POTENTIAL OF AUTOMATIC CLASSIFICATION IN ALPINE AREAS FOR SUPPORTING THE UPDATE OF CORINE LANDCOVER

Klaus Steinnocher
Austrian Research Centers, Seibersdorf, Austria
klaus.steinnocher@arcs.ac.at
Ursula Schmitt
Joanneum Research, Graz, Austria
ursula.schmitt@joanneum.ac.at
Peter Aubrecht
Federal Environment Agency, Vienna, Austria
aubrecht@ubavie.gv.at

KEY WORDS: generalization, land use, land cover, monitoring, semi-automation, spatial databases

ABSTRACT

The paper discusses benefits and limitations of automated classification results as support for updating land cover maps. The classification results of an Alpine Monitoring project (ALPMON) are compared to the respective land cover maps from the European CORINE land cover programme. While the CORINE land cover maps are based on visual interpretation of analogue satellite imagery, the ALPMON approach relies to a high extent on automated processing of digital data. Two major differences have to be considered when comparing these data sets. First the CORINE land cover units have a minimum size of 25 ha, while in ALPMON the pixel is the minimum mapping unit. Second the nomenclature of CORINE includes general land cover classes, while ALPMON focuses on typical alpine land cover types. The first problem was solved by applying a spatial generalization method, that aggregates small patches of land cover types to larger land cover units. In order to overcome the second difference transition rules between the nomenclatures were defined.
Results of the comparison lead to the impression that the different approaches cannot be directly combined, at least not on a detailed thematic level. Comparing the land cover maps on an aggregated level might help the interpreter to concentrate on areas that are likely to have changed, or have been incorrectly assigned in the first version of the CORINE land cover map. The final decision on the land cover type, however, should be left to the interpreter.

1 INTRODUCTION

The project “Inventory of alpine relevant parameters for an alpine monitoring system using remote sensing data” (ALPMON) aimed at the compilation of a basic landscape register for an alpine monitoring system by means of the analysis of TM, SPOT and other high resolution satellite images. The information system shall serve as the basis for planning tasks. Five Alpine landscapes were selected as test sites due to their typical characteristics. The components of the alpine monitoring system were, firstly, derived from the results of a classification of satellite images and, secondly, extrapolated from thematic maps. The information levels were harmonized in the different test sites. Primary information on the nature and state of vegetation for ALPMON were compiled solely by means of remote sensing. Remote sensing data plays a key role in the construction of information systems which use a small scale. An important objective, therefore, was the testing of operational and semi operational processing methods which permit as precise a compilation of specified parameters as possible. Additional parameters derived from existing maps were integrated as auxiliary information in the processing of satellite images and provided the basis for specific GIS applications. In order to demonstrate the feasibility of the Alpine monitoring system specific applications were performed in close co-operation with national customers (ALPMON, 1997, Waser et al., 2000).

The presented paper concentrates on one out of six applications, i.e. the integration of the classification results into the CORINE land cover model. It is restricted to the results of one test site, the Dachstein test site, located in the center of Austria.

2 MOTIVATION

The aim of the CORINE („Co-ordination of Information on the Environment“) land cover project was to map the land use / land cover of Europe within the framework of the CORINE Program of the European Union. The European Environment Agency (EEA) in Copenhagen was in charge of the CORINE Program; co-ordination and support was performed by the „European Topic Centers on Land Cover“ (ETC/LC). The Project was carried out by so called „National Reference Centres“ (NRC) in the member states. The „Co-ordination of Information on the Environment“ for Europe offers the possibility of comparing environmental data between the Member States. Furthermore the data can be a useful tool for environmental applications and research at European level.

The methodology – developed by the „Joint Research Centre“ (JRC) in Ispra, Italy (EUR 1993) – is based on visual interpretation of satellite data including the use of ancillary data (topographical and thematic maps, statistics etc.). The photo-interpretation was performed on the basis of color hardcopies of geocoded multispectral satellite data in a scale of 1 : 100.000. The different cover types were mapped according to the CORINE land cover Nomenclature, which consists of 44 different land cover / land use types. The surface area of the smallest unit mapped is 25 hectares. For linear elements the minimum width is 100 meters. Finally, the interpretation results were scanned and integrated into a geographical information system.

