Semantic data harmonisation of land-cover class sets using the FAO/UNEP Land Cover Classification System - Translator Module (LCCS-TM)

Louisa J.M. Jansen [a], Geoff Groom [b], Giancarlo Carrai [c]

[a] Land/natural resources consultant, Via Girolamo Dandini 21, 00154 Rome, Italy.
Tel.: +39-065757719, E-mail: louisa.jansen@tin.it

[b] National Environmental Research Institute, Kalø, Grenåvej 14, 8410 Rønde, Denmark. E-mail: gbg@dmu.dk

[c] Managing director, SVALTEC S.r.l., Via del Campofiore 106, 50136 Florence, Italy. E-mail: gc.carrai@svaltec.it

Abstract

With the growing interest in modelling and monitoring environmental change, it becomes important to re-examine existing class sets in order to harmonise the data. Data harmonisation has different dimensions and in this paper semantic data harmonisation is examined using two groups of class sets, totalling 190 classes. These classes have been translated into the terminology of the FAO/UNEP Land Cover Classification System (LCCS) and subsequently compared in the LCCS Translator Module (TM) tool. Translation encountered problems of the use of other than land-cover parameters in the definition of classes, differences in threshold values and the use of ranges of values to define boundary conditions that do not correspond to the LCCS implemented parameters. The analysis of the use of the LCCS classifiers in the class sets shows that some parameters were used very little in the data translation. Within the LCCS-TM, classes were compared and their similarities calculated but the implemented algorithm seems to be strongly biased. Similarity values are calculated per class but not per class set. Furthermore, it is impossible to show how good a data translation from the original class into LCCS terminology is, and accordingly suggestions for improvement are made.

Keywords: Data harmonisation; Land cover; Class comparison; Landscape monitoring.

1. Introduction

Land cover defined as "*the observed (bio) physical cover on the earth's surface*" (Di Gregorio and Jansen, 2000) is widely perceived as an important component of environmental and ecological monitoring. Many countries, and other geographic entities, have undertaken survey of land cover but often on the basis of a particular nomenclature. Nowadays emphasis is shifting from static land-cover mapping towards more dynamic environmental modelling in order to understand the past, monitor the present situation and to predict future trajectories (Lambin et al., 2000; McConnell and Moran, 2001; Dolman et al., 2003). Results from different surveys may need to be harmonised over time and space (e.g., for trans-boundary air and water pollution), or existing information may be used to verify new results (e.g., urbanisation, desert locust monitoring). With the growing interest in modelling environmental change, it is important to re-examine existing land-cover data sets and attempt to harmonise them to make comparisons between countries and to compile time series with which to analyse the change dynamics and detect trends. Data harmonisation will be required as it is unrealistic to work with new standardised class sets, with major financial and intellectual investments having been made in existing class sets and survey programmes that use established methods of classification (UNEP/FAO, 1994; Wyatt et al., 1997; Wyatt and Gerard, 2001). Harmonisation is mentioned in most data collection efforts but is hardly ever put into practice, probably because of the limited compatibility and comparability between data sets. Data harmonisation is a priority for the European Commission (e.g., INSPIRE initiative) and the UN (e.g., UNCED's Agenda 21 and the World Summit on Sustainable Development (Johannesburg 2002)).

In the era of geo-informatics, spatial data harmonisation involves consideration of several distinct aspects that should be considered in a comprehensive manner. These aspects are described in brief below:

1. The *geometric aspect* of the data set consists of the description of the form of the entities through geographic primitives or through a structured geometry (e.g., topology). In general this aspect considers cases of different representations of the same object. For instance, one can imagine a road network represented by polygons of the road surface, or as a network of road axes and nodes, or as an information level where the road is represented by the sequence of borders such as walls and façades of buildings (e.g., cadastre). These three ways of representing a road network each follow a different set of conceptual and technical practices. If data harmonisation is required it will be difficult, if not impossible, to obtain a continuous spatial basis with the same level of confidence unless the geometric aspect is also harmonised. However, data sets that are represented by either polygons or rasters, such as many thematic data sets, do not represent significant problems of geometric harmonisation since usually these can be restructured using topological procedures.
2. The *distribution* and *spatial aspects* describe the dimensions of the geometry (i.e. two dimensions for areas, one dimension for lines and no dimension for points) in relation to the scale and projection in a geodetic reference system. Differences in scale can be overcome by geometric generalisation but this may imply loss of information; generalisation means also reorganisation of the semantic attributes (see point 4). For instance, elaborating the above-mentioned example one could imagine a matching of road maps of two neighbouring countries. In country A, the roads are depicted at scale 1:5,000 by polygons projected in a local coordinate system, whereas in country B the roads are depicted at scale 1:100,000 by lines projected in UTM WGS84. Data harmonisation in such a case should consider three aspects: (1) how to depict the roads (e.g., as polygons or lines), (2) the scale to be adopted and (3) the geodetic reference system to be used.
3. The *temporal aspect* of the data sets is to be considered because certain themes undergo more changes with time than others and data harmonisation between class sets covering the same subject but from different periods may lead to lower levels of confidence (e.g., the first CORINE land-cover data set for Europe spans a period of 10 years).
4. The *semantic aspect* depends not only on the type of coding system or the nomenclature applied, but it depends primarily on the definitions used because these imply the parameters used in the formation of classes. Comparison of these parameters permits determination of whether a change in coding between data sets through a semantic translator will be sufficient for harmonisation. Otherwise it is necessary to reclassify the nomenclatures into a third system, a so-called reference system that functions like a bridge between two data sets; each class in the original data set will find its more or less corresponding class in the reference system. Class descriptions contribute to the definition of boundary conditions that should be applied unequivocally and consistently when establishing correspondence between class sets in order to avoid errors in data interpretation. The level of confidence with which such class correspondence is established is highest when the same parameters have been applied; differences in the applied parameters, and thus in boundary conditions, produce lower confidence levels. Complete correspondence is not always obtainable when harmonising data, thus there is a need to establish rules in order to reach the highest level of confidence possible.

This paper concentrates on the semantic aspect of data harmonisation. Two different groups of class sets have been used. The first group includes both specific purpose and general-purpose class sets from the Nordic countries within Europe, whereas the second group includes general-purpose class sets only from Afghanistan and Lebanon. Comparisons are made within each group, not between them. The two groups of class sets comprise:

1. Five Nordic class sets used in relation to environmental resource assessment and landscape monitoring programmes comprising a total of 152 classes (Jansen, 2004):
   - Area Information System - Land-Cover Map (AIS-LCM) of Denmark;

- Land-Cover Classification Scheme (EELC) of Estonia;
- Monitoring Agricultural Landscapes (3Q) of Norway;
- Digital Field Basis Map (Digitalt Markslagskart, DMK) of Norway; and
- Land-Cover Data (SMD) of Sweden.

2. Two class sets from countrywide land-cover inventories of the post civil war situation in Lebanon (FAO, 1991) and Afghanistan (FAO, 1999) together comprising 38 classes. Land cover has been mapped at 1:50,000 scale in Lebanon and at 1:100,000 in Afghanistan, in both cases by visual interpretation of hardcopy Landsat TM false colour composites and KFA-1000 space photographs.

The semantic data harmonisation has been performed using a specific bridging tool, the FAO/UNEP Land Cover Classification System – Translator Module (LCCS-TM). Without a semantic translator, comparison of n data sets requires $n(n-1)/2$ comparisons to be made. If $n=5$, such as in the first group of class sets, 10 comparisons per class would be needed. The usefulness of the LCCS-TM to act as a reference system was examined together with analysis of the type of problems encountered and solutions applied while finding the corresponding classes for class sets from different countries with similar types of landscapes but different approaches to land cover. At the same time the consistency and use of parameters in the data sets can be examined and the similarity between class sets (intercomparison) or within a class set (intracomparison) is illustrated with a number of examples.

