

# The thematic accuracy of Corine land cover 2000

## Assessment using LUCAS (land use/cover area frame statistical survey)

ISSN 1725-2237





# The thematic accuracy of Corine land cover 2000

## Assessment using LUCAS (land use/cover area frame statistical survey)

Cover design: EEA  
Layout: EEA

**Legal notice**

The contents of this publication do not necessarily reflect the official opinions of the European Commission or other institutions of the European Communities. Neither the European Environment Agency nor any person or company acting on behalf of the Agency is responsible for the use that may be made of the information contained in this report.

**All rights reserved**

No part of this publication may be reproduced in any form or by any means electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without the permission in writing from the copyright holder. For translation or reproduction rights please contact EEA (address information below).

Information about the European Union is available on the Internet. It can be accessed through the Europa server ([www.europa.eu](http://www.europa.eu)).

ISBN 92-9167-844-9  
ISSN 1725-2237

© EEA, Copenhagen, 2006

European Environment Agency  
Kongens Nytorv 6  
1050 Copenhagen K  
Denmark  
Tel.: +45 33 36 71 00  
Fax: +45 33 36 71 99  
Web: [www.eea.europa.eu](http://www.eea.europa.eu)  
Enquiries: [www.eea.europa.eu/enquiries](http://www.eea.europa.eu/enquiries)

# Contents

---

<b>Acknowledgements .....</b>	<b>4</b>
<b>Executive summary .....</b>	<b>5</b>
<b>1 Introduction.....</b>	<b>7</b>
1.1 Background .....	7
1.2 Purpose .....	8
<b>2 Databases .....</b>	<b>9</b>
2.1 CLC2000 data .....	9
2.2 Image2000 data .....	9
2.3 LUCAS 2001–2002 data .....	9
2.4 Setting up the database .....	9
2.5 CLC2000 cross-tabulated with LUCAS.....	11
2.6 Representativeness .....	15
<b>3 Methodology .....</b>	<b>18</b>
3.1 Automatic comparison .....	18
3.2 Evaluation based on reinterpretation of LUCAS photographs.....	19
<b>4 Results of the automatic approach.....</b>	<b>21</b>
<b>5 Results of the reinterpretation.....</b>	<b>24</b>
5.1 Reliability of CLC2000 .....	24
5.2 Analysis of errors .....	26
5.3 Subjectivity of photo interpretation.....	28
<b>6 Remarks on LUCAS data quality .....</b>	<b>29</b>
<b>7 Conclusions.....</b>	<b>30</b>
<b>References .....</b>	<b>32</b>
<b>Annexes overview .....</b>	<b>33</b>
<b>Annex 1 Corine land cover nomenclature.....</b>	<b>34</b>
<b>Annex 2 LUCAS land cover nomenclature.....</b>	<b>35</b>
<b>Annex 3 LUCAS land use nomenclature.....</b>	<b>36</b>
<b>Annex 4 Results of cross-tabulating CLC2000 and LUCAS LC .....</b>	<b>37</b>
<b>Annex 5 Results of cross-tabulating CLC2000 and LUCAS LU .....</b>	<b>38</b>
<b>Annex 6 Correspondence table between CLC classes and LUCAS land cover and land use classes .....</b>	<b>39</b>
<b>Annex 7 Reclassification of Image2000 using LUCAS: error matrix.....</b>	<b>40</b>
<b>Annex 8 Reclassification of Image2000 using LUCAS: subjective matrix .....</b>	<b>41</b>
<b>Annex 9 Illustrations (screen shots taken during reinterpretation – see Section 3.2) .....</b>	<b>42</b>

# Acknowledgements

---

The CLC2000 project is part of the work programme of the European Topic Centre on Terrestrial Environment (ETC-TE), working under contract with the European Environment Agency (2001–2005). LUCAS data for the study were provided by Eurostat and the national statistical offices in Estonia, Hungary and Slovenia. The authors express their gratitude to Chris Steenmans, EEA project manager, Adriana Gheorghe, EEA-Eionet coordinator, and Stefan Kleeschulte, ETC-TE manager for continuous cooperation

and support. Thanks also to Vanda Lima and Javier Gallego (JRC) for the useful comments on the validation methodology. Special thanks to Mária Bíró, who performed the reinterpretations and László Oszkó for his help on statistical considerations. Linguistic corrections were provided by Barbara Kosztra.

Prepared by George Büttner and Gergely Maucha, ETC/TE partner FÖMI.

# Executive summary

---

Corine land cover 2000 (CLC2000) was produced by the European Environment Agency (EEA) and its member countries in the European environment information and observation network (Eionet), based on the results of IMAGE2000, a satellite imaging programme undertaken jointly by the Joint Research Centre of the European Commission and the EEA. CLC2000 is an update for the reference year 2000 of the first Corine land cover database which was finalised in the early 1990s as part of the European Commission programme to COoRdinate Information on the Environment (Corine). The resulting national land cover inventories were integrated into a seamless land cover map of Europe, which is based on a standard methodology and nomenclature, providing consistent information on land cover across Europe.

LUCAS (European Land Use/Cover Area Frame Statistical Survey) is a project managed by the Eurostat, which main purpose is to provide harmonised information on agri-environment for Europe. LUCAS records land use (LU) and land cover (LC) information and stores digital field photographs in a two-level regular grid.

The purpose of this report is to assess the thematic accuracy of CLC2000 by means of an independent data source, the LUCAS data (not used in the compilation of the target database – CLC2000) and statistical methods homogeneously applied to as many of the participating countries as possible. The purpose of validating geographical data is to derive final accuracy/reliability figures by means of independent, high resolution and more accurate data, which is related to a similar data acquisition period.

LUCAS data is the only information that is available in many countries (18) for a European-wide validation of CLC2000. LUCAS fulfils the criteria of validation data as a result of its following characteristics:

- high geometric accuracy;
- data acquisition nearly coincident with CLC2000;
- independent, as not used during production of CLC2000.

The specific questions that the validation procedure is supposed to answer are:

- Was the 85 % thematic accuracy target of CLC2000 fulfilled?
- Which classes were determined with high accuracy and which ones with low?
- What were the reasons for misinterpretations?
- How subjective was photo-interpretation?

Two kinds of methods were used for validating CLC2000 data:

- Reinterpretation of IMAGE2000 data based on LUCAS codes and photographs;
- Automatic comparison of CLC2000 codes and LUCAS LU and LC codes.

The main result of the reinterpretation approach is that the total reliability of CLC2000 is  $87.0 \pm 0.8$  %, which leads us to the conclusion that the 85 % accuracy requirement specified in the Technical Guidelines of CLC2000 was fulfilled.

The result of the automatic comparison was that the percentage of total agreement (PTA) found between CLC2000 and LUCAS LU and LC is  $74.8 \pm 0.6$  %, meaning that CLC2000 approximates LUCAS thematic data with a 74.8 % average accuracy.

The analysis of results at class level shows that only 22 of the 44 classes of CLC2000 could be validated. The 22 classes that could not be validated belonged mostly to artificial surfaces, wetlands and water.

The highest class-level reliability (> 95 %) was obtained for rivers, lakes, industrial and commercial units and discontinuous urban fabric.

The two largest CLC classes (arable land and coniferous forest) were estimated to have a high level of reliability (between 90 and 95 %). Two other agricultural classes also enjoyed a high level of reliability: agro-forestry and permanently irrigated land.

The lowest class-level reliability (below 70 %) was obtained for the sparse vegetation class, which highlights the difficulties in interpreting this category.

Main sources of mistakes included using the wrong code (commission errors) and not interpreting enough details (omission errors). Geometrical inaccuracies played a less important role.

The majority of classification errors (78 %) occurred at level-3 and level-2 of the classification. Level-1 misclassifications mostly occur between the agriculture and the forest and semi-natural classes.

The analysis revealed that subjectivity of photo-interpretation could be noticed in 18 % of the samples. The most subjective CLC classes are, as follows: agriculture with significant amount of natural vegetation, transitional woodland, shrub, complex cultivation patterns and mixed forest.

The validation of CLC2000 leads us to think of some potential applications of the results. In the next update of CLC, special attention should be paid to the less accurate classes which mean that there is a need for improvement of the definition of mapping rules and the use of multi-temporal satellite data during interpretation. Also, positive examples of simultaneous display of IMAGE2000, CLC2000 and LUCAS data could be used in CLC related training/demonstration activities. And finally, the European LUCAS survey in 2006 could benefit from the results by improving consistency of LU/LC codification and producing better field photographs.



# 1 Introduction

## 1.1 Background

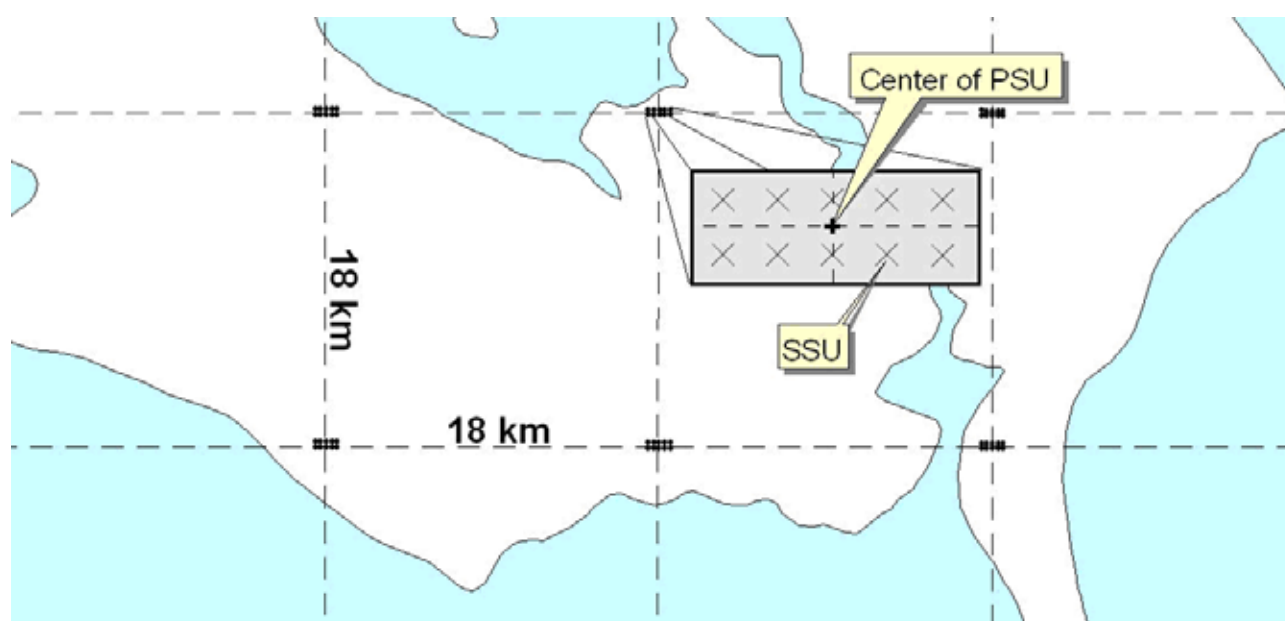
Corine land cover 2000 (CLC2000) is a project jointly managed by the Joint Research Centre (JRC) and the European Environment Agency (EEA). Its aim is to update the Corine land cover database in Europe for the year 2000. Landsat-7 Enhanced Thematic Mapper (ETM) satellite images were used for the update and were acquired within the framework of the Image2000 project. The first Corine land cover (CLC) inventory for the EU-15 and most of the new Member States was implemented between 1985 and 1996. It was carried out in order to characterise the land surface. A uniform nomenclature across Europe at scale 1:100 000 was used. The CLC nomenclature basically includes land cover items though land use elements can also be found. This is especially the case for built-up environments (Heymann *et al.*, 1994; Bossard *et al.*, 2000). The CLC database is a digital map covering countries with a seamless polygon database with 25 ha minimum mapping unit (MMU). The CLC nomenclature includes 44 categories in five major groups (Annex 1). In addition to the updated CLC2000 there are also other national deliverables: land cover changes between the 1990s and 2000 (CLC-Change), revised