The interest of the NRCs related to the ALPMON project concentrates on the possibilities to reduce the effort of visual interpretation and thus the costs of up-dating the CORINE land cover maps. This could be achieved by applying an automated approach as performed in the ALPMON project as alternative to photointerpretation. However, it has to be noticed that formalized methods of pattern recognition cannot currently replace the ability of the human vision system. On the other hand they provide reproducible results that are more consistent than the results of visual interpretation. In addition it might be possible to extract more detailed information to be added as level 4 classes to the CORINE nomenclature. The particular interest of the Austrian Environment Agency, who acts as the end user in this study, lies in the high alpine regions which are considered the most crucial areas for visual interpretation. This includes above all the 'Forests and semi-natural areas' classes of the CORINE nomenclature. A clear distinction of bare rock, sparsely vegetated areas, natural grassland and different forested areas is of major interest for users of the CORINE land cover maps. In addition separation of dwarf-pines from other coniferous forests would significantly improve the available information.

3 METHODOLOGY

3.1 Classification

Compared to the CORINE land cover maps the ALPMON data base has been derived from semi-automatic classification of digital satellite imagery. According to the processing chain developed in the course of the ALPMON project the following procedures have been applied:

- Geocoding
- Topographic normalization
- Atmospheric correction
- Image fusion (optional)
- Numerical classification

The resulting thematic maps refer to the nomenclature defined in the ALPMON project (Table 1). Minimum mapping unit is the single pixel. Direct comparison between the CORINE and the ALPMON land cover maps (Figure 3) shows three fundamental differences:

- Differences in the nomenclatures
- Different minimum mapping unit
- Different data representation (raster versus vector)

While the problem of data representation will not be discussed in this paper achievements will be presented that have been made in setting up rules for the transition between the nomenclatures and in overcoming the problem of the different minimum mapping units.

Table 1: ALPMON aggregated land cover classes

Forest type	Forest age	Forest canopy closure
Broad-leaved (>25% coniferous) Mixed (25-75% coniferous) Coniferous (>75% coniferous)	Culture No age defined (for pinus mugo, greenalder) Thinning to pole Timber and old timber	Crown closure >10 to 30% Crown closure >30 to 60% Crown closure >60%
Non-forest classes	Sealed surface	
Rock, gravel, soil, sealed surface Water Meadow, pasture Shrubs (rhod.sp., junip.sp., etc.) and "open forest" (crown closure up to 10%) Wet land Snow, ice Shadow	Sealed surface >75% Sealed surface 20-75%	

3.2 Transition of nomenclatures

Discussions with the Austrian Environment Agency were used to define the ALPMON classes in the context of the CORINE nomenclature. A table was set up that contains the corresponding CORINE land cover class for all ALPMON classes (Table 2).

As can be seen from Table 1 the ALPMON approach concentrated on the classification of different types, crown closure and age of forest. In addition a number of non forest classes were derived. No focus was set on the differentiation of artificial surfaces. For the integration into CORINE the forest types were aggregated to broad-leaved, coniferous and mixed forest. The forest age was not considered, but the crown closure was taken into account in order to investigate the transition between non-forest and forest areas in the CORINE land cover. Grassland comprises both natural grassland and pastures. Shrubland was assigned to transitional woodland shrub. Rock, gravel and bare soil might comprise both rocks and sparsely vegetated areas. Wetlands and water bodies could be limited to the corresponding classes of inland marches and inland waters. The lack of most Artificial Surface and Agricultural classes can be explained by the alpine environment of the test sites – the altitude of most areas is too high for growing agricultural crops and large urban areas do not occur.

Table 2: Transition rules between ALPMON and CORINE land cover classes

ALPMON classes	CORINE classes
Sealed surfaces	1.1 Urban fabric
Broad-leaved forest (aggregated)	3.1.1 Broad-leaved forest
Coniferous forest (aggregated)	3.1.2 Coniferous forest
Mixed forest (aggregated)	3.1.3 Mixed forest
Grassland	2.3.1 Pastures 3.2.1 Natural grassland
Shrubs (juniperus, rhododendron)	3.2.4 Transitional woodland shrub
rock/gravel/soil	3.3.2 Bare rock 3.3.3 Sparsely vegetated areas
Snow/ice	3.3.5 Glaciers and perpetual snow
Wet land	4.1.1 Inland marshes
Water	5.1 Inland waters

Applying the transition rules from Table 2 allowed for a direct comparison of the two data sets. In order to enable this comparison the vector based CORINE land cover maps were rasterized with the pixel size of the ALPMON land cover maps. Confusion matrices of the two raster layers give a first idea of the level of agreement between the data sets. They were also used as reference for generalizing the ALPMON classification results as described in the following chapter.