2. The Land Cover Classification System – Translator Module (LCCS-TM) tool

LCCS (Di Gregorio and Jansen, 1998 and 2000) was developed by FAO and UNEP out of the necessity to have a consistent and pragmatic methodology when collecting land-cover data in several countries representing different types of environments. Subsequently the methodology and its software application have been endorsed by the Land Use and Cover Change (LUCC) core project of the International Geosphere Biosphere Programme (IGBP) and International Human Dimensions Programme on Environmental Change (IHDP) (McConnell and Moran, 2001). As basis of a harmonisation strategy LCCS-TM is recommended by the Global Observations of Forest Cover – Global Observation of Land Dynamics (GOFC-GOLD) and the Global Terrestrial Observing System (GTOS).

LCCS has been applied for country-wide land-cover data collection in 10 East African countries, in Mozambique and a number of Central and Eastern European and CIS countries (e.g., Albania, Azerbaijan, Bulgaria, Moldova and Romania) (Travaglia et al., 2001; Jansen et al., 2003b, 2006a and 2006b). Several projects and case studies using LCCS have focussed on land-cover change dynamics (e.g., Albania, Mozambique, Niger and Senegal) (Mahamadou, 2001; Jansen et al., 2003a, 2006a and 2006b; CSE, 2005; Jansen and Ndiaye, 2006) and LCCS has been used as the basis for map legends in GLC2000 (Mayaux et al., 2004). LCCS not only provides a systematic method to classification based upon a parametric approach, but also through its Translator Module (TM) a systematic procedure for the comparison of land-cover class sets. In the TM the parameters used for definition of a land-cover class can be used to "dissect" existing land-cover classes. The selected parameters in LCCS are based upon a thorough analysis of numerous existing nomenclatures (Jansen and Di Gregorio, 2002). The LCCS parameters allow definition of thousands of land-cover classes at various levels of detail; consequently the user is not limited by a finite number of global or national categories. The software application assists the user in the definition of the appropriate class (Di Gregorio and Jansen, 2000). To find the corresponding class in LCCS, the user is obliged to redefine the classes of his or her class set in the LCCS Classification Module and consequently save them in the LCCS Legend Module before importing the class sets, class by class, into the LCCS-TM. The approach is illustrated in Figure 1.

Insert Figure 1.

For the calculation of the similarity between classes within LCCS-TM, there are a number of rules that influence the computation. Each parameter used in the definition of the classes has the same weight. Weighting is not required since the relative importance of the individual parameters in LCCS is linked to their hierarchical order. The parameters at the top levels of the classification system are those that define broad classes (e.g., the parameters Life Form and Canopy Cover are used to define "Closed trees"); subsequent parameters further refine the defined class (e.g., "Broadleaved deciduous closed trees" or "Multi-layered closed trees"). The top-level parameters are easily identifiable in the field and defined with more accuracy than subsequent parameters. The order of the parameters in this way assists class comparisons because comparison will first relate to the broadly defined land-cover type to which the class belongs and then relate to differences within the land-cover type. For example, once a "Closed trees" is defined the use of additional parameters will make a distinction between for instance "Multi-layered closed trees" and "Closed trees with shrubs". The land-cover type refers to a specific structural-physiognomic arrangement within a land-cover category (e.g., Forest, Woodland, Thicket, Shrubland, Grassland and Sparse Vegetation are distinguished within the land-cover category of (Semi-) Natural Vegetation, whereas Tree Crops, Shrub Crops and Herbaceous Crops are distinguished for Cultivated Areas).

In the similarity computation the values are either 1 (similar) or 0 (dissimilar). Some parameters consist of two levels, i.e. a more general level further divided in detailed options (e.g., the Life Form "Woody" is subdivided into "Trees" and "Shrubs"). For such parameters the arbitrary value of 0.5 (half-similar or half-dissimilar) is introduced (e.g., the similarity between "Trees" and "Woody" equals 0.5). The number of parameter options having a similarity of 0.5 is very limited (Di Gregorio and Jansen, 2000).

During the computation, the software application will look for the same parameter from the reference class and the class being compared and then look at the parameter option selected. This means that if the identical parameter is present in both classes, the second step consists of comparison of the options selected (e.g. if the same option is found twice the parameter is similar, if two different options are found the parameter is dissimilar). A number of examples will be discussed to illustrate the LCCS-TM computation later on in this paper.

3. Difficulties encountered when structuring correspondence between classes

The data translation of the class sets into the corresponding LCCS terminology forced the translators working with the two groups of class sets at times to make compromises from among the choices and options available within LCCS. Although LCCS contains in general more options than the parameters used in the original class sets, the correspondence between original criteria and the LCCS parameters was not always 100 percent. The main problems encountered are discussed below as well as the solutions adopted. None of these solutions, however pragmatic, led to a data set translation that was fully matching the original. Some of the solutions represent merely a basis for further discussion.

3.1 Differences in threshold values of parameters

The main problems encountered while establishing correspondence between original classes and LCCS were related to differences in thresholds, especially crown cover. For instance, the Norwegian Monitoring of Agricultural Landscapes (3Q) uses 25 percent crown cover to distinguish between vegetated and non-vegetated ground. The crown cover value used in the Swedish Land-Cover Data (SMD) for distinction between forest classes is 30 percent. LCCS contains more options for indication of crown cover than most existing class sets but differences of 5 to 10 percent occur between the LCCS thresholds and those used by the class sets examined here (Table 1).

Insert Table 1.

Different thresholds also exist for artificial surfaces such as urban areas. Table 2 provides an overview of the minor differences that exist between LCCS and SMD, which have been ignored here, since a difference of five percent is difficult to establish in practise. Furthermore, differences exist between tree heights used in the various class sets. Table 3 shows the differences between LCCS and SMD and shows that certain tree heights would lead to a different class (e.g., 6m). Such differences have not been ignored.

Insert Tables 2 and 3.

In the Norwegian Digitalt Markslagskart (DMK) class set the thresholds for “Mixed forest” are unusual, with 20 percent crown cover as the lower limit and 50 percent crown cover as the higher limit for needleleaved trees; thus a forest area with 40 percent broadleaved trees is classified as needleleaved forest. In LCCS the thresholds follow the UNESCO vegetation classification (Di Gregorio and Jansen, 2000).

In the DMK class set an area is called “forest” also when it is for the time being without trees. This situation is analogous to the description of cultivated areas without crops in LCCS. Furthermore, if an area is covered with needleleaved trees under a canopy of broadleaved trees, it is classified in DMK as “needleleaved”. This is probably due to the larger economic value of needleleaved trees and therefore of prime interest to the forester. A cultivated area where trees are also present is described as “forest” in the DMK class set, whereas this would be a mixed class in LCCS.

A class such as “Other soil covered land” in the DMK class set was impossible to translate, because this class does not describe the actual land cover. The same is the case with the DMK class “Land with shallow topsoil”. These classes are examples of classes that could not be included in the data translation.

As noted earlier, an important difference in threshold values occurs between LCCS and 3Q for determining whether a class belongs to vegetated or non-vegetated land-cover types. In LCCS the value is very low, i.e. 4 percent, whereas in 3Q it is set at 25 percent. This notable difference may be due to the objectives with which the classes were defined as well as the concept of a vegetated area. Vegetation changes with the seasons and only the permanent life forms, i.e. trees and bushes, remain. Furthermore, if only 4 percent of an area is vegetated it is assumed that the rest of the area is empty, i.e. there are no other structures or occupied surfaces. If this was not so one would have to speak of a mixed class in which the vegetated area is subordinate to other land-cover classes (e.g., bare surfaces). Thus, the definition used by 3Q will either encourage creation in LCCS of mixed classes or else disregard extremely sparse vegetation.