first CLC inventory (revised CLC90) and metadata (EEA-ETC/TE, 2002).

European land use/cover area frame statistical survey (LUCAS) is a project managed by Eurostat. Its main purpose is to provide harmonised information on agri-environment for Europe. LUCAS (as implemented in 2001–2003) recorded land use and land cover information in a two-level regular grid. The grid size was 18 x 18 km (primary sampling unit, PSU). Each PSU included 10 secondary sampling units (SSUs). SSUs were placed in two parallel rows, with five points in each row. The distance between SSU points was 300 m (Figure 1). The field surveyor located the point with a high level of accuracy (specification states 1–3 m), and registered LC and LU information related to this point. This was conducted according to strict guidelines (Avikainen *et al.*, 2003). The area of observation was usually a circle with a 1.5 m radius. However, in exceptional cases an area with a 20 m radius was considered (Duhamel *et al.*, 2003).

Usually, in each central SSU point of the first row (SSU13), digital field photographs were taken in east, west, north and southwards directions.

**Figure 1** The LUCAS two-stage sampling (LUCAS technical document No 1: sampling plan)



The LUCAS LC nomenclature includes 57 categories in seven major groups (Annex 2), while LUCAS LU codification applies 14 land use classes (Annex 3).

A new LUCAS survey with a denser grid will take place in 2006 in 23 Member States (Eurostat, 2005).

### 1.2 Purpose

The purpose of validating geographical data is to derive final accuracy/reliability figures by means of independent, high resolution and more accurate data, which are related to a similar data acquisition period. Furthermore, they were not used in the compilation of the target database. Some countries performed a national validation of their CLC2000 database using country-specific approaches. In most cases, more spatially and thematically detailed data were used for that purpose (for example, see final reports in the Netherlands, Italy, Finland, Sweden, Portugal and the United Kingdom). Validation in this context means thematic validation as the geometry of CLC2000 is too trivial to check by comparison with Image2000.

The general aim of validating CLC2000 is to assess its thematic accuracy by means of a statistical method which is homogeneously applied to as many of the participating countries as possible.

LUCAS data are the only information available in many countries for a European-wide validation of CLC2000. LUCAS fulfils the criteria of validation data as a result of its following characteristics:

- high geometric accuracy;
- high thematic accuracy (LC/LU data based on field checking or, in exceptional cases, on photo interpretation of aerial ortho-photographs);
- data acquisition nearly coincide (although in the worst case a three-year difference is possible);
- independent, as usually not used during production of CLC2000.

The specific questions of the validation are the following.

- Is the 85 % target accuracy (Corine land cover update, I&CLC2000 project, Technical guidelines, Final version, August 2002) fulfilled?
- Which classes are determined with high accuracy and which ones with low accuracy?
- What are the reasons for misinterpretations?
- How subjective is photo interpretation?

Due to lack of similar reference data with wide coverage for the period of the first CLC inventory (1985–1996), the two other deliverables of the CLC2000 project (CLC-Change and revised CLC90) are not validated.

The results of the validation are made available for the users of CLC2000, and could be especially valuable in planning the next update of Corine land cover.

## 2 Databases

Three European databases were used in the validation process.

### 2.1 CLC2000 data

By the time this report was prepared, 29 European countries had finished CLC2000 (EU-25, Bulgaria, Croatia, Liechtenstein and Romania). National CLC2000 databases were produced by national teams based on Image2000 satellite imagery. Most countries used computer-assisted photo interpretation. Finland, Sweden and the United Kingdom applied a more computer-oriented approach, tailored to the geography and available databases in those countries.

On behalf of the EEA, the European Topic Centre on Terrestrial Environment (ETC/TE) was responsible for quality control, quality assurance and management of national CLC2000 databases and, finally, the production of the European products. Geometric precision (localisation accuracy) of CLC2000 is better than 100 m, according to specifications. Planned overall thematic accuracy is better than 85 % (EEA-ETC/TE, 2002). CLC2000 data are available for the public on the EEA data service (<http://dataservice.eea.europa.eu>). Technically controlled final national CLC2000 databases for the purpose of this work were provided by ETC-TE partner, GISAT.

### 2.2 Image2000 data

Image2000 data means orthorectified Landsat-7 satellite imagery, taken in the period 1999–2001 (2002) covering countries participating in the CLC2000 project. Image2000 data were produced by Metria (covering EU-15) and GISAT (covering the new Member States, Bulgaria and Romania). Metria processed images covering Croatia within the framework of the EU Life programme. Image2000 data in national projections were provided by the EEA.

### 2.3 LUCAS 2001–2002 data

The first national LUCAS surveys were implemented in 13 EU Member States in 2001. The

United Kingdom and Ireland were left out from this survey because of foot and mouth disease. These countries were surveyed in 2002 together with three (at that time) accession countries (Estonia, Hungary, and Slovenia). Eurostat was responsible for coordinating the national LUCAS surveys. The total land area defined by administrative boundaries of countries and coastlines has been divided into grids of 18 km at national level in Universal Transverse Mercator (UTM) projection. Comparison of some parameters of CLC and LUCAS can be found in Table 1.

Data covering the EU-15 were provided by Eurostat for the purposes of this validation. LUCAS data covering the three other countries have been provided by the national statistical offices.

The geographic coverage of the study is shown in Figure 2. The combined area of the 18 countries represents 77.5 % of the total area of 29 countries participating in CLC2000. The number of LUCAS PSUs in each country can be found in Table 2. The last column in Table 2 ('LUCAS, % country') shows that 79 % of the area of 18 countries is covered by LUCAS PSUs with at least one photo. Over larger lakes, in high mountains, and in areas of the far North, no LUCAS field photographs were taken.

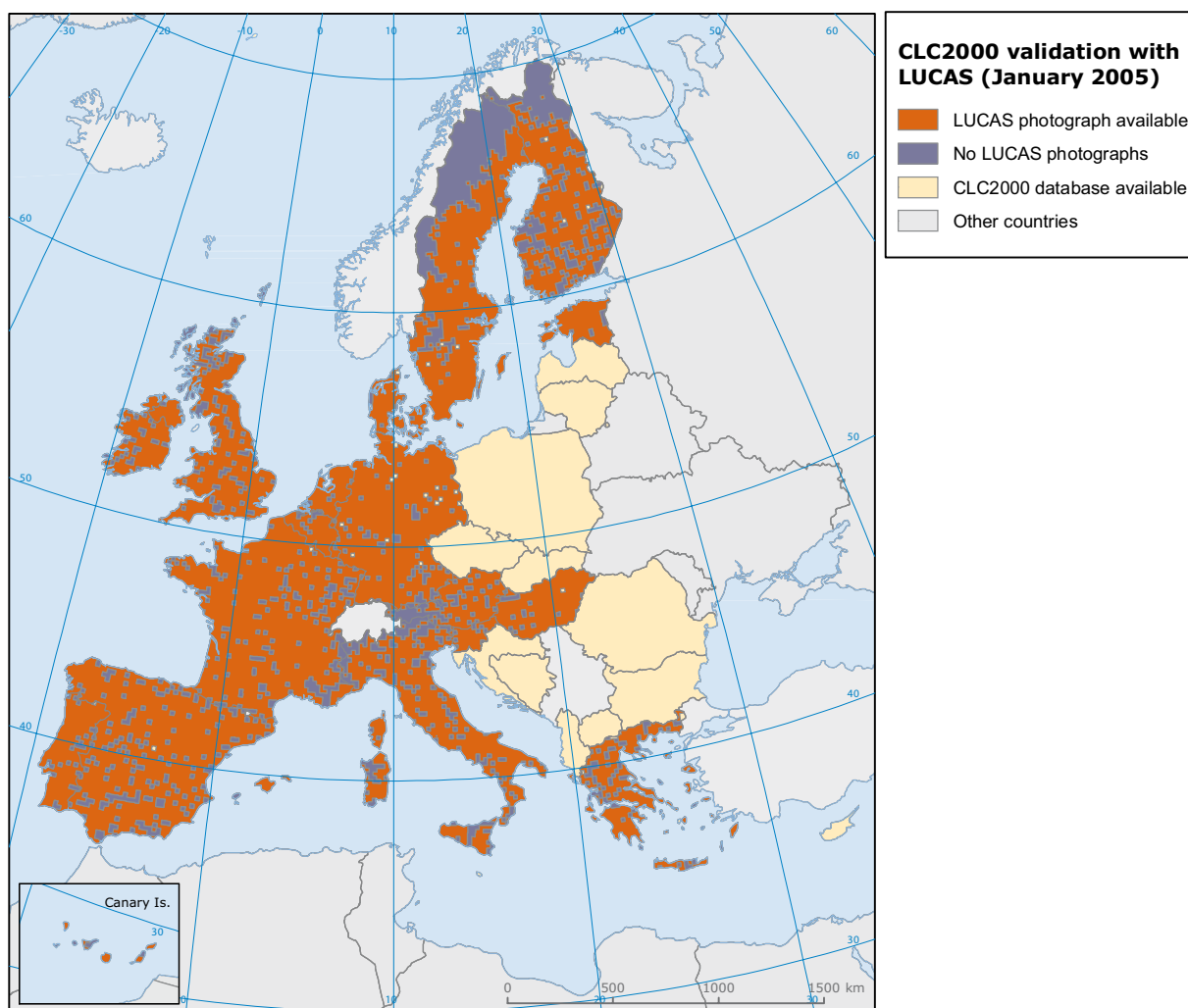
### 2.4 Setting up the database

In order to be able to view and handle LUCAS and CLC data, a GIS environment was set up under ESRI ArcView 3.2 software. Main steps of this process were as follows (Maucha *et al.*, 2003):