3.3 Spatial generalization of land cover maps

When comparing thematic maps on a pixel basis the minimum mapping units have to be considered. Areas that are composed of different land cover types in one map might be generalized to one land cover type in the other map. In the actual case the minimum mapping unit of CORINE equals 25 hectares, whereas the ALPMON results are classified on a pixel basis. Assuming a pixel size of 10m all regions smaller than 2500 pixels have to be generalized, e.g. an area composed of patches of coniferous and broad-leaved forest each being smaller than 2500 pixel might be assigned mixed forest after generalization.

In order to perform the generalization two aspects have to be considered. First only those pixels are processed that belong to regions smaller than 25 hectares. Second the neighborhood of these pixels has to be investigated in order to find the proper generalized class for the pixels. This neighborhood analysis will be performed applying a so-called postclassification algorithm (Steinnocher 1996).The algorithm works within a local neighborhood which is defined by a moving window. Within this window a standardized histogram of the input classes is calculated, representing the spatial composition i.e. the frequency of the input classes found in the local neighborhood. The histogram is then compared to a set of rules which represents the expected frequency of input classes for each generalized class. As soon as a rule is found to be true the corresponding generalized class is assigned to the center pixel of the window. Each rule defines the minimum frequency of one or more input classes for one generalized class. When compared to the corresponding elements in the histogram the frequency values represent thresholds. If all thresholds are exceeded within a rule, it is recognized as true and the corresponding secondary class will be assigned. The rules allow for a combination of several sub-rules within one major rule. Each sub-rule defines a threshold for one or more input classes and all sub-rules have to be true to accept the major rule. Processing of the rule-set is performed step by step starting at the top of the set. As soon as a rule is accepted and therefore applied, the rest of the rule-set will not be considered any more. If no rule is found to be true a rejection class is assigned.

Apart from the design of the rule-set the size of the analyzed neighborhood represents a crucial parameter in the postclassification process. Choosing a small window size will lead to a ‘noisy’ result since only high frequency structures will be recognized. If the window is too large the smoothing effect will become very strong, thus leading to a loss of detail. At this point it has to be noted that the presented postclassification is a generalization process and will always suppress details. However, for the actual application this effect is required for deriving the generalized representation of the CORINE land cover maps.

Application of the postclassification algorithm will lead to patches that might still be smaller than the minimum mapping unit (although it is likely that they will be larger than the patches in the original classification). By simply shrinking these areas a lot of information would be lost. Therefore a second generalization process is needed after the postclassification. This is achieved by applying an iterative shrinking process that affects only the regions originally smaller than the minimum mapping unit, while the larger ones are not changed at all. First all regions smaller than a certain threshold are shrunk leading to increasing larger regions at the expense of the smallest ones. In the following iterations this process is repeated applying an increased threshold until the threshold equals the desired minimal mapping unit. Performing this iterative process will lead to a generalized map without loosing significant information.

4 RESULTS

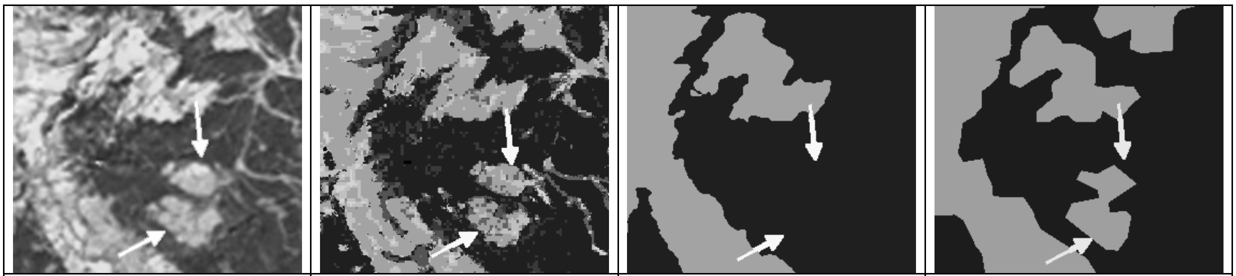
The process of postclassification and subsequent iterative shrinking was applied to the classification results of the test site. In order to show the effect of this process both the original classifications and the resulting generalized maps were compared to the respective CORINE land cover maps. Visual comparison was performed by means of maps (Figure 3), while a numerical comparison was based on confusion matrices (Table 3). Looking at the results of the comparison a general trend can be seen. While the agreement between the original classifications and the CORINE land cover maps is relatively low, it increases significantly after the application of the postclassification process. This effect is due to the increased level of generalization that is provided by the post-classified images. In addition there is a significant increase of agreement when reducing the degree of thematic detail in the CORINE land cover maps by aggregating to level 2 (15 classes) and level 1 (5 classes). This indicates that the major confusion between the maps result from detailed thematic differences, that are likely to be caused by mis-interpretation/-classification and differences in nomenclatures. In order to better understand the reason of the remaining confusions selected areas of the test site were compared to aerial photographs.