3.2 Occurrence of land-use terminology and environmental parameters

The occurrence of land-use terminology in some classes as well as the importance of the class sets for monitoring of environmental change (e.g., development of land-use patterns) calls for compromises in the data translation. The two Norwegian class sets have a land-use oriented focus: 3Q is a class set developed to monitor agricultural land-use patterns, biodiversity and cultural heritage whereas DMK focuses on the land (e.g., soil) as a resource for agriculture and forestry. LCCS is designed for the description of land cover although in some places land-cover related (management) practices can be accommodated. This is especially true for artificially covered areas such as cultivated and built-up areas. Grasslands that are managed belong to the cultivated areas category of LCCS and some land-cover related cultural practises could be described. Grasslands that are abandoned and invaded by natural vegetation belong to the (semi-) natural vegetation category and land-cover related cultural practises cannot be described. The user could make user-defined parameters to cover the land-use related aspects but these parameters are not standardised and consequently not considered in the Translator Module.

Classes that gave rise to considerable discussion during data translation were “Clear felled areas” and “Burned areas” where the class names refer to a land-use practise or an environmental event. The resulting land-cover type can be manifold, depending on the species invading the terrain after the event and the time that has passed after tree felling or fire. The land-cover class(es) created in LCCS are neutral with respect to any applied interpretation of the land cover, including for example its ecological meaning. Therefore, these phenomena, and also damage due to hail storms or wind (e.g., the use of the “vi” code in 3Q), cannot be accommodated by any other means than adding user-defined parameters.

In the SMD class set, which is a general-purpose land-cover class set with some land-use criteria also present, the classes concerning burned and clear felled areas were translated as a closed woody vegetation type with an added LCCS user-defined parameter explaining that it refers to a clear felled or burned area. This solution is very subjective because it depends heavily on who is making the data translation.

Another SMD class relating to a future cover, “Construction sites”, was difficult to translate because the type of construction and the time before the construction will become visible determine the description of the actual land cover with LCCS parameters. As a result a mix was chosen between built-up areas, unconsolidated and consolidated materials, but this choice is arbitrary. One cannot expect to find a perfect match between an actual and a potential land cover.

In the general-purpose class sets of Lebanon and Afghanistan some classes contain information about the particular environment in which they occur and this information is translated into so-called Environmental Attributes, such as landform or climate (e.g., class “5b. Sparse grassland and forbs in mountains or desertic areas” in Lebanon, classes “4a. Cultivation in flat areas” and “4b. Cultivation in sloping areas” in Afghanistan). Information related to land-use instead of land cover will be lost in the semantic data harmonisation process as described above (e.g., “old fallows” and “abandoned fields” in class “5a Grassland and forbs from open to closed, or abandoned fields, or old fallows in agricultural areas” in Lebanon, and terms like “intensively” and “occasionally cultivated” in classes “3a Intensively cultivated” and “3c Occasionally cultivated” in Afghanistan).

3.3 Mixed classes

In a number of cases the correspondence to the classes and the available options in LCCS version 1 was imperfect and this was especially true when a range was included in the definition (e.g., from open to closed). This occurred in the vegetation types, “Forest”, “Moors and heath land”, “Grass heath” and “Meadow” (SMD class set), and more explicitly in the classes “Grassland and forbs from open to closed, or abandoned fields, or old fallows in agricultural area” and “Scrubland and other types of degenerate forest” of the Lebanon class set and the classes “Degenerate forest/ high shrubs” and “Rangelands (low shrubs/grasslands and forbs (open to closed cover))” of the Afghanistan class set. The forest of DMK can be either closed or open forest in LCCS. To directly translate “Forest”, it should be possible to define a forest category that combines closed and open forest but in LCCS one is forced to create a mixed class to accommodate this range in cover. However, a mixed class may be explained in various ways and it will be difficult for a user of LCCS translated classes to know if the original forest is either closed or open or if there is a range from open to closed. In LCCS version 2, released in 2006, this problem has been solved by adding the option “closed to open”.

In other class names two clearly distinct types of objects co-occur, such as “Fruit trees and berry plantations” in the SMD class set, in various classes of the Danish Area Information System’s Land-Cover Map (AIS-LCM) class set, the class “Field crops and fallow land” of the Lebanon class set and the class “Gardens (ornamental and fruit trees)” of the Afghanistan class set. In such cases various options are possible; taking the SMD class as an example:

- A mixed class of fruit tree plantations with berry plantations is created because due to the mapping scale and/or the arrangement of fields these two types of fields cannot be spatially distinguished.
- A single class is created containing the dominant crop fruit trees with the berries as a second crop because they occur in the same field. In this case it is a single class containing a multiple crop (N.B. it may also be possible that a single class exist in which the berries are dominant over the trees).
- A mixed class is created combining the two above-mentioned options.

The best practise in such cases depends on how the two components are arranged spatially but this information was not available for the above-mentioned class. It may also happen that all options occur in practise but that this is not reflected in the original class set and thus poses problems when translating.

Furthermore, whilst class definitions of a general nature such as the class "Deciduous fruit trees" in the Lebanon data set can be translated without problems, its detailed description "*This sub-category includes mainly apple, pear, peach and apricot orchards*" cannot be translated as this would require in LCCS the specification of the dominant crop type followed by a second and/or third crop. This refinement may be possible at the level of the individual polygon but it cannot be represented in an overall class set translation.

Sparsely vegetated areas in the SMD class set were translated as a mixed class containing unconsolidated materials and herbaceous open vegetation. The definition of "sparsely" as used in the SMD class set is lacking and depending on it, it could be argued that "sparse herbaceous vegetation" should have been selected for the data translation. The translated class gives the impression that there are two elements present: (1) bare areas/bare soils with (2) open vegetation. But the concept of sparse vegetation in LCCS is not the same as a mixed class of bare soil with vegetation.

3.4 Vegetation description concept

The occurrence of lichen-dominated areas with trees cannot be accommodated in LCCS by creation of a mixed class. The concept adopted in LCCS for lichens is very limited and it is impossible to link this life form with any occurring stratification. The same would happen if the user would start with the definition of open trees, in which case the lichen component cannot be added. This is a real limitation for the correct class set translation of lichen-dominated vegetation types in the Nordic countries. Lichens also occur in mires but the combination of a waterlogged area with lichens is also not considered in LCCS.

Mires are a very important feature in Nordic landscapes and they are usually not well represented in nomenclatures. Nowadays it is important to report on mires especially in the context of the UN Wetland Convention. In the Estonian Land-Cover Classification Scheme (EELC) class set the 4th level is used for description of wetlands because the 3rd level, which follows CORINE LC, has no possibility to describe these classes well. LCCS provides enough options to describe the EELC classes apart from those that relate to the presence of lichen.

3.5 Existence of user-defined parameters and codes

Classes containing a user-defined code caused problems when importing them into the Translator Module. The category to which the classes belonged can be identified but the Boolean formula and the class name do not appear in the appropriate boxes of the LCCS-TM when they are processed and hence cannot be saved. An attempt to remove the user-defined codes and adding them after processing of the class worked but the classes as such could not be further used in the similarity assessment because an error message ("Invalid use of Null") was generated. Therefore these codes were moved manually into the environmental attributes boxes, where they did not cause further problems since the attributes are not considered in the computation of the similarities. The decision to move these codes was based upon trial and error as the user is not supposed to move manually any of the codes from one box to another.

4. Semantic data harmonisation results

4.1. Overview of the use of classifiers by the class sets

One of the things to examine as part of class translation is whether the classifiers present in LCCS are the same as those commonly used to represent the translated class set. If this is so, it is also interesting to evaluate how the different classifiers have been used. If particular LCCS classifiers are not used the explanation may simply be that the class set contained only very limited land-cover parameters that could be translated and that the other parameters used belong to the criteria that are not inherent to land cover. Differences in the use of classifiers may also be related to the scale at which the original class set is applied: a small-scale (e.g. 1:1M) mapping class set will not use as many classifiers as a large-scale (e.g. 1:50,000) mapping class set.