- the position of LUCAS SSUs was transformed from ETRS89 into national projection, in each of the 18 countries;
- LUCAS data tables provided in MS/Access database (EU-15) or a formatted text file (Estonia, Hungary, Slovakia) were imported into ESRI database format and joined to coordinates using IDs of SSU points;
- landscape photographs were joined to the coordinates of the corresponding SSU point, so photos could be displayed automatically based on SSU IDs (see Illustration 35 in Annex 9);

**Table 1 Comparison of main parameters of CLC1990/2000 and LUCAS 2001–2002**

	<b>Corine land cover</b>	<b>LUCAS</b>
Coverage	EU-25, Bulgaria, Croatia, Liechtenstein, Romania	EU-15 (2001 and 2003), Estonia 2002, Hungary 2002/03, Slovenia 2002
Database characteristics	Single land cover database with land use elements	Separate land use and land cover databases; field photographs
Method of database production	Mapping, based on satellite images and topographic maps	Area frame sampling in 18 x 18 km grid (PSU), 10 sampling points in each grid element (SSU)
Nomenclature	Land cover (with some land use elements); 44 classes in five major groups	Land cover: 57 classes in seven groups, emphasis on agriculture Land use: 14 classes in four groups
Observation unit	MMU of 25 ha (250 000 m <sup>2</sup> ), minimum width 100 m	7 m <sup>2</sup> (0.1 ha = 1 000 m <sup>2</sup> in exceptional cases)
Coding	Single level-3 code in each polygon	One (or two) LC code, One (or two) LU code in each SSU + additional environmental information
Geometric (localisation) accuracy	Better than 100 m	1–3 m
Thematic accuracy	Better than 85 %	Not known

**Figure 2 Geographic coverage of the study**

Source: CLC2000, EEA; LUCAS, Eurostat.

- the CLC2000 database and the Image2000 database were prepared for easy display simultaneously with LUCAS data. To reduce data volume, 20 km x 20 km size sub-images around LUCAS PSUs have been created from Image2000.

## 2.5 CLC2000 cross-tabulated with LUCAS

To have a rough idea how the CLC2000 and LUCAS inventories relate to each other, the two databases were cross-tabulated. This means that once the databases were in the same projection, the CLC2000 code and the corresponding LUCAS LC code for each of the SSUs were collected and summarised (Annex 4). A similar table is collected with CLC2000 codes and LUCAS LU codes (Annex 5). Most of the CLC codes have a rather long list of corresponding LUCAS codes. This is due to the differences in nomenclatures, different mapping rules (generalisation in CLC), geometrical shifts between the two datasets and, equally important, mistakes/inconsistencies in both databases.

Tables 3–7 show how the CLC classes correspond to LUCAS LC and LU codes. To provide an easier overview, only the majority (including at least 80 % of all SSUs) of LUCAS codes belonging to a given CLC class are included. Under the heading 'represented class' we see, in descending order of importance, which LUCAS codes (LC and LU) coincide with a CLC code. For example, 112 (continuous urban fabric) is represented in LUCAS LC mostly as buildings with one to three floors. This is followed by permanent grassland, non-built-up areas, non-built-up linear features (i.e. roads), woods, forests and buildings with more than three floors. For LUCAS LU, the most important uses are residential, then agriculture, transport and recreation, leisure and sport. All these correspond to the definition of CLC class 112.

Although all classes in 'Artificial surfaces' fit LUCAS data (Table 3) rather well, three points need to be made:

- ports (123) and dump sites (132) are represented by a very small number of SSUs only;

**Table 2** Number of LUCAS PSUs with photos used to validate CLC2000

Country	Area (km <sup>2</sup> )	Year of LUCAS data	No of PSUs with photos	PSUs, % of total	LUCAS, % country
Austria	83 850	2001	185	2.25	71
Belgium	30 520	2001	90	1.09	96
Denmark	43 090	2001	126	1.53	95
Estonia	45 226	2002	135	1.64	97
Finland	338 130	2001	687	8.35	66
France	551 500	2001	1 500	18.22	88
Germany	356 910	2001	1 016	12.34	92
Greece	131 990	2001	326	3.96	80
Hungary	93 030	2002	272	3.30	95
Ireland	70 283	2002	185	2.25	85
Italy	301 270	2001	708	8.60	76
Luxembourg	2 590	2001	7	0.09	88
Netherlands	35 398	2001	105	1.28	96
Portugal	92 390	2001	248	3.01	87
Slovenia	20 273	2002	55	0.67	88
Spain	504 780	2001	1 078	13.10	69
Sweden	449 960	2001	899	10.92	65
United Kingdom	244 880	2002	609	7.40	81
<b>Total</b>	<b>3 396 070</b>		<b>8 231</b>	<b>100</b>	<b>79</b>

- only a small number of LUCAS samples (< 100) represent five other CLC classes;
- it is noticeable, that the CLC class 133 (construction sites) does not have a perfectly corresponding LUCAS LU code (U33 = construction is missing from the list).

All classes in 'Agricultural areas' fit well with LUCAS data (Table 4). The match in LU is especially clear, for example, class 244 (agro-forestry) is represented in LUCAS LU as agriculture and forestry. The following remarks are also important.

- The significant share (12 %) of LUCAS permanent grassland class in the CLC 'arable land' class (211) seems to indicate an interpretation problem.
- The appearance of LUCAS 'maize' class under CLC class 213 (rice fields) is definitely a mistake (however, the number of points is small).

- The appearance of LUCAS LC forest classes in most agricultural classes of CLC can be explained by the generalisation, and not necessarily by mistakes.
- It is worth mentioning that the most frequent LUCAS LC class in CLC class 241 (annual crops associated with permanent crops) is olives. This class fulfils the expectations (olives inside the arable land field). However, a large share of 'unused' LUCAS LU class might indicate mistakes.
- It also shows clear agreement that class 244 (agro-forestry) is dominated by LUCAS LC class 'broadleaved forest'. This conforms to the definition of 244 (forest trees inside arable land).
- The LUCAS LC class, 'temporary pasture' appears in six of the 11 CLC agriculture classes, which illustrates the difficulties of interpreting this class in CLC.

**Table 3 CLC2000 'Artificial surfaces' cross-tabulated with LUCAS (figures in parentheses refer to percentage of SSUs)**

CLC	SSU	LUCAS LC	LUCAS LU
		Represented classes	Represented classes
111	129	Buildings with one to three floors (34), buildings with more than three floors (24), non-built-up areas (12), non-built-up linear features (12)	Residential (50), transport (18), commercial (5), construction (5), unused (5)
112	2 738	Buildings with one to three floors (23), permanent grassland without (14) and with (12) shrubs, non-built-up areas (12), non-built-up linear features (12), woods (4) forest (2), buildings with more than three floors (2)	Residential (43), agriculture (16), transport (12), recreation-leisure-sport (10)
121	424	Non-built-up areas (25), buildings with one to three floors (19), non-built-up linear features (13), buildings with more than three floors (8), permanent grassland without (7) and with (7) tree/shrub, woods (4)	Industry (21), transport (21), commerce (13), agriculture (11), unused (9), residential (8)
122	41	Non-built-up linear features (37), permanent grassland with and without tree/shrub (24), non-built-up areas (12), buildings with one to three floors (7)	Transport (54), unused (24), agriculture (12)
123	8	Non-built-up areas (25), non-built-up linear features (25). Other LLC classes represented only by one sample	Transport (50), industry (25), recreation-leisure-sport (13)
124	49	Permanent grassland without tree/shrub (61), non-built-up linear features (23)	Transport (69), community services (20)
131	90	Bare land (33), broadleaved forest (11), non-built-up areas (13), inland water bodies (9), forest (7), permanent grassland without tree/shrub (6), shrubland without tree cover (6), non-built-up linear features (6)	Mining and quarrying (38), unused (14), forestry (13), agriculture (10), water and waste treatment (7)
132	12	Bare land (58). Other LLC classes represented only by one sample	Mining and quarrying (42), unused (25), agriculture (17)
133	45	Permanent grassland with and without tree/shrub (36), bare land (11), fallow land (9), non-built-up area (9), shrubland (7)	Unused (27), recreation-leisure-sport (22), agriculture (18), residential (16)
141	52	Permanent grassland with sparse tree/shrub (35), forest (29), non-built-up linear features (12), buildings with one to three floors (6)	Recreation-leisure-sport (38), community services (17), residential (12), forestry (12), transport (10)
142	222	Permanent grassland without tree/shrub (33), permanent grassland with sparse tree/shrub (21), non-built-up area features (8), non-built-up linear (7), broadleaved woods (6), mixed forest (10)	Recreation-leisure-sport (50), agriculture (19), residential (12)
<b>Total</b>	<b>3 810</b>		

All classes in the 'Forests and semi-natural areas' match LUCAS data (Table 5). The match with LU is especially clear, for example, class 324 (transitional woodland shrub) is represented in LUCAS LU as forestry (i.e. under forest management) and unused (i.e. not under forest management). The following remarks are also important.

- Concerning the forest classes, 311 (broadleaved) and 312 (coniferous) are most frequently represented by their LUCAS LC counterpart. This is not the case with 313 (mixed forest), where the most frequent coincidence is 'coniferous'. This means that there is a difference if the area is observed from above (CLC) and on the field (LUCAS).
- Burnt area (334), glaciers (335) and beaches, dunes and sand (331) are represented in LUCAS by a few samples only.