Table 3: Confusion between CORINE land cover map and ALPMON results

	CORINE Level 3	CORINE Level 2	CORINE Level 1
ALPMON classification	53%	62%	79%
ALPMON postclassification	64%	70%	88%

In general the following reasons can be given for the confusion between the two data sets:

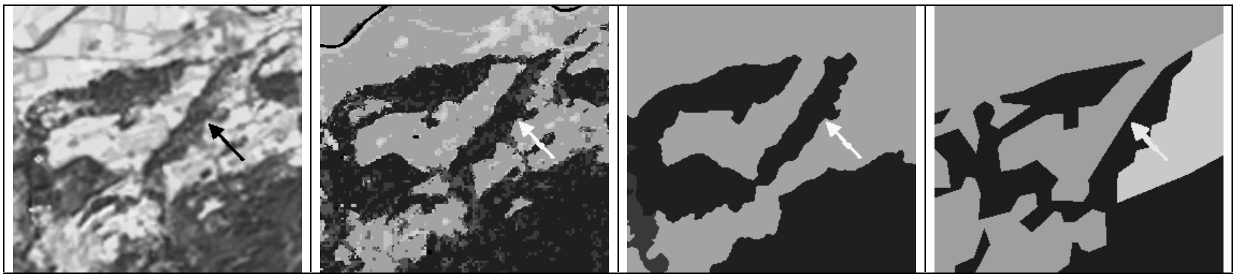
Different generalization algorithms: whereas CORINE is based on the delineation and interpretation of more or less homogeneous areas with a minimum size of 25ha, the ALPMON result is based on a pixel-wise classification which is generalized/aggregated in a second step. In the course of this generalization, due to their size and class mixture some small areas that are surrounded by dominant classes disappear, whereas they are delineated in the CORINE data set (Figure 1).



Landsat TM 5,4,3	ALPMON classification	ALPMON generalized	CORINE land cover

Figure 1: Confusion of meadow (light gray) and coniferous forest (dark gray).

Location error*s* or shift of objects in the CORINE land cover maps compared to the Landsat TM image cause disagreement in the thematic class of an area in both directions of the shift. The reason for this geometric problem is not clear, as it occurs only in few parts of the image (Figure 2).



Landsat TM 5,4,3	ALPMON classification	ALPMON generalized	CORINE land cover

Figure 2: Location error (the arrow indicates identical location), and interpretation of agricultural land (bright gray) in CORINE.

Different nomenclature and definition of classes in the two data sets sometimes makes a direct comparison of the classes difficult. E.g. in the ALPMON classification no difference between natural and agricultural grassland has been made as is the case for CORINE. The CORINE category *heathland* comprises dwarf mountain pine, which in ALPMON is part of *coniferous forest.* Also the definition of classes is a critical factor for the comparability. This is especially the case in the alpine region, where a transition between different land cover categories without clear borders and a mixture of land cover types within small areas are typical. E.g. the ALPMON definition of *rock/gravel/soil* is that more that 60% of an area are covered with this category. The rest (up to 40%) may be covered by grass or shrub. These areas may be interpreted as *natural grassland* in CORINE.

Interpretation errors in CORINE, partly due to subjectivity of interpretation, and **classification errors** in ALPMON occur, but play a minor role in the comparison of the two data sets.