The two major LCCS land-cover categories of main interest for terrestrial landscape level monitoring are Cultivated Vegetated Areas and (Semi-) Natural Vegetated Areas. The relevant classes of the five Nordic class sets for those two categories and their use of classifiers are represented in Tables 4 and 5. Table 4 shows the use of the terrestrial Cultivated Area classifiers and also two relevant attributes. It is immediately evident that the AIS-LCM and 3Q class sets use a range of available LCCS classifiers, whereas DMK and SMD only use the classifier Life Form, which is the first encountered and most important classifier within LCCS. The DMK class set could only be translated in part because it is based upon a specific concept (land-use potential) for which no classifiers exist in LCCS. Certain classifiers or attributes have not been used at all or with only very limited use by these five class sets (e.g. Field Size and Crop Type). The interpretation of this could be that the classifier is difficult to apply, not clearly defined and/or explained or is not considered to be of (any) importance. The latter would justify moving such a classifier further down the LCCS hierarchy or ensure that a “skip” option is available, as is indeed the case already within LCCS for Field Size.

Insert Tables 4 and 5.

Table 5 shows the use of classifiers and specific technical attributes for the two LCCS categories of (Semi-) Natural Vegetation, i.e. terrestrial and aquatic or regularly flooded environments (the “≠” means that the classifier is not applicable for one or other of these two LCCS categories). The classifiers Life Form and Cover are obligatory for any class definition. Height is almost always used because the range of values contained in the classifier does not result in any obstacle. The classifier Spatial Distribution is used much less; apparently not much emphasis is placed on the horizontal patterns of vegetation. The classifiers Leaf Type and Leaf Phenology are occasionally used. One of the problems, which occurred in the data translation, was that one cannot skip Leaf Type to define Leaf Phenology only. As a result, if Leaf Type could not be defined, consequently one could not add Leaf Phenology. The almost complete absence of the use of the classifier Stratification is notable. One explanation of this is that the applications for which the Nordic class sets were created are not interested in the vertical arrangement, i.e. layering of the groups of life forms. Or, it may be that due to the climate not very much layering is present in the described vegetation types. Or, possibly the use of the Stratification classifier in LCCS not seen as straightforward and therefore passed-over by the translators? The answer is probably a mixture of these three. In many forests undergrowth is systematically removed by either wild animals or management practises, or many of the described vegetation types are low in height and consequently no or not much layering can be distinguished or at most two layers can be distinguished but this is not reflected in the translated classes. Part of the answer may also be that these class sets are used with remote sensing data applications in which any layering underneath the highest canopy cover cannot be identified on the image or photo.

4.2. Class set intracomparison

The AIS-LCM and the Lebanon class sets have been selected for intracomparison. The aim of class set intracomparison is to quantify the similarity of classes belonging to the same class set, and thereby assess their distinctiveness. The similarity values will give an indication of how similar the physiognomic-structural descriptions of the classes in a class set are. The more similar the classes, the more difficult their correct distinction.

Table 6 shows the results for the intracomparison of the AIS-LCM class set. In the computation, only one element of a mixed class can be selected as the reference class. When selecting the reference class one can set which element of the mixed class will be the reference class (e.g., see division of mixed classes 6, 7, 10 and 11 where the figures 1 and 2 refer to the first or second element of the mixed class). No reference class can be compared to the other element(s) of a mixed class.

Along the diagonal axis one finds in general the highest values, especially in the group of classes belonging to the same major LCCS land-cover category as indicated by the boxes in the table. However, there are also values showing high similarity further from the diagonal axis. Most of these values are linked to the aquatic (semi-) natural vegetation type of class "08-marshland". This class can be regrouped with the terrestrial vegetation classes under the category of (Semi-) Natural Vegetation independent of the environment (e.g., aquatic or terrestrial) in which the vegetation type occurs (dotted line in Table 6). The other, lower (but non-zero) off-diagonal similarity values are caused by the occurrence of similar Life Forms between classes.

Insert Tables 6 and 7.

The intracomparison of the Lebanon class set shows the same patterns: the highest similarity values and found, as expected, along the diagonal axis (Table 7). It is clearly shown that the level of detail in distinction of the crop classes is greater than in any other group of land-cover classes. Higher similarity values are found on the left and right side of the diagonal axis. Sometimes the structural physiognomic description of the classes is identical but the specification of crop type and/or crop cover differs or some classifiers are not specified in the reference class though they appear in the classes being compared. These classes need to be interpreted with great care because of their resemblances. The more similar the land-cover features, the more difficult the correct designation of a class. The difference may be based on a single classifier or the use of additional classifiers. The same applies to the forest classes where a distinction is made based upon leaf type, such as could be more easily detected with multi-date satellite imagery. There are some high values found outside the boxes indicated in Table 7 where one would expect much lower values. An example is the comparison between classes from Cultivated Area(s) and (Semi-) Natural Vegetation where the high similarity can be explained by the coincidence of the classifiers Life Form, Leaf Type and Leaf Phenology with identical options selected.

Comparison of Tables 6 and 7 illustrates that the variation within a land-cover category is very much dependent on the occurrence and definition of the classes present in the class set. The similarity within a category is in general higher than between categories. High similarity values between categories are found when these categories share a number of parameters. It could be helpful in future land-cover survey programmes to intracompare the classes in the preliminary phase of class definition before actual data collection takes place.

4.3. Class sets intercomparison and similarity

The comparison of two class sets of neighbouring countries, such as the 3Q and SMD class sets, gives an indication of how similar land-cover classes are in these class sets. In the 3Q class set several Cultivated Area classes are distinguished whereas the SMD is geared towards description of general-purpose land cover hence the limited number of Cultivated Area classes. The results of this comparison are discussed in detail in Jansen (2004). The pattern of values is basically identical to that described for class intracomparisons: the majority of high values are found within the land-cover categories and several high values are found outside these categories due to the co-occurrence of classifiers.

Of more interest than the full correspondence matrix is how the class comparison works within LCCS-TM. Similarity between Cultivated Area classes and (Semi-) Natural Vegetation classes is due to the occurrence of a limited number of joint classifiers since both land-cover categories are dominated by plants and their vertical and horizontal arrangements. As a consequence, a similarity can be found between for instance graminoid crops with a herbaceous or graminoid type of (semi-) natural vegetation.

Insert Tables 8 and 9.

Tables 8 and 9 show how the similarity is calculated and illustrate at the same time that the reference class greatly determines the type and number of classifiers in the computation. When there are few classifiers (Table 8) the resulting similarity values are high, whereas with more classifiers in the reference class (Table 9) the resulting similarities are much lower.

Two classes of the Lebanon and the Afghanistan class sets respectively belonging to the same land-cover category (Cultivated Area), thus using the same set of classifiers, are compared in Table 10. The classifiers present in the reference class, i.e. the first class, are determining the computation because they influence the type and number of classifiers compared. In this example one can also see that a so-called modifier, i.e. a parameter that modifies a classifier, is included in the parameters but not considered in the computation because no modifier is present in the reference class. One should note that if a modifier had been present the parameter Field Size would have entered twice into the similarity computation. Crop type and Crop Cover belong to the so-called LCCS attributes, which are optional, and never included in the similarity assessment because only land-cover classifiers are used in the computation. Computation between (Semi-) Natural Vegetation and Cultivated Area classes is more complicated as they have some classifiers in common but ordered differently so that the computation takes more time in order to establish whether or not the same classifiers occur and the options selected are comparable. Table 11 demonstrates that although many classifiers may be present, the number and type of classifiers of the reference class determine the similarity. The attribute Landform is shared but it is not considered in the computation.

Insert Table 10 and 11.

Tables 8 - 11 illustrate a bias that is present in the algorithm implemented in LCCS: the fewer the classifiers in the reference class the higher the similarity. It is important that the current algorithm be changed to one that also takes into account the number of classifiers used.