- The appearance (10 %) of LUCAS LC wetlands in CLC class 324 (transitional woodland shrub) is explained by Nordic landscapes, where peat-bogs under natural colonisation are frequent.

All classes in 'Wetlands' fit well with LUCAS data (Table 6). The match in LU is especially clear with the dominance of 'unused' class. The following remarks are important.

- Very few SSUs characterise CLC classes 422 (salines) and 423 (intertidal flats).
- The presence of LUCAS LC vegetation and water classes are explained by CLC generalisation.
- The LUCAS LU code 'mining and quarrying' in case of CLC class 422 (salines) relates to salt extraction.

**Table 4** CLC2000 'Agricultural areas' cross-tabulated with LUCAS (figures in parentheses refer to percentage of SSUs)

CLC	SSU	LUCAS LC Represented classes	LUCAS LU Represented classes
211	21 388	Common wheat (18), barley (13), maize (9), permanent grassland without tree/shrub (9), temporary pastures (7), fallow land (6), durum wheat (4), rape seed (4), permanent grassland with sparse tree/shrub (3), sunflower (2), non-built-up linear features (2), broadleaved forest (2), oats (2), sugar beet (2)	Agriculture (87)
212	826	Maize (17), cotton (12), fallow land (8), barley (8), temporary pastures (6), durum wheat (5), non-built-up linear features (4), common wheat (4), sunflower (3), bare land (2), other fresh vegetables (2), vineyard (2), sugar beet (2), inland running water (2), buildings with one to three floors (2)	Agriculture (86)
213	104	Rice (70), maize (8), fallow land (6)	Agriculture (89)
221	904	Vineyards (55), shrubland without and with trees (7), permanent grassland without sparse tree/shrub (5), fallow land (4), other fruit trees (3), non-built-up linear features (3), olives (2), barley (2)	Agriculture(82)
222	574	Other fruit trees and berries (14), oranges (13), apples (11), nuts (10), shrubland without and with trees (7), permanent grassland with and without sparse tree/shrub (6), other citrus (4), fallow land (3), maize (3), pear (3), olives (3), non-built-up linear features (2), broadleaved forest (2)	Agriculture (78), unused (13)
223	1 196	Olives (65), shrubland without and with trees (6), fallow land (3), other fruit trees (2), durum wheat (2), permanent grassland with sparse tree/shrub (2)	Agriculture (84)
231	8 616	Permanent grassland without and with sparse tree/shrub (68), temporary pastures (6), broadleaved forest (3), maize (2), broadleaved woodland (2)	Agriculture (82)
241	271	Olives (12), shrubland with and without trees (13), durum wheat (8), permanent grass with and without sparse tree/shrub (7), fallow land (6), vineyards (6), temporary pastures (6), maize (6), non-built-up linear features (5), buildings with one to three floors (3), nuts (3), broadleaved forest (2), poplars, eucalyptus (2)	Agriculture (67), unused (17)
242	5 655	Permanent grassland without and with sparse tree/shrub (28), temporary pastures (7), maize (6), common wheat (5), shrubland with and without sparse trees (5), vineyards (4), fallow land (4), barley (4), olives (4), broadleaved forest (3), non-built-up linear features (3), broadleaved woodland (2), buildings with one to three floors (2), other fruit trees and berries (2)	Agriculture (76), unused (8)
243	3 937	Permanent grassland without and with sparse tree/shrub (21), shrubland with and without trees (13), broadleaved forest (12), temporary pastures (7), coniferous forest (6), mixed forest (5), fallow land (5), barley (3), olives (3), non-built-up linear features (3), common wheat (2)	Agriculture (51), forestry (22), unused (19)
244	875	Broadleaved forest (42), permanent grassland with and without sparse tree/shrub (32), fallow land (6)	Agriculture (58), forestry (37)
<b>Total</b>	<b>44 346</b>		

**Table 5 CLC2000 'Forests and semi-natural areas' cross-tabulated with LUCAS (figures in parentheses refer to the percentage of SSUs)**

CLC	SSU	LUCAS LC	LUCAS LU
		Represented classes	Represented classes
311	8 805	Broadleaved forest (56), mixed forest (11), shrubland with and without sparse trees, (8) coniferous forest (7)	Forestry (67), unused (20)
312	15 931	Coniferous forest (65), mixed forest (15)	Forestry (81)
313	6 441	Coniferous forest (37), mixed forest (29), broadleaved forest (14)	Forestry (75), unused (18)
321	2 777	Permanent grassland without tree/shrub (30), shrubland without trees (18), permanent grassland with sparse tree/shrub (16), bare land (15), shrubland with sparse tree cover (7)	Agriculture (53), unused (35)
322	2 709	Shrubland without trees (44), permanent grassland without tree/shrub (16), bare land (10), shrubland with sparse tree cover (8), wetland (5)	Unused (58), agriculture (25)
323	2 380	Shrubland without trees (36), shrubland with sparse tree cover (17), deciduous forest (10), bare land (8), permanent grassland with sparse tree/shrub (7), coniferous forest (5)	Unused (48), agriculture (30), forestry (16)
324	5 727	Coniferous forest (39), mixed forest (13), deciduous forest (10), wetlands (10), shrubland with sparse tree cover (7), wetland (7), shrubland without trees (7)	Forestry (59), unused (28)
331	75	Bare land (51), inland running water (15), shrubland without trees (9), coniferous forest (5)	Unused (60), community services (13), forestry (9)
332	564	Bare land (74), glaciers, permanent snow (8)	Unused (87)
333	957	Bare land (40), shrubland without trees (25), permanent grassland without tree/shrub (11), glaciers, permanent snow (6)	Unused (74), agriculture (16)
334	27	Coniferous forest (33), shrubland without trees (26), mixed forest (11), poplars, eucalyptus (7), permanent grassland with sparse tree/shrub (7)	Unused (48), forestry (33)
335	41	Glaciers, permanent snow (63), bare land (37)	Unused (100)
<b>Total</b>	<b>46 434</b>		

**Table 6 CLC2000 'Wetlands' cross-tabulated with LUCAS (figures in parentheses refer to percentage of SSUs)**

CLC	SSU	LUCAS LC	LUCAS LU
		Represented classes	Represented classes
411	181	Wetland (35), permanent grassland without and with sparse tree/shrub (19), inland water bodies (18), shrubland with and without sparse trees (11)	Unused (61), agriculture (19)
412	2 152	Wetland (61), coniferous forest (13), permanent grassland without shrubs (7)	Unused (68), forestry (13)
421	83	Wetland (59), permanent grassland without sparse tree/shrub (13), coastal water bodies (7), inland water bodies (5)	Unused (67), agriculture (24)
422	22	Coastal water bodies (73), wetland (18)	Mining and quarrying (73), unused (23)
423	5	Permanent grassland without sparse tree/shrub (40), bare land (40)	Unused (100)
<b>Total</b>	<b>2 443</b>		

**Table 7 CLC2000 'Water bodies' cross-tabulated with LUCAS (figures in parentheses refer to percentage of SSUs)**

CLC	SSU	LUCAS LC	LUCAS LU
		Represented classes	Represented classes
511	188	Inland running water (57), inland water bodies (10), coastal water bodies (5), broadleaved forest (4), bare land (3), broadleaved wood (3)	Unused (48), transport (27), recreation-leisure-sport (11)
512	2 646	Inland water bodies (91)	Unused (45), recreation-leisure-sport (33), fishing (14)
521	47	Coastal water bodies (70), wetland (17)	Unused (64), fishing (21)
522	27	Inland running water (48), coastal water bodies (41)	Unused (67), transport (30)
523	68	Bare land (31), coastal water bodies (26), shrubland without trees (13), inland water bodies (6), permanent grassland without tree/shrub (6)	Unused (54), fishing (19), recreation-leisure-sport (10)
<b>Total</b>	<b>2 976</b>		



Most classes in 'Water bodies' fit well with LUCAS data (Table 7). Concerning land use, the 'unused class' dominates. The following remarks are important.

- Few SSUs characterise CLC classes 521 (coastal lagoons) and 522 (Estuaries).
- The dominance of LUCAS LC class 'inland running water' in 522 (estuaries) and the presence of LUCAS LC 'coastal water bodies' class in CLC class 511 (rivers) probably relates to a definition problem, in other words, being unsure of where the river ends.
- The appearance of LUCAS LC vegetation classes in CLC classes 511 (rivers) and 523 (sea and ocean) could relate to geometric mis-registration.
- Slight geometric mis-registration might explain that the most frequent coincidence with CLC 523 class (sea and ocean) is LUCAS LC 'bare land'. It is important to note, however, that no LC class for sea water exists in LUCAS, and samples falling under the classification 'sea and ocean' are usually marked as 'not classified' (and excluded from this analysis).

## 2.6 Representativeness

When looking for the reliability estimates of CLC2000 using LUCAS data, we have to consider the representativeness of the estimates at class level or country level. The representativeness was calculated based on the assumption of binomial distribution. The binomial distribution represents a good approximation for the PSU level LUCAS sampling (see Section 3.2 on reinterpretation of LUCAS photographs). However, it is distorted at SSU level (see Section 3.1 on automatic comparison) due to the two levels of sampling, in other words, the samples cannot be considered as totally independent. The SSU level case study for Hungary shows that the standard deviation of the resulting distribution can be approximated by doubling the standard deviation of the binomial distribution (corresponding to the 'N' total number of samples) (Maucha *et al.*, 2003).

The expected value of the binomial distribution:  
 $M(x) = N \cdot p(x)$

The standard deviation:  $D(x) = \sqrt{N \cdot p(x) \cdot (1-p(x))}$

- where N is the total number of samples;

- p is the probability of success (corresponding to the relative area represented by a given class or a given country).

The confidence level of the estimation is 95 % if we consider an error of  $M \pm 2 \cdot D$ . Therefore we calculate:

the absolute error:

$$AE(x) = 2 \cdot D(x) / N$$

the relative error:

$$RE(x) = 2 \cdot D(x) / M(x)$$

We express both error rates in percentages (AE and RE multiplied by 100). In the case of the automatic comparison (SSU level), the above error rates are multiplied by a factor of two.

As seen above, the representativeness depends on the size of the class (or country) and the number of LUCAS samples available for validation. In case of a small CLC class (or country), the small number of LUCAS samples might provide misleading results because the error (standard deviation) of the estimation is high.