Major confusion between the results of the ALPMON classification and the CORINE land cover interpretation occurs with the following classes:

CORINE	ALPMON (in descending order)
natural grassland	meadow, shrub, forest, rock
heathland	coniferous f., shrub, meadow, broad-leaved forest
coniferous forest	broad-leaved f., meadow, shrub, mixed forest
rock	shrub
meadow	coniferous forest

The confusion between ***natural grassland*** (CORINE) and *meadow* (ALPMON) is due to the nomenclature, as in ALPMON no differentiation between these classes exists. The confusion with *shrubs* (small bushes in ALPMON) mainly occurs in the transition zone between forest, shrubs, and alpine pastures. By aggregation the class mixture is replaced by the class *shrubs* which therefor is overemphasized in the generalized ALPMON classification. In the same way, confusion with *forest* (ALPMON) partly is caused by the aggregation. But in general, the CORINE forest border is set lower than the ALPMON forest border, which might be a problem of nomenclature, as in ALPMON forest is defined with a relatively low tree canopy closure (> 10%). Confusion with *rock* may be partly due to the definition of the ALPMON rock class (including up to 40% vegetation), but the subjectivity of the CORINE interpretation seems to be another reason.

Heathland in the CORINE nomenclature is defined as dwarf mountain pine. This category in ALPMON belongs to *coniferous forest*. Thus confusion between these two categories is as expected. The main reason for confusion with *shrubs* and *meadows*, which mainly occurs in the transition zone between forest (including low growing trees), shrubs, and meadows, is most probably the definition of these classes and the subjectivity of interpretation in CORINE. Confusion with *broad-leaved forest* in ALPMON is concentrated on areas covered by *greenalder*. It seems reasonable to assume that *greenalder* in the CORINE data set also has been assigned to *heathland*.

In general it has to be stressed that no *mixed forest* and very little *broad-leaved forest* has been identified in CORINE. These categories in most cases are assigned to ***coniferous forest***. For the confusion with *meadow* three reasons were found. By generalization of the ALPMON classification some forest areas are changed or disappear completely but are identified in CORINE. Main reason for the differences are location errors/shifts in CORINE (Figure 2). Finally, some interpretation errors could be detected in CORINE. Confusion with *rock/gravel/soil* mainly occurs in the transition zone between forest, other vegetation, and areas with sparse or no vegetation. This again is a matter of class definition and interpretation, but also of aggregation. In general, CORINE tends to overemphasize *forest* with respect to *shrubs*.

Confusion of ***rock/gravel/soil*** with *coniferous forest* is mainly restricted to areas with a mixture of forest and rock/gravel. The reason for the differences may be found in the different definition of forest (in ALPMON canopy cover > 10%). Confusion with *shrubs* is mainly concentrated on the transition areas between rock, shrubs and meadow. Partly it is caused by different definition of the rock and shrub classes, partly by loss of area due to aggregation. Some differences are caused by classification errors in ALPMON.