5. Discussion

The semantic data harmonisation of five Nordic data sets and two other national data sets using LCCS-TM reveals that the problems in finding the correspondence between the original criteria and the LCCS classifiers are mainly related to: (1) differences in the definition of threshold values; (2) use of land-use terminology that cannot be translated in LCCS classifiers; and (3) use of mixed classes when the original classes contain a range of values that do not correspond with the options available in LCCS. Some information, especially when not closely linked to land cover, was inevitably lost in the data translation process. The use of user-defined attributes resolves such losses only partially as these attributes maintain the information in the class descriptions but are ignored when making the class comparisons. In the examples given, the user-defined attribute was not used. Whilst it is unrealistic to assume that no information losses will occur it is important that such losses are within acceptable limits. Some of the solutions adopted in the data translation of classes may require discussion and would, more importantly, require consensus between translators of different countries.

Comparison of the parameters used in the class sets shows that major class distinctions are made with more or less the same set of parameters although classes occur in different environments and class sets have been made for different purposes. Specific-purpose class sets use almost the full range of classifiers to describe the Cultivated Area classes, whereas for (Semi-) Natural Vegetation classes they show more variation in the use of classifiers. The use of the classifier Spatial Distribution is limited and Stratification is mainly not used, possibly because the class sets considered here are used in remote sensing applications in which the layering cannot, or to a very limited extent, be detected. However, the actual reasons for these patterns in classifier use are not evident.

In general, the similarity of land cover is highest in a group of classes describing the same main land-cover category. Values between classes belonging to different land-cover categories are in general small with the exception of Cultivated Area(s) and (Semi-) Natural Vegetation because these land-cover categories contain a number of identical classifiers to describe plants. Other land-cover categories, such as Bare Areas, Water Bodies and Snow and Built-up Areas can be easily distinguished because they are dissimilar to any other land cover.

The intracomparison of a class set indicates which classes are similar, and inversely dissimilar, and what types of parameters are used for smaller class distinctions. It will also reveal if a class set is geared towards a specific purpose: similarities found in Cultivated Area(s) will be higher than those in (Semi-) Natural Vegetation if the class set is geared towards agriculture. This type of analysis may contribute to a better design of the data validation effort because it explains what type of parameters is used for class boundary differentiation. It also shows whether or not these characterising elements can be linked to the interpretation of classes in the field and/or to any type of imagery. This could be a helpful tool because it shows which class distinctions should be supported by a significant validation effort.

It is important to note that once data translation into LCCS of a class belonging to a class set has been made, the user does not know if this data translation completely embraces the original class or that only part of the original class could be translated. The impression given in the LCCS-TM is that the data translation result is 100 percent, whereas frequently it will be (much) less. This is important to know if the data translation results are to be linked to land-cover change dynamics and the boundary conditions verified in the data validation effort are involved.

6. Conclusion

Though the selected land-cover class sets show that data translation and semantic data harmonisation can be carried out, much more study and discussion is needed to assess whether or not LCCS includes the classifiers commonly used, as well as the establishment of agreed rules for the computation of LCCS similarity in class comparisons. In certain cases very high values of similarity are found because few classifiers are shared with identical options selected, whereas the other classifiers used are neglected in the computation. The current algorithm results in values that do not reflect the number of classifiers used (e.g. if two classifiers are identical the fact whether there are three or five other classifiers is neglected). Therefore the resulting similarity values are biased with the tendency to result in very high values when few classifiers are compared. The current algorithm should be modified to include the detail with which a class has been defined. At present, there is no overall score of the similarity between two class sets. Such a value would be particularly important when translating existing class sets into

LCCS terminology when such a value could indicate if the data translation is close to the original class set. A quality assessment of the data translation per class as well as per class set is suggested analogous to the semantic data accuracy assessment (e.g., classification error matrix). More class sets need to be re-evaluated, translated, compared and intracompared to evaluate the efficiency of LCCS-TM. Feedback from the user community will be indispensable in order to assess and enhance the current methodology.

The distinct aspects of spatial data harmonisation that this paper did not focus on, i.e. geometric aspect, distribution and spatial aspects and temporal aspect, should be considered in parallel to the semantic aspect. If the two different data sets to be harmonised are seen as two different "objects", the harmonisation is the establishment of relationships between the two objects. The relationship between any two objects encompasses the assumptions that each makes about the other, including what operations can be performed and what behaviour results (Booch, 1994).

Acknowledgements

The authors would like to gratefully acknowledge Kiira Aaviksoo of the Estonian Environmental Information Centre (EEIC), Wendy Fjellstad and Arnt Kristan Gjertsen of the Norwegian Institute of Land Inventory (NIJOS) and Eva Ahlcrona of Swedish Metria Miljöanalys for making the correspondences of the land-cover class sets with LCCS-TM, their contributions and Michael Ledwith's (Metria Miljöanalys, Sweden) to the workshops and discussion forum of the Nordic Landscape Monitoring (NordLaM) Project of the Nordic Council of Ministers of Environment. Furthermore, Saverio Stoppioni of the Italian Istituto Agronomico per l'Oltremare (IAO) is gratefully thanked for the discussions on semantic data harmonisation with LCCS versions 1 and 2.

References

Booch, G., 1994. Object-oriented analysis and design with applications. Second edition. Benjamin/Cummings Publishing Company Inc., Santa Clara, California.

CSE [Centre de suivi écologique], 2005. Une étude diachronique de l'agriculture urbaine et péri-urbaine dans la région de Dakar. CSE, Dakar. Pp. 35. (author: D.S. Ndiaye)

Di Gregorio, A., Jansen, L.J.M., 1998. A new concept for a land-cover classification system. *The Land* 2.1:55-65.

Di Gregorio, A., Jansen, L.J.M., 2000. Land Cover Classification System (LCCS): classification concepts and user manual. FAO/UNEP, Rome.

Dolman, A.J., Verhagen, A., Rovers, C.A. (Eds.), 2003. Global environmental change and land use. Kluwer Academic Publishers, Dordrecht.

FAO, 1991. Land-cover mapping of Lebanon. NECP/LEB/001/SAU Project. FAO Remote Sensing Centre, Rome.

FAO, 1999. Provincial land cover atlas of the Islamic State of Afghanistan. AFG/90/002 Utilization of remote sensing for the inventory and monitoring of agricultural land in Afghanistan. FAO, UNDP and Afghan Geodesy and Cartography Head Office (AGCHO). FAO, Rome.

Jansen, L.J.M., 2004. Thematic harmonisation and analyses of Nordic data sets into Land Cover Classification System (LCCS) terminology. In: Groom, G. (Ed.), 2004. Developments in strategic landscape monitoring for the Nordic countries. Nordic Council of Ministers of Environment. ANP 2004/705:91-118.

Jansen, L.J.M., Bagnoli, M., Chimenti, S., Dell'Agnello, A., Focacci, M., Monaco, S.A., Sorini, P., Stoppioni, S., Urbano, F., 2006a. Satellite image interpretation of land-cover types in Manica and Maputo Provinces at nominal scale of 250,000 and at National level at nominal scale of 1:1,000,000. Technical report 1. Integrated Assessment of Mozambican Forests (AIFM) project. Agriconsulting S.p.A., Rome. 175p.

Jansen, L.J.M., Carrai, G., Morandini, L., Cerutti, P.O., Spisni, A., 2003a. Analysis of the spatio-temporal and semantic aspects of land-cover/use dynamics. World Bank Albanian National Forest Inventory (ANFI) project. Technical report 2. AGROTEC S.p.A. Consortium, Rome. 40p.

Jansen, L.J.M., Carrai, G., Morandini, L., Cerutti, P.O., Spisni, A., 2006b. Analysis of the spatio-temporal and semantic aspects of land-cover/use change dynamics 1991-2001 in Albania at national and district levels. *Environmental Monitoring & Assessment*, in print.

Jansen, L.J.M., Cerutti, P.O., Spisni, A., Carrai, G., Morandini, L., 2003b. Interpretation of remote sensing imagery for land-cover/use mapping. World Bank Albanian National Forest Inventory (ANFI) project. Technical report 1. AGROTEC S.p.A. Consortium, Rome. 62p.

Jansen, L.J.M., Mahamadou, H., Sarfatti, P., 2003c. Land-cover change analyses using LCCS. *Journal of Agriculture and Environment for International Development*, 97(1-2): 47-68.