At the time of writing, the seamless European database was not yet available. The CLC2000 statistics (i.e. the area of each class) was computed covering the 18 countries using national data sets. As the photo interpretation continued a few hundred meters through the borders, these statistics included some inaccuracies (12 % over-estimation of total area) because the countries were not clipped around national boundaries. This may cause only a slight distortion in the probability of a given class. CLC class 523 (sea and ocean) was omitted in the calculations as not sampled by LUCAS.

Class level representativeness to support the LUCAS SSU-based comparison (see Section 3.1) is presented in Table 8. As mentioned above, the error of estimation is about double the figure that would be valid for the same amount of samples with regular distribution.

Table 8 shows that seven of the smallest European CLC2000 classes (highlighted in grey) can be estimated with LUCAS SSUs with only a relative error higher than 50 %.

Class level representativeness to support LUCAS PSU-based comparison (see Section 3.2) is presented in Table 9. As the number of PSUs is smaller than the number of SSUs, there are more CLC classes with low representativeness. In Table 9, the light blue rows highlight those classes that cannot be tested

**Table 8** Class level representativeness for CLC2000 validation based on LUCAS SSUs

Code	CLC2000 area (ha)	% area	Mean sample number (M)	Standard deviation (D)	Absolute error (AE)	Relative error (RE = 4D/M)
111	583 033	0.17	169	13.0	0.03	30.7
112	9 784 165	2.84	2 837	52.5	0.10	7.4
121	1 530 668	0.44	444	21.0	0.04	18.9
122	153 823	0.04	45	6.7	0.01	59.9
123	93 812	0.03	27	5.2	0.01	76.7
124	245 737	0.07	71	8.4	0.02	47.4
131	479 983	0.14	139	11.8	0.02	33.9
132	65 112	0.02	19	4.3	0.01	92.1
133	105 034	0.03	31	5.5	0.01	72.5
141	222 688	0.06	65	8.0	0.02	49.8
142	728 315	0.21	211	14.5	0.03	27.5
211	72 639 835	21.06	21 059	128.9	0.26	2.4
212	3 126 654	0.91	906	30.0	0.06	13.2
213	562 089	0.16	163	12.8	0.03	31.3
221	3 279 611	0.95	951	30.7	0.06	12.9
222	1 889 948	0.55	548	23.3	0.05	17.0
223	3 941 481	1.14	1 143	33.6	0.07	11.8
231	28 835 584	8.36	8 360	87.5	0.18	4.2
241	963 233	0.28	279	16.7	0.03	23.9
242	19 390 659	5.62	5 622	72.8	0.15	5.2
243	13 389 164	3.88	3 882	61.1	0.12	6.3
244	3 203 231	0.93	929	30.3	0.06	13.1
311	30 115 724	8.73	8 731	89.3	0.18	4.1
312	53 562 307	15.53	15 528	114.5	0.23	3.0
313	21 851 224	6.33	6 335	77.0	0.15	4.9
321	10 339 366	3.00	2 998	53.9	0.11	7.2
322	8 824 825	2.56	2 558	49.9	0.10	7.8
323	9 417 468	2.73	2 730	51.5	0.10	7.6
324	19 630 349	5.69	5 691	73.3	0.15	5.1
331	316 378	0.09	92	9.6	0.02	41.7
332	2 208 747	0.64	640	25.2	0.05	15.8
333	3 733 043	1.08	1 082	32.7	0.07	12.1
334	135 302	0.04	39	6.3	0.01	63.9
335	198 683	0.06	58	7.6	0.02	52.7
411	609 334	0.18	177	13.3	0.03	30.1
412	7 013 193	2.03	2 033	44.6	0.09	8.8
421	310 718	0.09	90	9.5	0.02	42.1
422	71 449	0.02	21	4.6	0.01	87.9
423	1 051 782	0.30	305	17.4	0.03	22.9
511	718 872	0.21	208	14.4	0.03	27.7
512	8 916 799	2.58	2 585	50.2	0.10	7.8
521	390 317	0.11	113	10.6	0.02	37.6
522	338 382	0.10	98	9.9	0.02	40.4
<b>Total</b>	<b>344 968 121</b>	<b>100.00</b>	<b>100 009</b>	<b>0.0</b>	<b>0.00</b>	<b>0.0</b>

**Note:** CLC codes printed on grey means > 50 % relative error.

(i.e. relative error higher than 100 %), while the grey rows show those classes that can be tested only with

low representativeness (relative error higher than 50 %).

Table 9 shows that six of the smallest European CLC2000 classes (highlighted in light blue) cannot be validated with LUCAS PSUs as the relative error exceeds 100 %.

Furthermore, 11 of the CLC2000 classes (highlighted in grey) can only be weakly validated as the relative error exceeds 50 % (but is less than 100 %).

**Table 9** Class level representativeness for CLC2000 validation based on LUCAS PSUs

Code	Area (ha)	% area	Mean sample number (M)	St.Dev. (D)	Absolute error (AE)	Relative error (RE = 2D/M)
111	583 033	0.17	13	3.7	0.09	54.4
112	9 784 165	2.84	226	14.8	0.37	13.1
121	1 530 668	0.44	35	5.9	0.15	33.5
122	153 823	0.04	4	1.9	0.05	106.0
123	93 812	0.03	2	1.5	0.04	135.7
124	245 737	0.07	6	2.4	0.06	83.8
131	479 983	0.14	11	3.3	0.08	60.0
132	65 112	0.02	2	1.2	0.03	162.9
133	105 034	0.03	2	1.6	0.04	128.2
141	222 688	0.06	5	2.3	0.06	88.1
142	728 315	0.21	17	4.1	0.10	48.7
211	72 639 835	21.06	1 681	36.4	0.91	4.3
212	3 126 654	0.91	72	8.5	0.21	23.4
213	562 089	0.16	13	3.6	0.09	55.4
221	3 279 611	0.95	76	8.7	0.22	22.8
222	1 889 948	0.55	44	6.6	0.16	30.2
223	3 941 481	1.14	91	9.5	0.24	20.8
231	28 835 584	8.36	667	24.7	0.61	7.4
241	963 233	0.28	22	4.7	0.12	42.3
242	19 390 659	5.62	449	20.6	0.51	9.2
243	13 389 164	3.88	310	17.3	0.43	11.1
244	3 203 231	0.93	74	8.6	0.21	23.1
311	30 115 724	8.73	697	25.2	0.63	7.2
312	53 562 307	15.53	1 240	32.4	0.80	5.2
313	21 851 224	6.33	506	21.8	0.54	8.6
321	10 339 366	3.00	239	15.2	0.38	12.7
322	8 824 825	2.56	204	14.1	0.35	13.8
323	9 417 468	2.73	218	14.6	0.36	13.4
324	19 630 349	5.69	454	20.7	0.51	9.1
331	316 378	0.09	7	2.7	0.07	73.9
332	2 208 747	0.64	51	7.1	0.18	27.9
333	3 733 043	1.08	86	9.2	0.23	21.4
334	135 302	0.04	3	1.8	0.04	113.0
335	198 683	0.06	5	2.1	0.05	93.2
411	609 334	0.18	14	3.8	0.09	53.2
412	7 013 193	2.03	162	12.6	0.31	15.5
421	310 718	0.09	7	2.7	0.07	74.5
422	71 449	0.02	2	1.3	0.03	155.5
423	1 051 782	0.30	24	4.9	0.12	40.5
511	718 872	0.21	17	4.1	0.10	49.0
512	8 916 799	2.58	206	14.2	0.35	13.7
521	390 317	0.11	9	3.0	0.07	66.5
522	338 382	0.10	8	2.8	0.07	71.4
<b>Total</b>	<b>344 968 121</b>	<b>100.00</b>	<b>7 985</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>

**Note:** CLC codes printed in light blue are non-representative, while grey means > 50 % error.

## 3 Methodology

A two-tier approach was applied to validate CLC2000. The two methods measure a completely different kind of reliability. These methods were:

- automatic comparison of CLC2000 and LUCAS LU/LC codes;
- reinterpretation of Image2000 with the help of LUCAS data (LU and LC codes and photographs) by following interpretation rules of CLC.

### 3.1 Automatic comparison

The methodology is introduced in the pilot study done by the ETC/TE in 2003 (Maucha *et al.*, 2003).

Although the nomenclatures of CLC and LUCAS are similar, they are not directly compatible. A correspondence table between CLC and LLC classes can be created. In this table only basic constituents of each CLC class are considered and not those additional elements that might be present in a given polygon because of the generalisation. As the CLC is not a pure land cover nomenclature – especially considering the first group of classes (artificial surfaces) – the LLU information should also be considered during comparison. Therefore, a similar table is constructed between CLC and LLU.

The automatic comparison includes three steps:

- Creation of the correspondence table between CLC and LLC, and CLC and LLU. Some elements of the correspondence table are

presented and explained in Table 10, while the full table is shown in Annex 6.

- GIS overlay of all SSUs and CLC2000. This is a similar step to that in the cross-tabulation (see Section 2.5).
- Comparison of codes using the correspondence table and calculation of the degree of agreement between the two databases. Each SSU is checked to see whether the CLC code corresponds to one or more LLC category(ies) in the correspondence table. If 'yes', the value of the 'lc\_ok' variable is set to one. Otherwise the value of 'lc\_ok' remains zero. Land use is handled in a similar way. Therefore, 'lu\_ok' is set to one if the CLC code corresponds to one or more of the LLU code(s) in the correspondence table. Otherwise, the value of 'lu\_ok' remains zero. The value of agreement ('ssu\_ok') is considered 'OK', if both the value of 'lc\_ok' and the value of 'lu\_ok' are equal to one. This means that both LUCAS codes (LLC and LLU value) should correspond to the CLC code. Finally, the 'ssu\_ok' values are calculated, and the percentage total agreement (PTA) is obtained.

