somewhat more, generally by 10–30%.

The confusion matrices reveal that the spatial coincidence of urban land use between both databases is around 80%. There are rather more omission errors than commission errors, which is consistent with the CLC scale and its digitization rules. CLC does not show small developments of less than 25 ha and linear infrastructures less than 100 m wide. However, when classes of urban land are differentiated, the error rate is considerably higher, generally between 20% and 60%. There are more omission errors than commission errors for most of the classes, but the opposite occurs in the Urban Residential and Non-Urban classes. This is basically due to the fact that, because of its scale, CLC includes small polygons in Urban Residential or Non-Urban areas that really belong to other classes, such as Industrial/Commercial/Services, Infrastructure or Urban Green.

Results obtained in the dynamic analysis (changes to urban use) are discouraging. There is not much divergence between the two sources in the total number of hectares that change to urban land, but the confusion matrix shows higher rates of error than those registered in the static analysis. This is because the polygons changing to urban land are much smaller than urban land polygons in 2000 and 2006. In fact, the commission error rate rises to 58.0% and that of the omission error to 56.4%. When urban land classes are differentiated, errors are greater and more asystematic.

There is no doubt CLC provides highly valuable information on land use and its evolution in European Union countries. Its average scale (1:100,000) is suitable for analysing land-use dynamics at national or even regional level. Its use is inappropriate, however, for local studies on urban land use. In the Madrid region, the results of CLC over the total urban land area in 2000 and 2006 and on the area converted to urban between 2000 and 2006 are relatively accurate. Nevertheless, the confusion matrices show very high error rates, especially in dynamic analyses (changes between 2000 and 2006) and when the different urban land uses are differentiated. Care should be taken in generalizing the Madrid region results to other regions. Given the results obtained in this study, it follows that researchers, planners and decision-makers are recommended to proceed with caution when using this source for analysing urban land use and its dynamics for local approaches.

Acknowledgements

MLU geodatabase was built by researchers from the Department of Human Geography, Universidad Complutense de Madrid with financial support from the Ministerio de Ciencia e Innovación (project TRA2008-06682).
Note
1. Unlike CLC, MLU focuses on differentiating the uses of urban land. This means that artificial structures that are not explicitly of an urban nature, such as dumps, areas of mineral extraction or construction sites, are identified only as ‘non-urban land’ and no further classification is made.

References
Bach, M., Breuer, L., Frede, H. G., Huisman, J. A., Otte, A., & Waldhardt, R. (2006). Accuracy and congruency of three different digital land-use maps. *Landscape and Urban Planning, 78*, 289–299.

Bielecka, E. (2005). A dasymetric population density map of Poland. In *Proceedings of the 22nd Cartographic International Conference, La Coruña, Spain*.

Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: The social and environmental costs of different patterns of urban expansion. *Ecological Economics, 40*, 199–216.

Catalá, R., Bosque, J., & Plata, W. (2008). Análisis de posibles errores en la base de datos CORINE Land Cover (1990–2000) en la comunidad de Madrid. *Estudios Geográficos, 69*, 81–104.

Congalton, R. G., & Green, K. (1999), *Assessing the accuracy of remotely sensed data: Principals and practices*. Boca Raton, FL: Lewis.

EEA. (2000). *CORINE Land Cover technical guide – Addendum 2000* (Technical Report No. 40). Copenhagen: EEA.

EEA, Eurostat, DG Agriculture, DG Environment, and Joint Research Center. (2001). *Towards agri-environmental indicators: Integrating statistical and administrative data with land cover information* (p. 6). Copenhagen: EEA.

Feranec, J., Hazeu, G., Christensen, S., & Jaffrain, G. (2006). Corine Land Cover change detection in Europe (case studies of the Netherlands and Slovakia). *Land Use Policy, 24*, 234–247.

Fuller, R., & Gaston, K. (2009). The scaling of green space coverage in European cities. *Biology Letters, 5*, 352–355.

Gallego, J., & Peedell, S. (2001). *Using CORINE Land Cover to map population density* (p. 6). Copenhagen: EEA.

Goldewijk, K. (2001). Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles, 15*, 417–433.

Guérois, M., & Pumain, D. (2008). Built-up encroachment and the urban field: A comparison of forty European cities. *Environment and Planning A, 40*, 2186–2203.

Hewitt, R., & Escobar, F. (2011). The territorial dynamics of fast-growing regions: Unsustainable land use change and future policy challenges in Madrid, Spain. *Applied Geography, 31*, 650–667.

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: Developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment, 85*, 7–24.

Janssen, S., Dumont, H., Fierens, F., & Mensink, C. (2008). Spatial interpolation of air pollution measurements using CORINE Land Cover data. *Atmospheric Environment, 42*, 4884–4903.

Kiage, L. M., Liu, K.-B., Walker, N. D., Lam, N., & Huh, O. K. (2007). Recent land-cover/use change associated with land degradation in the Lake Baringo catchment, Kenya, East Africa: Evidence from Landsat TM and ETM+. *International Journal of Remote Sensing, 28*, 4285–4309.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelens, A., Bruce, J. W., . . . Xu, J. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change, 11*, 261–269.

Loveland, T., Reed, C., Brown, J., Ohlen, D., Zhu, Z., Yang, L., & Merchant, J. (2010). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing, 21*, 1303–1330.

Siedentop, S., & Meinel, G. (2004). CORINE Land Cover 2000 in nation-wide and regional monitoring of urban land use and land consumption. In *UVA-Texte. 04/04 ISSN 0722-186X, Workshop CORINE Land Cover 2000*, Leibniz-Institut für Ökologische Raumentwicklung e.V, Berlin 2004 (pp. 162–169).

Thinh, N. X., Arlt, G., Heber, B., Hennersdorf, J., & Lehmann, I. (2002). Evaluation of urban land-use structures with a view to sustainable development. *Environmental Impact Assessment Review, 22*, 475–492.
Turner, B. (2001). Land-use and land-cover change: Advances in 1.5 decades of sustained international research. *GAIA, 10*, 269–272.

Turner, B., & Meyer, W. (1994). Global land-use and land-cover change: An overview. In W. Meyer, B. Turner (Eds.), *Changes in land use and land cover: A global perspective* (pp. 3–10). Cambridge: Cambridge University Press.

Turner, B., Skole, D., Sanderson, Sc., Günther, F., Fresco, L., & Leemans, R. (1995). *Land-use and land-cover change* (IGBP/HDP report). IGBP/HDP, 35/7.

Wackernagel, M. (Ed.). (1996). *Our ecological footprint: Human impact on the earth*. Gabriola Island, BC: New Society.