Jansen, L.J.M., Ndiaye, D.S., 2006. Land-cover change dynamics 1978-1999 of (peri-) urban agriculture in the Dakar region. Submitted to *Journal of Agriculture and Environment for International Development.*

Lambin, E.F., Rounsevell, M.D.A. and Geist, H.J., 2000.Are agricultural land-use models able to predict changes in land-use intensity? *Agriculture, Ecosystems and Environment*, 82: 321-331.

Mahamadou, H., 2001. Étude diachronique des changements de couverture des terres au Niger (scène de Mayahi) des années 1990 à 1999. Gestion des ressources naturelles et sécurité alimentaire au Niger – interprétation des images satellitales pour la cartographie de couverture des terres. Istituto Agronomico per l'Oltremare (IAO)/Centre régional Agrhymet/Cooperazione italiana, Florence.

Mayaux, P., Bartholome, E., Fritz, S. and Belward, A., 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31: 861-877.

McConnell, W.J., Moran, E.F. (Eds.), 2001. Meeting in the middle: the challenge of meso-level integration. An international workshop on the harmonization of land-use and land-cover classification. LUCC Report Series No. 5. Anthropological Center for Training and Research on Global Environmental Change - Indiana University and LUCC International Project Office, Louvain-la-Neuve.

Travaglia, C., Milenova, L., Nedkov, R., Vassilev, V., Milenov, P., Pironkova, Z., 2001. Preparation of land-cover database of Bulgaria through remote sensing and GIS. Environment and Natural resources Working paper No. 6. FAO, Rome.

UNEP/FAO, 1994. Report on the UNEP/FAO expert meeting on harmonizing land-cover and land-use classifications. GEMS report series No.25. Nairobi.

Wyatt, B.K., Billington, C., De Bie, K., De Leeuw, J., Greatorex-Davies, N., Luxmoore, R., 1997. Guidelines for land-use and land-cover description and classification. ITE/WCMC/ITC/UNEP/FAO Internal unpublished working document. ITE, Monkshood, UK.

Wyatt, B.K., Gerard, F.F., 2001. What's in a name? Approaches to the inter-comparison of land-use and land-cover classifications. In: Groom, G., Reed, T. (Eds.), 2001. Strategic landscape monitoring for the Nordic countries, TemaNord 2001:523. Nordic Council of Ministers, Copenhagen.

Figure 1. Overview of the process of definition of corresponding classes in LCCS and their comparison

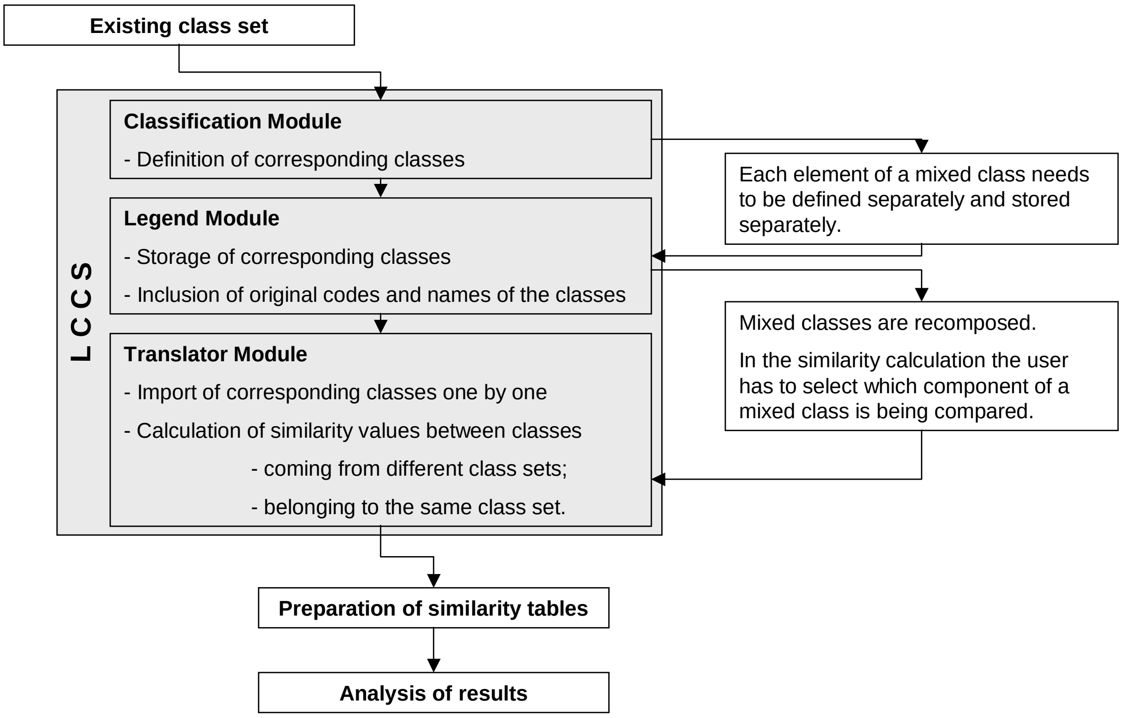

Table 1. Differences in crown cover thresholds between LCCS and SMD

Land-cover type	Crown cover (in %)	
	LCCS	SMD
Closed forest	>65	
Open forest/Woodland	40-65	
	15-40	>30
Sparse trees	1-15	

Table 2. Differences in densities of urban areas

Density classes	Density (in %)	
	LCCS	SMD
High density/continuous urban fabric	>75	>80
Medium density/discontinuous urban fabric	50-75	50-80
Low density/discontinuous urban fabric	<50	30-50

Table 3. Differences in tree height

Distinguished tree height classes	Height (in m)	
	LCCS	SMD
High/Coniferous forest >15 m	>14-30	>15
Medium/Forest	7-14	>5
Low/Young forest	3-7	2-5

Table 4. Overview of the use of classifiers and two specific attributes in the major land-cover category Cultivated Terrestrial Areas (A11) by the different class sets[a]

		Land-cover classifiers						Attributes	
Class set	Class codes	A. Life form	B. Field size	B. Distribution	C. Crop combination	D. Water supply	D. Cultivation time factor	S. Crop cover	W. Crop type
AIS-LCM[b]	18	x	-	x	-	x	x	-	-
(Denmark)	07	x	-	x	-	x	x	-	-
3Q	A1pl	x	-	x	x	x	x	x	x
(Norway)	A2fr	x	-	x	x	x	x	-	x
	A2bu	x	-	x	x	x	x	-	-
	A1ko	x	-	x	-	x	x	x	-
	A1br	x	-	x	x	x	x	-	-
	A3be	x	-	x	x	x	x	x	-
	A1en	x	-	x	x	x	x	x	-
	A1be	x	-	x	x	x	x	x	-
	A1in	x	-	x	x	x	x	x	-
	A1na	x	-	x	x	x	x	-	x
	A2jo	x	-	x	x	x	x	-	-
	A3sb	x	-	x	x	x	x	x	-
	A3kr	x	-	x	x	x	x	x	-
	A1gr	x	-	x	x	x	x	x	-
DMK[c]	1	x	-	-	-	-	-	-	-
(Norway)	2	x	-	-	-	-	-	-	-
	3	x	-	-	-	-	-	-	-
	4	x	-	-	-	-	-	-	-
SMD	211	x	-	-	-	-	-	-	-
(Sweden)	231	x	-	-	-	-	-	-	-
	222	x	-	-	-	-	-	-	-

[a] The EELC class set is not represented as it contains no relevant classes.
[b] Class codes AIS-LCM class set: 18-Seasonally vegetation covered ground, 07-Grazed or mown grass.
[c] Class codes DMK class set: 1. Agricultural area, 2. Arable land, 3. Surface cultivated area, 4. Fertilized pasture.