For example, an SSU falling within CLC class 112 (discontinuous urban fabric) would be considered 'OK', if:

- the corresponding LLC class is one of the following: A11 (buildings with one to three floors), A12 (buildings with more than three floors), A2 (non-built-up area and linear features), B43 (other fresh vegetables), B44

**Table 10** Examples of correspondence between CLC and LUCAS classes

CLC code	Class name	Main constituents	Corresponding LLC codes (see Annex 2)	Corresponding LLU codes (see Annex 3)
112	Discontinuous urban fabric	Houses (irrespective of the number of floors), urban greenery, roads, parking lots, gardens around houses, roads	A11, A12, A2, B43, B44, C1, C2, E	U11, U31, U34, U35, U36, U37
242	Complex cultivation patterns	Arable land, pastures, vineyards, fruit trees, olives, smaller houses	A11, A13, A22, all B, except B17, B83 and B84; E	U11, U37
324	Transitional woodland, shrub	Clear cuts, young forest, diseased or damaged forest, forest nurseries	B83, C1, C2, C30, D01, E	U12, U40
512	Water bodies	Water, wetland	G01	U13, U21, U31, U36, U40

(floriculture and ornamental plants), C1 (forest area), C2 (other wooded area) or E (permanent grassland); and

- the corresponding LLU class is one of the following: U35 (community services), U37 (residential), U34 (commerce, finance, business), U31 (transport, communication, storage, protective works) or U40 (unused).

The PTA measure represents the level CLC2000 approximates in reality, which is modelled by LUCAS LC and LU codes. The benefits lie in the fast, automatic computation, and in the inclusion of a large number of points (more than 100 000 SSUs). These benefits arise because comparisons are made at SSU level, which provides better representativeness when compared with the other approach. A definite weakness is the dependence on the content of the correspondence table. PTA is expected to provide a pessimistic assessment of the CLC2000 data quality, as the inherent differences between CLC and LUCAS (observation unit, difference of scale, point versus area sampling) cannot be considered. Mistakes and inconsistencies in LUCAS LU/LC codes further decrease the value of PTA (see Chapter 6). As PTA is just a sum of binary decisions, it cannot distinguish between omission and commission errors, and an error (misclassification) matrix cannot be created. Results of automatic comparison are presented in Chapter 4.

### 3.2 Evaluation based on reinterpretation of LUCAS photographs

At least one LUCAS landscape photograph is available for each of the 8 231 locations out of the 10 258 PSUs observed in the 18 countries implementing CLC2000 (Table 2). Validation is based on a visual reinterpretation of Image2000 around the LUCAS PSU, in combination with the reference information provided by LUCAS: landscape photographs and the LLC and LLU information for the corresponding SSUs. As the sample number is smaller, a fewer number of CLC2000 classes can be validated with acceptable error, in other words, the number of representative classes will be smaller than in the case of the automated method (see Section 2.6). This method is much more time consuming than the automated one presented in Section 3.1 because of the involvement of a photo interpreter expert. A further weakness is that no other reference data are available for the validating expert (e.g. topographic maps, aerial photographs). The main advantages of this method are:

- CLC generalisation rules and complex class definitions can be considered;
- LLC and LLU codes in the 1 200 x 300 m area surrounding the central SSU (where the photos were taken) provide the spatial context;
- LUCAS field photographs provide a valuable 'ground truth' for reinterpretation.

Results provided by this method can be considered as thematic reliability figures of CLC2000, as national CLC2000 data are compared with an independent CLC2000 interpretation.

The reinterpretation includes the following three steps.

- (1) The validating expert provides the CLC code by visual interpretation of the location around the LUCAS sampling point where the LUCAS landscape photographs were taken. The interpretation is based on Image2000 data, LLU and LLC data and landscape photographs. Additional environmental information collected by LUCAS can also be consulted (e.g. irrigation, natural hazards). No information from the CLC database is shown at this moment. The validating expert should respect all the CLC interpretation rules (MMU, class definitions, generalisation, etc.) when providing the code (abbreviated as val-cod1). After confirmation, the code cannot be changed.
- (2) The CLC polygons are displayed on top of the satellite image and LUCAS points, but still without class information. This situation is interpreted again by assigning a second CLC code (abbreviated as val-cod2). After confirmation, the code cannot be changed.
- (3) The actual CLC code is displayed (taken from the CLC2000 database, abbreviated as 'act-cod'). By comparing the actual CLC code (act-cod) and the control codes (val-cod1 and val-cod2), the validating expert has to evaluate the situation as one of four cases (or 'decisions').

**Clear agreement:** This decision is used when agreement between the control interpretations and the CLC2000 is not in doubt. (The most trivial cases are those where the two control interpretations are the same as the actual CLC2000 code.) See Illustrations 1–9 in Annex 9.

**Another interpretation is possible:** This decision is used when the interpretation is still

acceptable, but the control codes do not fully agree with the CLC2000 code. The validating expert should select the explanation as to whether other delineation or another code or LC boundary is possible. See Illustrations 10–13 in Annex 9.

**Wrong interpretation:** This decision is used when the CLC2000 interpretation is not acceptable. (This is mostly in the case where none of the control interpretations are the same as the CLC2000 code.) The validating expert should select the explanation as to whether there is a wrong code, wrong generalisation, not enough details (omission) or inaccurate delineation. Inaccurate delineation means that the geometrical matching of Image2000 and CLC2000 deviates more than 100 m. See Illustrations 14–32 in Annex 9.

**Not enough information:** This decision is used in those rare cases when the available information is insufficient to make a well-based decision (e.g. cloudy Image2000, bad quality LUCAS photos, etc.). See Illustration 33 in Annex 9.

This methodology should give a more realistic evaluation of the database quality as it considers the mapping rules and the landscape context in which the CLC2000 database has been created. Cases of a combined 'Clear agreement' and 'Another interpretation is possible' provide cases of accepted photo interpretation. The methodology offers the possibility of measuring the inherent subjectivity of photo interpretation. Cases of 'Wrong interpretation' provide misclassifications and distinguish between omission and commission errors. The result of this analysis provides a reliability figure, in other words, at what level an independent CLC interpretation – based on the LUCAS photos and LU and LC codes – would provide the same interpretation of Image2000.

### 3.2.1 Implementation

The task of validation is rather complicated because the 18 countries represent highly varied natural

landscapes, different land use and agricultural and forestry practices. It was understood during this work that a certain 'learning period' was necessary for almost all the countries to be able to evaluate CLC2000. To reduce subjective differences, the entire validation was carried out by two experts using the following three phases.

- Base validation phase: The validating expert checked all the points in a country. In the mean time ('learning phase'), the expert discovered that some of the decisions were not correct. These could be corrected in the second step.
- Self-control phase: The validating expert had an optional possibility to modify the first evaluation. For this purpose it was possible to select specific CLC codes (e.g. act-code = 244, if the difficulties were associated with the 'agro-forestry' class). In this phase the interpreter always saw the codes (val-cod1 and val-cod2) he had given in the base validation phase.
- Independent control: In order to reduce subjectivity, another expert checked a reduced number of samples which were considered as 'problematic':
  - all samples which were evaluated as not 'clear agreement' in any of the previous two phases;
  - all samples which were evaluated as 'clear agreement', but neither 'val-cod1' nor 'val-cod2' were equal to 'act-cod'.

During this last phase, the former decisions are hidden, so as not to influence the evaluation. The evaluation given by this phase is considered as final (together with the 'non-problematic' cases of the first two phases).

This process means that over 32 % of the samples were evaluated independently by two experts.

## 4 Results of the automatic approach

The correspondence table between CLC and LUCAS LC and LU classes was built in the framework of the pilot study (Maucha *et al.*, 2003). It has been refined using the results of the cross-tabulation (see Section 2.5). The following remarks are related to the correspondence table (see Annex 6).

- Most of the CLC classes can be described in a relatively clear way by LLC and LLU, if combined.
- The distinction of the two kinds of shrubs (322: moors and heathland, and 323: sclerophyllous vegetation) is not possible by means of LUCAS. However, this is not really a problem as these two classes are usually geographically separated.
- There is no LUCAS class to characterise the CLC class 324 (transitional woodland shrub).
- Wetlands and peatlands are not distinguished in LUCAS.

CLC2000 and LUCAS LU and LC covering 18 countries have been compared in a correspondence table using almost 100 000 SSUs. SSUs with no coincident LLU codes were omitted. Results are shown in Tables 11 and 12 and Figure 3.

CLC classes 523 (sea and ocean) and 423 (intertidal flats) are omitted from this analysis as no LUCAS LC class exists for sea water, due to the fact that LUCAS does not sample the sea. Table 11 shows that out of 99 936 SSUs, 80.1 % corresponded to LUCAS LC, 88.6 % corresponded to LUCAS LU and 75.6 % of the SSUs agreed with both LUCAS codes. This was based on the correspondence table.

The calculation of representativeness is now a more complex lesson. Considering the entire territory of analysis one can calculate a probability 'P' as a combination of the probability of the given class and the probability of the agreement (PTA):

$$P = P_{\text{CLC}} * P_{\text{PTA}}$$

By analysing results at class level it can be seen that:

- three of the CLC classes (highlighted in light blue) cannot be validated with LUCAS SSUs, as the error of estimation exceeds 100 %;
- eight of the CLC classes (highlighted in grey) are weakly representative (relative error exceeds 50 %) because of their too few appearances in CLC2000 (see Section 2.6);
- the largest CLC2000 class in Europe (see Table 8) is arable land (211), which holds a 74.8 % PTA value — this means that if one selects a point in a polygon with a CLC2000 code 211, there is a 74.8 % chance that one will find one of the crops (or fallow land), or greenhouses, or non-linear features (i.e. roads with agricultural land use) that are listed in the correspondence table;
- the second largest CLC2000 class in Europe (see Table 8) is coniferous forest (312) with 83.3 % PTA value — this means that if we select a point in a polygon which has a CLC2000 code of 312, there is an 83.1 % chance of finding coniferous or mixed forest, non-built-up linear features (i.e. roads), other coniferous and mixed wooded land, shrubland with sparse tree cover or natural grassland under one of the following land-use classes: forestry, recreation, leisure, sport or unused, as shown by the correspondence table;
- the highest PTA value (91 %, after glaciers — 335 — which is weakly representative) characterises class 243 (agriculture with significant amount of natural vegetation). This high level of agreement is not a surprise as the correspondence table includes a rather broad list of LLC classes. With this 'complex' class it is not really possible to test by means of point sampling as its definition includes the spatial context. The non-correspondence means a definite mistake, but the correspondence is not necessarily an agreement. On this basis, all 'complex' classes (2.4.x) are removed from Table 11 to provide a modified PTA value, which is easier to interpret (Table 12);
- very high PTA value (> 85 %) was obtained for the CLC2000 classes 512 (lakes), 522 (estuaries) and 521 (coastal lagoons) — only considering classes with < 50 % relative error;
- for CLC classes with acceptable representativeness (< 50 % relative error), the two CLC classes with the lowest PTA value (below 50 %) are: 333 (sparse vegetation) and 131 (mineral extraction sites). The weak accuracy for