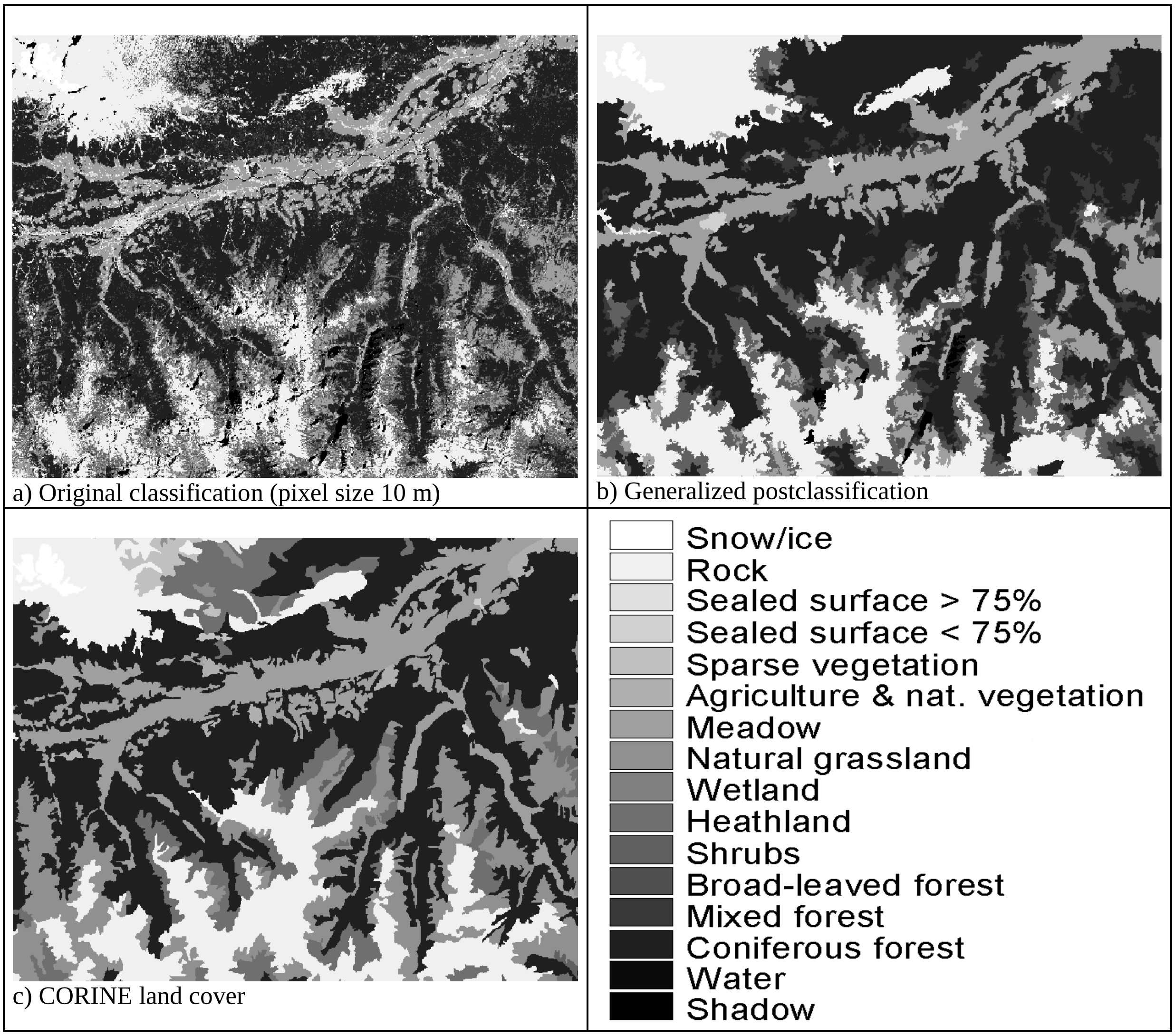
Confusion of ***meadow*** with *coniferous forest* in most cases is caused by interpretation errors of CORINE. But also the different generalization of both data sets leads to significant deviations. Figure 1 gives an example of two meadows in the lower part of the image, which have been identified in CORINE and correctly classified in ALPMON, but which disappear after post-classification/ generalization of the ALPMON classification result.

Additionally, the following differences or peculiarities attract attention:

- *Sealed surfaces* are completely different in the two data sets. This on one hand is due to the aggregation/generalization process applied to the ALPMON classification result, which preserves only relatively large settlements (small towns). In the CORINE data set additionally some smaller settlements (villages) were delineated. On the other hand obvious errors were discovered in the CORINE data set: the largest town is missing and sealed surface was indicated at a place covered by forest, meadow, and an artificial lake.

- *Water bodies* in both data sets are dropped due to the generalization or mapping units respectively.

- *Agriculture* hardly occurs in the test area, apart from meadows and pastures. Nevertheless, some areas have been interpreted as agricultural land in the CORINE land cover map, even though they don't differ from the areas interpreted as meadows (e.g. in Figure 2).



a) Original classification (pixel size 10 m)
b) Generalized postclassification
c) CORINE land cover

Figure 3: ALPMON and CORINE land cover maps

5 CONCLUSIONS

The agreement between the original classifications and the CORINE land cover maps of the test sites equals 53%. This "poor" agreement is partly due to the different degrees of generalization of the land cover maps. While in CORINE the minimum mapping unit is 25 ha (= 250000 m^2), the ALPMON classifications refer to the single pixel, i.e. an area between 25 m^2 and 900 m^2 depending on the image data applied. In order to overcome this contradiction of scale the ALPMON classification results were post-processed using a generalization technique. Generalization of the ALPMON classification has its clear impact on the degree of agreement that could be increased by post-processing to 64%. Although an increase of 11% can be considered significant, and proves the effectiveness of the post-processing method, there is still about 40% of confusion between the data sets that have to be explained. A part of it might be caused by the different approaches towards generalization. Visual interpretation usually defines the dominant class in an image as "background" and "cuts out" the remaining classes. The post-processing technique is always limited to the local neighborhood and therefore does not consider dominant classes in a larger environment.