Table 5. Overview of the use of classifiers and two specific attributes in the major land-cover categories (Semi-) Natural Vegetation (A12 and A24) by the different class sets

		Land-cover classifiers									Attribute
Class set	Class codes	A. Life form	A. Cover	B. Height	C. Spatial distribution (A12)	C. Water seasonality (A24)	D. Leaf type	E. Leaf phenology	F./G. Stratification 2nd layer	F./G. Stratification 3rd layer (A12)	T. Floristic aspect
AIS-LCM[a]	14	x	x	x	x	≠	-	-	-	-	-
(Denmark)	11	x	x	x	x	≠	-	-	-	-	-
	05	x	x	x	-	≠	-	-	-	-	-
	10	x	x	x	-	≠	-	-	-	-	-
	15	x	x	x	-	≠	x	x	-	-	-
	16	x	x	x	-	≠	x	x	-	-	-
	08	x	x	x	≠	x	-	-	-	≠	-
EELC[b]	6	x	x	x	≠	x	-	-	-	≠	x
(Estonia)	19	x	x	x	≠	x	-	-	-	≠	x
	20	x	x	x	≠	x	-	-	-	≠	x
	21	x	x	x	≠	x	-	-	-	≠	x
	10	x	x	x	≠	x	-	-	-	≠	x
	13	x	x	x	≠	x	-	-	-	≠	x
	18	x	x	x	≠	x	-	-	-	≠	x
	17	x	x	x	≠	x	-	-	-	≠	x
	12	x	x	x	≠	x	-	-	-	≠	x
	9	x	x	x	≠	x	-	-	-	≠	x
	8	x	x	x	≠	x	-	-	-	≠	x
	11	x	x	x	≠	x	-	-	-	≠	x
	14	x	x	x	≠	x	-	-	-	≠	x
	16	x	x	x	≠	x	-	-	-	≠	x
	15	x	x	-	≠	-	-	-	-	≠	x
3Q	S1la	x	x	x	x	≠	x	x	-	-	-
(Norway)	S2bl	x	x	x	x	≠	x	x	-	-	-
	S3ba	x	x	x	x	≠	x	x	-	-	-
	S2bl	x	x	x	x	≠	x	x	-	-	-
	F1kr	x	x	x	-	≠	-	-	x	-	-
	F1vi	x	x	x	-	≠	-	-	-	-	-
	F1sb	x	x	x	x	≠	-	-	x	-	-
	F1sb	x	x	x	x	≠	-	-	x	x	-
	F2ri	x	x	x	≠	-	-	-	-	≠	-
	F2sk	x	x	x	≠	-	-	-	-	≠	-
	M2bu	x	x	x	≠	x	x	x	-	≠	-
	M1gr	x	x	x	≠	-	-	-	-	≠	-
	F3st/F3dr	x	x	x	≠	x	-	-	-	≠	-
DMK[c]	5	x	x	x	-	≠	x	x	-	≠	-
(Norway)	6	x	x	x	-	≠	x	x	-	≠	-
	7	x	x	x	-	≠	x	x	-	≠	-
	8	x	x	-	≠	-	-	-	-	≠	-
SMD[d]	3242	x	x	-	-	≠	-	-	-	-	-
(Sweden)	224	x	x	-	-	≠	-	-	-	-	-
	3243	x	x	x	-	≠	-	-	-	-	-
	3241	x	x	-	-	≠	-	-	-	-	-

		Land-cover classifiers									Attribute
Class set	Class codes	A. Life form	A. Cover	B. Height	C. Spatial distribution (A12)	C. Water seasonality (A24)	D. Leaf type	E. Leaf phenology	F./G. Stratification 2^{nd} layer	F./G. Stratification 3^{rd} layer (A12)	T. Floristic aspect
	322	x	x	x	-	≠	-	-	-	-	-
	3212	x	x	-	-	≠	-	-	-	-	-
	3211	x	x	-	-	≠	-	-	-	-	-
	3111	x	x	x	-	≠	x	x	-	-	-
	3112	x	x	x	-	≠	x	x	-	-	-
	3113	x	x	x	-	≠	x	x	-	-	-
	31211	x	x	x	-	≠	x	x	-	-	-
	312122	x	x	x	-	≠	x	x	-	-	-
	312121	x	x	x	-	≠	x	x	-	-	-
	3122	x	x	x	-	≠	x	x	-	-	-
	3123	x	x	x	-	≠	x	x	-	-	-
	3131	x	x	x	-	≠	x	x	-	-	-
	5122	x	x	-	≠	-	-	-	-	≠	-
	411	x	x	x	≠	x	-	-	-	≠	-
	421	x	x	x	≠	-	-	-	-	≠	-
	5232	x	x	-	≠	-	-	-	-	≠	-
	4121	x	x	x	≠	x	-	-	-	≠	-
	4122	x	x	x	≠	x	-	-	-	≠	-

[a] Class codes AIS-LCM class set: 14-Scrub or woodland, 4. 11-Shrub dominated heathland, 5. 05-Grass dominated heathland, 6. 10-Shrub and grass heathland, 7. 15-Deciduous forest, 8. 16-Evergreen forest, 9. 08-Marshland.

[b] Class codes EELC class set: 6. Coastal and shore reed bed, 19. Minerotrophic swamp forest, 20. Transitional bog forest, 21. Bog forest, 10. Treed fen, 13. Treed transitional bog, 18. Treed bog, 17. Dwarf shrub bog, 12. Shrub transitional bog, 9. Shrub fen, 8. Open fen, 11. Open transitional bog, 14. Bog lagg, 16. Complex bog, 15. Open bog.

[c] Class codes DMK class set: 5. Broadleaved forest, 6. Needle leaved forest, 7. Mixed forest, 8. Mire.

[d] Class codes SMD class set: 3242 Cleared-felled areas (before woody vegetation), 334 Burned areas (before woody vegetation), 3243 Younger forest, 3241 Thickets, 322 Moors and heathland, 3212 Meadow, 3211 Grass heath, 3111 Broadleaved forest not on mires or bare rock, 3112 Broadleaved forest on mires, 3113 Broadleaved forest on bare rock, 31211 Coniferous forest on lichen dominated areas, 312122 Coniferous forest >15m, 312121 Coniferous forest 5-15m, 3122 Coniferous forest on mires, 3123 Coniferous forest on bare rock, 3131 Mixed forest not on mires or bare rock, 5122 Water bodies, vegetation covered water area, 411 Inland marches, 421 Salt marshes, 5232 Sea and ocean, vegetation covered water area, 4121 Wet mires, 4122 Other mires.

Table 6. Intracomparison AIS-LCM class set (Denmark)

LCCS order Class codes	1 18	2 07	3 14	4 11	5 05	6 10	7 15	8 16	9 08	10 03	11 01
1	100	100	0	0	25	0	0	0	25	0	0
2	100	100	0	0	25	0	0	0	25	0	0
3	0	0	100	40	0	0	60	60	0	0	0
4	0	0	40	100	0	60	20	20	0	0	0
5	25	25	0	0	100	25	0	0	75	0	0
6.1	0	0	0	75	25	100	0	0	25	0	0
6.2	25	25	0	0	100	25	0	0	75	0	0
7.1	0	0	60	20	0	0	100	60	0	0	0
7.2	0	0	40	0	20	20	80	40	20	0	0
8	0	0	60	20	0	0	60	100	0	0	0
9	25	25	0	0	75	25	0	0	100	0	0
10.1	0	0	0	0	0	0	0	0	0	100	0
10.2	0	0	0	0	0	0	0	0	0	0	0
11.1	0	0	0	0	0	0	0	0	0	0	100
11.2	0	0	0	0	0	0	0	0	0	0	100

Class codes: 1. 18-Seasonally vegetation covered ground, 2. 07-Grazed or mown grass, 3. 14-Scrub or woodland, 4. 11-Shrub dominated heathland, 5. 05-Grass dominated heathland, 6. 10-Shrub and grass heathland, 7. 15-Deciduous forest, 8. 16-Evergreen forest, 9. 08-Marshland, 10. 03-Unvegetated ground, 11. 01-Open water.