**Table 11** Automatic comparison of CLC2000 and LUCAS LU and LC for Europe

CLC code	SSU	LLC_OK	% LLC_OK	LLU_OK	% LLU_OK	SSU_OK	PTA	RE
111	129	106	82.2	98	76.0	94	72.9	36.0
112	2 738	2 405	87.8	2 376	86.8	2 125	77.6	8.4
121	424	337	79.5	297	70.0	267	63.0	23.9
122	41	30	73.2	22	53.7	20	48.8	85.8
123	8	6	75.0	5	62.5	5	62.5	97.0
124	49	42	85.7	34	69.4	33	67.3	57.7
131	90	58	64.4	49	54.4	42	46.7	49.6
132	12	8	66.7	1	8.3	1	8.3	318.9
133	45	16	35.6	5	11.1	5	11.1	217.5
141	52	41	78.8	25	48.1	22	42.3	76.5
142	222	206	92.8	128	57.7	125	56.3	36.7
211	21 388	16 397	76.7	18 558	86.8	15 997	74.8	2.9
212	826	492	59.6	709	85.8	444	53.8	18.1
213	104	86	82.7	93	89.4	80	76.9	35.7
221	904	576	63.7	829	91.7	543	60.1	16.7
222	574	350	61.0	524	91.3	340	59.2	22.2
223	1 196	856	71.6	1 107	92.6	835	69.8	14.1
231	8 616	6 671	77.4	7 652	88.8	6 381	74.1	4.9
241	271	178	65.7	182	67.2	155	57.2	31.6
242	5 655	4 651	82.2	4 511	79.8	4 329	76.6	6.0
243	3 937	3 847	97.7	3 643	92.5	3 583	91.0	6.6
244	875	793	90.6	824	94.2	772	88.2	13.9
311	8 805	7 279	82.7	7 822	88.8	6 652	75.5	4.8
312	15 931	13 747	86.3	15 174	95.2	13 268	83.3	3.3
313	6 441	5 597	86.9	6 052	94.0	5 354	83.1	5.4
321	2 777	1 764	63.5	2 465	88.8	1 690	60.9	9.3
322	2 709	1 960	72.4	2 305	85.1	1 742	64.3	9.8
323	2 380	1 542	64.8	1 875	78.8	1 349	56.7	10.1
324	5 727	4 364	76.2	5 023	87.7	3 947	68.9	6.3
331	75	38	50.7	48	64.0	25	33.3	72.3
332	564	416	73.8	492	87.2	394	69.9	18.9
333	957	509	53.2	725	75.8	453	47.3	17.6
334	27	23	85.2	22	81.5	24	88.9	67.7
335	41	41	100.0	41	100.0	41	100.0	52.7
411	181	130	71.8	147	81.2	111	61.3	38.4
412	2 152	1 845	85.7	1 841	85.5	1 722	80.0	9.8
421	83	68	81.9	58	69.9	50	60.2	54.3
422	22	16	72.7	16	72.7	16	72.7	103.1
511	188	107	56.9	165	87.8	107	56.9	36.7
512	2 646	2 405	90.9	2 539	96.0	2 384	90.1	8.2
521	47	41	87.2	47	100.0	41	87.2	40.2
522	27	24	88.9	26	96.3	24	88.9	42.8
<b>Total</b>	<b>99 936</b>	<b>80 067</b>	<b>80.1</b>	<b>88 626</b>	<b>88.6</b>	<b>75 598</b>	<b>75.6</b>	<b>0.7</b>

**Note:** CLC codes printed in light blue are non-representative, while grey means > 50 % error.

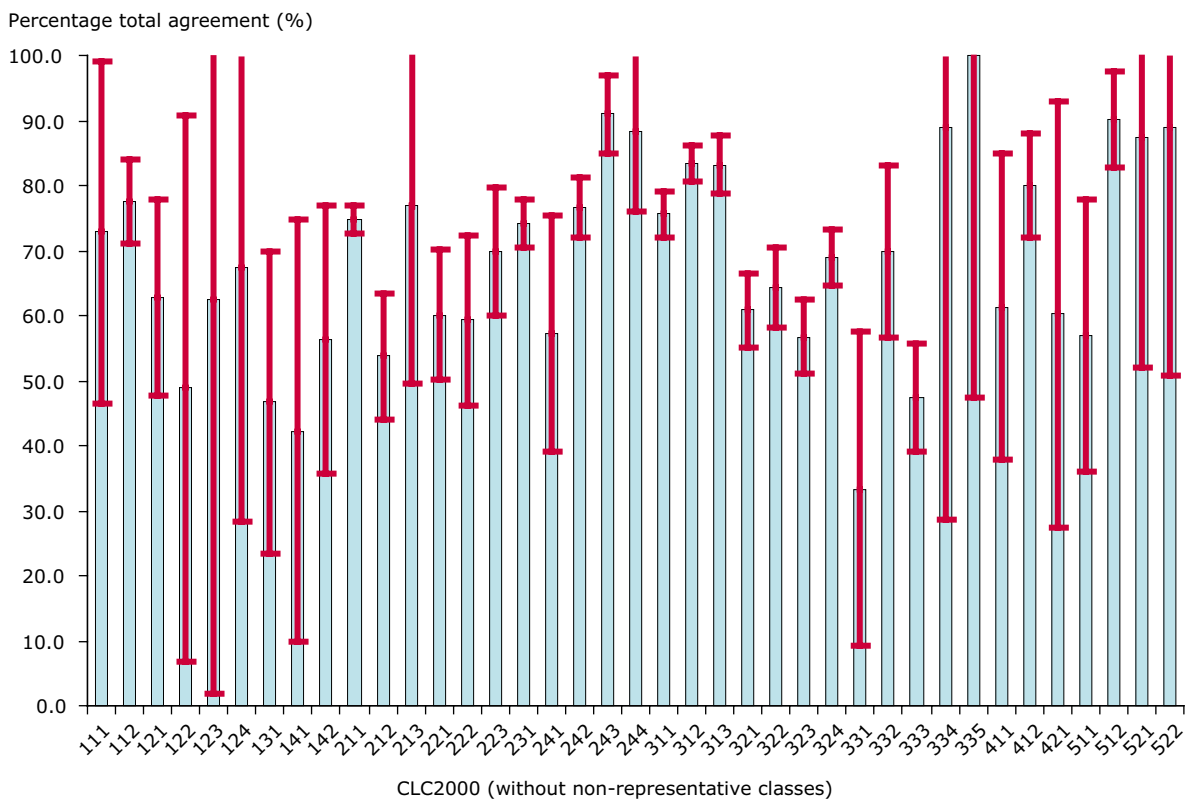


333 is explained by thematic confusion. This is confirmed also by the reinterpretation approach (see Section 5.2), while the explanation for 131 is their small characteristic size.

three samples are expected to be correctly classified by CLC2000. Considering the different sizes of the observation units (CLC2000: 25 ha versus LUCAS: circle of 3 m diameter), this result seems to be rather positive.

Table 12 means that CLC2000 approximates reality with  $74.8 \pm 0.6$  % total agreement (PTA) value. This means that out of four randomly selected locations,

**Figure 3 Automatic comparison of LUCAS and CLC2000: class-level PTA values (without non-representative CLC classes)**



**Note:** Absolute error ranges are displayed in red.

**Table 12 Summarised results of comparison for Europe without the 'complex' CLC classes**

PSU	LLC_OK	% LLC_OK	LLU_OK	% LU_OK	SSU_OK	PTA	AE
89 198	70 598	79.1	79 391	89.0	66 749	74.8	0.6

## 5 Results of the reinterpretation

Using the methodology introduced in Section 3.2, 8 231 PSUs distributed in 18 countries were analysed.

### 5.1 Reliability of CLC2000

According to Figure 4 and Table 13:

- 81.2 % of samples (6 681 cases) showed 'clear agreement';
- 12.8 % of samples (1 057 cases) were considered as 'wrong interpretation';
- 4.6 % of samples (377 cases) were labelled as 'another interpretation is possible', while;
- 1.4 % of samples (116 cases) were not evaluated ('not enough information').

Combined first and third cases ('clear agreement' and 'another interpretation is possible') yield the amount of samples with acceptable CLC2000 classification (7 058 cases). This represented 87 % of all samples used in the verification (8 115 cases). This means that the reliability of CLC2000 at 95 % confidence level is  $87.0 \pm 0.7$  %, according to the independent interpretation.

Based on Table 13 and Figure 5, the class level reliability (for classes with relative error better than 50 %) can be summarised as follows:

- CLC classes estimated with very high reliability (> 95 %) are:
  - Rivers (511)
  - Lakes (512)
  - Industrial and commercial units (121)
  - Discontinuous urban fabric (112).
- CLC classes estimated with high reliability (between 90 and 95 %):
  - Agro-forestry (244)
  - Irrigated arable land (212)
  - Arable land (211)
  - Coniferous forest (312).
- CLC classes estimated with good (around average) reliability (between 85 and 90 %):

- Mixed forest (313)
- Olive groves (223)
- Pastures (231)
- Deciduous forest (311).

- CLC classes estimated with moderate reliability (between 80 and 85 %):

- Vineyards (221)
- Peat bogs (412)
- Transitional woodland shrub (324).

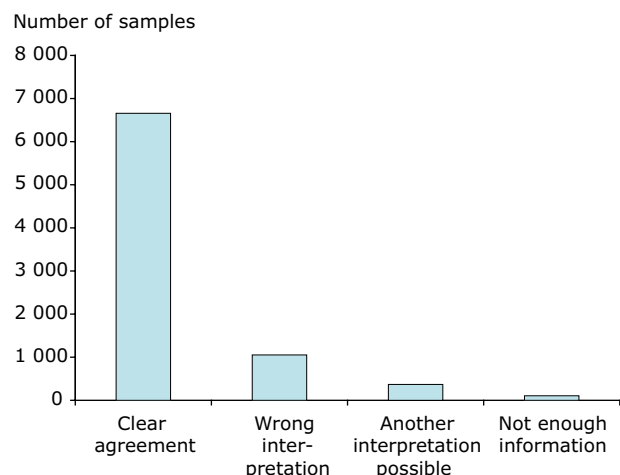
- CLC classes estimated with low reliability (between 70 and 80 %):

- Complex cultivation pattern (242)
- Sclerophyllous vegetation (323)
- Natural grassland (321)
- Primarily agriculture with significant amount of natural formations (243)
- Moors and heathland (322)
- Fruit trees and orchards (222).