A major reason for the confusion between the data sets are the differences in the nomenclatures, despite the effort of harmonization, or in the interpretation of the nomenclature. This is stated in the confusion between forest and shrubland or between meadows / pastures, sparse vegetation and rocks. Here the problem is a clear definition of class limits, both in terms of space and theme. Delineation of land cover types is critical, if the spatial transition is continuos, e.g. between forest and shrubland. Differentiation of land cover types is critical, if the thematic transition is continuos, e.g. between sparse vegetation and rocks. A significant difference can also be found in the interpretation of forest types. Mixed forest in ALPMON is to a high extent interpreted as coniferous forest in CORINE, while parts of coniferous forest in ALPMON are confused with natural grassland and heathland in CORINE. The latter is due to the fact that dwarf mountain pine was defined as heathland in the CORINE nomenclature. The over estimation of coniferous forest at costs of mixed forest in CORINE – as confirmed by aerial images – might be due to the differences in the original minimal mapping unit, i.e. the different approaches to generalization.

Final conclusions lead to the impression that the different approaches cannot be directly combined, at least not on level 3 of the CORINE nomenclature. Comparing the land cover maps on level 2 could lead to spatial indicators that might help the interpreter to concentrate on areas that are likely to have changed, or have been incorrectly assigned in the first version of the CORINE land cover map. The final decision on the land cover type, however, should be left to the interpreter. Such a combined approach would help to overcome the inconsistencies that appear in visual interpretation, such as overlooking single land cover units, but at the same time would benefit from the human vision system for detailed interpretation.

The introduction of level 4 classes to the CORINE nomenclature is considered possible for forest areas. While CORINE only differentiates three forest classes – coniferous, mixed and broad-leaved forest – the ALPMON classification takes into account forest types, age classes and crown coverage. However, it is to be noted that with an increasing level of detail the reliability of these classes is significantly decreasing. No conclusions can be given for other level 4 classes, such as for artificial surfaces or agricultural areas, as these land cover types hardly occur in alpine areas and thus were not of major interest for the ALPMON project.

6 OUTLOOK

The statements in this paper result from the comparison of two classifications (CORINE and ALPMON) which were based on different nomenclatures. If the semi-automatic classification was based on the CORINE nomenclature from the beginning, most of the problems due to different nomenclature and transfer of one classification result to another nomenclature could be solved. Although this was not investigated within the current study, significantly better results can be expected under this assumption.

As stated in section 2, a separation of the low growing dwarf mountain pines from other coniferous forests would significantly improve the available information. This separation could not be reached by classification of multi-spectral satellite data from summer due to similar spectral characteristics of these classes. However, in a small test area an attempt has been made to achieve this separation by classification of a panchromatic IRS-1D winter image. A situation was selected, where the snow has already fallen from the branches of the high growing trees, but the low growing dwarf mountain pine as well as greenalder were still covered by snow. The result was very promising, showing that with a multi-temporal approach the classification of dwarf mountain pine as well as greenalder may be reached with high accuracy.

ACKNOWLEDGEMENTS

This work has been supported by the European Commission, Contract No. ENV4-CT96-0359.

REFERENCES

ALPMON1997. ALPMON project homepage, Graz, Austria. http://dib.joanneum.ac.at/alpmon.

EUR 12585 - CORINE land-cover project - Technical guide, Luxembourg, 1993.

Steinnocher, K., 1996. Integration of spatial and spectral classification methods for building a land-use model of Austria. In: International Archives of Photogrammetry and Remote Sensing, Vienna, Austria, Vol. XXXI, Part B4, pp. 841-846.

Waser, L., Catalini, M., Schardt, M., Schmitt, U., Zini, E. 2000. Inventory of Alpine-relevant parameters for an Alpine monitoring system using remote sensing data. In: International Archives of Photogrammetry and Remote Sensing, Amsterdam, The Netherlands, Vol. XXXIII, in prep.