Explanation of grey shading used:

100% similarity
75-99% similarity
50-74% similarity
Less than 50% similarity

Table 7 Intracomparison matrix Lebanon land-cover class set (in percentages) (reference classes presented vertically)

LCCS categories			A11						A12 and A24						B15		B16		B28
	LCCS order [a]		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		Class codes [b]	4c	4d	4b	2	4a	3	6b	6a	6c	5a	5b	8	1a	7a	7b	7c	9
A11	1	4c	100	60	80	20	80	40	60	20	0	0	0	0	0	0	0	0	0
	2	4d	100	100	67	33	100	67	33	33	0	0	0	0	0	0	0	0	0
	3	4b	67	33	100	17	67	50	33	0	17	0	0	0	0	0	0	0	0
	4	2	11	11	11	100	11	11	0	0	0	6	6	6	0	0	0	0	0
	5	4a	67	50	67	17	100	50	33	33	0	0	0	0	0	0	0	0	0
	6	3	33	33	50	17	50	100	0	0	0	8	8	8	0	0	0	0	0
A12 and A24	7	6b	50	17	33	0	33	0	100	67	33	0	0	17	0	0	0	0	0
	8	6a	17	17	0	0	33	0	67	100	33	0	0	17	0	0	0	0	0
	9	6c	0	0	25	0	0	0	50	50	100	0	0	25	0	0	0	0	0
	10	5a	0	0	0	17	0	17	0	0	0	100	33	33	0	0	0	0	0
	11	5b	0	0	0	25	0	25	0	0	0	50	100	50	0	0	0	0	0
	12	8	0	0	0	10	0	10	20	20	20	20	20	100	0	0	0	0	0
B15	13	1a	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
B16	14	7a	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
	15	7b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
	16	7c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
B29	17	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

[a] The LCCS categories are: A11=Cultivated and Managed Terrestrial Areas, A12=(Semi-) Natural Terrestrial Vegetation, A24=(Semi-) Natural Aquatic or Regularly Flooded Vegetation, B15=Artificial Surfaces and Associated Areas, B16= Bare Areas, B28=Natural Water Bodies, Snow & Ice.

[b] The LCCS order refers to the order of classes in the translated class sets according to the automatic grouping of the LCCS Legend Module.

[c] Class code refers to the original class set coding system (FAO, 1991): 1. Urban areas, 2. Horticulture, 3. Field crops and fallow land, 4a. Olives, 4b. Vineyards, 4c. Deciduous fruit trees, 4d. Citrus or bananas, 5a. Grassland and forbs from open to closed, or abandoned fields, or old fallows in agricultural areas, 5b. Sparse grassland and forbs in mountains or desertic areas, 6a. Coniferous forest, 6b. Deciduous forest, 6c. Scrubland and other types of degenerate forest, 7a. Barren rocks, 7b. Highly dissected and eroded land, 7c. Beaches, 8. Swamp vegetation, 9. Water bodies.

Explanation of grey shading used:

100% similarity
75-99% similarity
50-74% similarity
Less than 50% similarity

Table 8. Similarity between two Cultivated Area classes of the 3Q class set (reference class) with two classes of the SMD class set

LCCS classifiers and attributes		3Q (Norway) *4. A1ko*	SMD (Sweden) *3211. Grass heath*	Similarity
Classifiers	A. Life form	Graminoid crops	Graminoids	1
	A. Cover	-	Closed	-
	B. Spatial distribution	Continuous		0
	D. Water supply:	Rainfed cultivation		0
	D. Cultivation time factor:	Permanent cropped		0
				25 %
		4. A1ko	*3212. Meadow*	
Classifiers	A. Life form:	Graminoid crops	A2 Herbaceous	0.5
	A. Cover:	-	A10 Closed	-
	B. Spatial distribution:	Continuous		0
	D. Water supply:	Rainfed cultivation		0
	D. Cultivation time factor:	Permanent cropped		0
				13%
		12. A3sb	*3211. Grass heath*	
Classifiers	A. Life form:	Graminoid crops	Graminoids	1
	A. Cover:	-	Closed	-
	B. Spatial distribution:	Continuous		0
	C. Crop combination:	Single crop		0
	D. Water supply:	Rainfed cultivation		0
	D. Cultivation time factor:	Permanent cropped		0
				20%
		12. A3sb	*3212. Meadow*	
Classifiers	A. Life form:	Graminoid crops	A2 Herbaceous	0.5
	A. Cover:	-	A10 Closed	-
	B. Spatial distribution:	Continuous		0
	C. Crop combination:	Single crop		0
	D. Water supply:	Rainfed cultivation		0
	D. Cultivation time factor:	Permanent cropped		0
				10%

Table 9. Similarity of 3Q class (A12 and A24) with (Semi-) Natural Terrestrial Vegetation (A12) classes of SMD

LCCS classifiers and attributes		3Q (Norway)	SMD (Sweden)	Similarity
		30. F2ri	*3241. Thickets*	
Classifiers	A. Life form:	Shrubs	Shrubs	1
	A. Cover:	Closed	Closed	1
	B. Height:	5-0.3m		0
Modifier	B. Height:	<0.5m		0
				50%
		32. M2bu	*322. Moors and Heathland*	
Classifiers	A. Life form:	Shrubs	Shrubs	1
	A. Cover:	Closed	Closed	1
	B. Height:	5-0.3m	5-0.3m	1
Modifier	B. Height:	3-0.5m	<0.5m	0
	C. Water seasonality:	Waterlogged soil		0
Classifiers	D. Leaf type:	Broadleaved		0
	E. Leaf phenology:	Deciduous		0
	F. Stratification:	single layer		0
				38%

Table 10. Similarity between of Cultivated Areas classes of the Lebanon and Afghanistan class sets in which also attributes occur

LCCS classifiers and attributes		4c. Deciduous Fruit Trees (Lebanon)	2b. Vineyards (Afghanistan)	Similarity [c]
Classifiers	A. Life form:	tree crop(s)	shrub crop(s)	0
		broadleaved	broadleaved	1
		deciduous	deciduous	1
	B. Field size	large-to-medium sized field(s)	large-to-medium-sized field(s)	1
Modifier	B. Field size	- [a]	medium-sized field(s)	#
Classifier	B. Spatial distribution	continuous	continuous	1
Classifier	C. Crop combination	-	single crop monoculture)	#
Attributes [b]	Crop type	fruit and nuts	fruit and nuts – grapes	≠
	Crop cover	open, orchards	closed, plantations	≠
				80 %

[a] Option not existent or not used in the reference class, and therefore not included in the computation (#).
[b] These so-called Environmental or Specific Technical Attributes, are not considered in the computation (≠).
[c] Five classifiers are considered in the computation of which 4 are similar hence a similarity of 80%.

Table 11. Similarity between classes of two different land-cover categories of the Lebanon and Afghanistan class sets

	5b. Sparse grassland and forbs in mountains or desertic areas (Lebanon)		4a. Cultivation in flat areas (Afghanistan)		Similarity [c]
Classifiers	A. Life form	herbaceous	A. Life form	herbaceous crop(s)	1
	A. Cover	sparse (20-10)-1%	-		0
		- [a]	B. Field size:	large-to-medium	#
			B. Spatial Distribution:	continuous	#
		-	C. Crop combination:	single crop	#
		-	D. Cover-related Cultural Practices:	rainfed	#
Attributes [b]	Landform:		Landform:		
	Major land class	steep land – high gradient hill	Major Land class	level land	≠
	Slope class	rolling	Slope class	flat to almost flat	≠
	Climate	subtropics winter rainfall – arid	-		≠
	-		Crop Type:	dominant crop – cereals - wheat	≠
					50 %

[a] Option not existent or not used in the reference class, and therefore not included in the computation (#).
[b] These so-called Environmental or Specific Technical Attributes, are not considered in the computation (≠).
[c] Only two classifiers are considered in the computation of which one is similar, hence a similarity of 50%.