- CLC class estimated with very low reliability (< 70 %):

- Sparse vegetation (333).

**Figure 4 Summarised results of the reinterpretation of CLC2000**

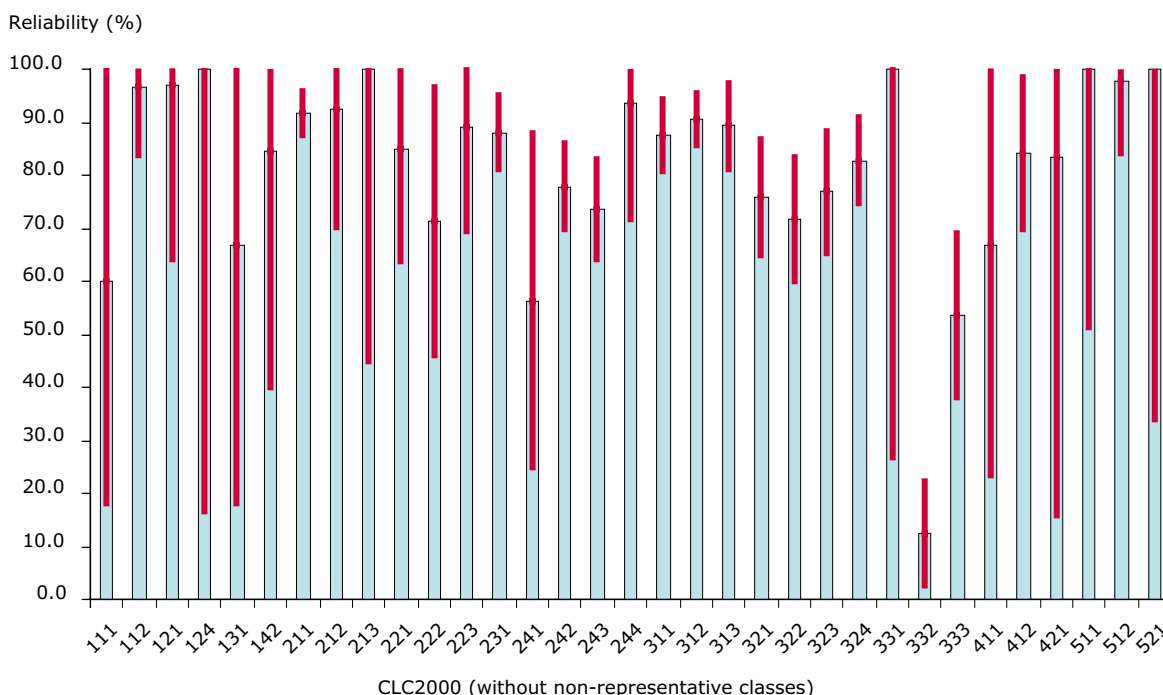


**Table 13** Class level results of reinterpretation of CLC2000

CLC code	PSU controlled	Clear agreement (a)	% (a)	Wrong (w)	% (w)	Another interpretation possible (p)	% (p)	Not enough info (n)	% (n)	Total points considered (a + w + p)	Percentage of points considered (a + w + p)/t	Percentage of points correct (a + p)/(a + w + p)	RE
111	10	6	60.0	4	40.0					10	100.0	60.0	69.7
112	257	239	93.0	9	3.5	7	2.7	2	0.8	255	99.2	96.5	13.2
121	37	31	83.8	1	2.7	2	5.4	3	8.1	34	91.9	97.1	33.8
122	1	1	100.							1	100.0	100.0	105.1
123	0												
124	2	2	100.							2	100.0	100.0	83.2
131	9	4	44.4	3	33.3	2	22.2			9	100.0	66.7	72.9
132	1	1	100.							1	100.0	100.0	161.6
133	5	1	20.0	3	60.0			1	20.0	4	80.0	25.0	254.5
141	2	1	50.0	1	50.0					2	100.0	50.0	123.6
142	27	21	77.8	4	14.8	1	3.7	1	3.7	26	96.3	84.6	52.5
211	2 039	1 828	89.7	168	8.2	33	1.6	10	0.5	2 029	99.5	91.7	4.5
212	82	66	80.5	6	7.3	7	8.5	3	3.7	79	96.3	92.4	24.2
213	8	8	100.							8	100.0	100.0	55.0
221	82	59	72.0	12	14.6	8	9.8	3	3.7	79	96.3	84.8	24.6
222	49	32	65.3	14	28.6	3	6.1			49	100.0	71.4	35.4
223	115	86	74.8	12	10.4	10	8.7	7	6.1	108	93.9	88.9	21.9
231	839	710	84.6	100	11.9	26	3.1	3	0.4	836	99.6	88.0	7.9
241	24	5	20.8	7	29.2	4	16.7	8	33.3	16	66.7	56.3	56.0
242	545	380	69.7	118	21.7	35	6.4	12	2.2	533	97.8	77.9	10.4
243	381	227	59.6	98	25.7	46	12.1	10	2.6	371	97.4	73.6	12.9
244	69	56	81.2	4	5.8	4	5.8	5	7.2	64	92.8	93.8	23.7
311	602	491	81.6	75	12.5	33	5.5	3	0.5	599	99.5	87.5	7.7
312	1 239	1 084	87.5	116	9.4	34	2.7	5	0.4	1 234	99.6	90.6	5.5
313	490	397	81.0	51	10.4	34	6.9	8	1.6	482	98.4	89.4	9.1
321	200	128	64.0	47	23.5	19	9.5	6	3.0	194	97.0	75.8	14.6
322	126	78	61.9	34	27.0	8	6.3	6	4.8	120	95.2	71.7	16.2
323	202	140	69.3	45	22.3	9	4.5	8	4.0	194	96.0	76.8	15.2
324	466	342	73.4	79	17.0	39	8.4	6	1.3	460	98.7	82.8	10.0
331	5	5	100.							5	100.0	100.0	73.3
332	8			7	87.5	1	12.5			8	100.0	12.5	78.4
333	29	11	37.9	13	44.8	4	13.8	1	3.4	28	96.6	53.6	29.1
334	4	3	75.0					1	25.0	3	75.0	100.0	112.1
335	0												
411	19	12	63.2	6	31.6			1	5.3	18	94.7	66.7	64.7
412	103	78	75.7	16	15.5	7	6.8	2	1.9	101	98.1	84.2	16.8
421	6	5	83.3	1	16.7					6	100.0	83.3	81.0
422	2	1	50.0					1	50.0	1	50.0	100.0	154.3
511	17	16	94.1			1	5.9			17	100.0	100.0	48.6
512	124	121	97.6	3	2.4					124	100.0	97.6	13.8
521	3	3	100.							3	100.0	100.0	66.0
522	0												
<b>Total</b>	<b>8 231</b>	<b>6 681</b>	<b>81.2</b>	<b>1.057</b>	<b>12.8</b>	<b>377</b>	<b>4.6</b>	<b>116</b>	<b>1.4</b>	<b>8 115</b>	<b>98.6</b>	<b>87.0</b>	<b>0.9</b>

**Note:** CLC codes printed in light blue are non-representative, grey means > 50 % relative error, while red means the absence of LUCAS field photographs.

**Figure 5** Reinterpretation of Image2000 with LUCAS: class-level accuracies (without non-representative CLC classes)



## 5.2 Analysis of errors

### 5.2.1 Reasons for errors in CLC2000

Cases classified as 'wrong interpretation' are further analysed. Using a set of standard explanations (see Section 3.2), the 1 057 cases are labelled as follows:

- 472 samples (44.7 %): wrong code;
- 435 samples (41.2 %): not detailed interpretation;
- 134 samples (12.8 %): wrong generalisation;
- 15 samples (1.8 %): inaccurate delineation (if the positional error in delineation exceeds 100 m).

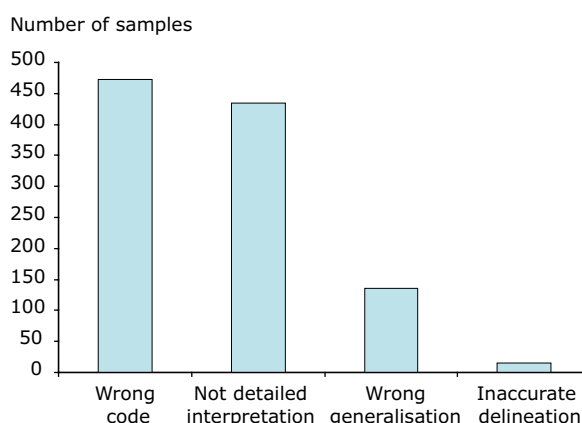
According to Figure 6, 'wrong code' (i.e. commission errors) and 'not detailed interpretation' (i.e. omission errors) are the two main reasons for mistakes that have a similar number of occurrences. Drawing accuracy (i.e. matching of Image2000 and CLC2000) is the least important source of error. This reflects the many efforts taken by national teams to improve the geometric accuracy of CLC90. Several examples of different cases of 'wrong interpretation' are shown in Annex 9.

### 5.2.2 Misclassifications in CLC2000

An error matrix has been constructed using the second validation code (cod-val2) and the CLC2000 database code (act-cod) for all cases of 'wrong interpretation'. The detailed error matrix is presented in Annex 7. Level 1 summary of the error matrix is shown in Table 14.

- Elements of the error matrix (Annex 7) above the diagonal means commission errors, that is, cases where the area around the LUCAS PSU belongs to class 'j' but is classified as class 'i'.

**Figure 6** The reasons for cases with 'wrong interpretation'



- Elements of the error matrix below the diagonal means omission errors, that is, cases where the area around LUCAS PSU is really class 'i' but is classified as a member of class 'j'.
  - Diagonal elements of the error matrix are expected to include zero values.
- The actual distribution of samples is as follows:
- commission errors: 594 cases (56.2 %);
  - omission errors: 457 cases (43.2 %);
  - diagonal elements (special non-agreements): six cases (0.6 %).

**Table 14 Error matrix aggregated to level 1**

	Artificial surfaces	Agriculture	Forests and semi-natural areas	Wetlands	Water	Total
Artificial surfaces	10	12	1	1	0	<b>24</b>
Agriculture	7	448	80	3	0	<b>538</b>
Forests and semi-natural areas	4	80	365	11	3	<b>463</b>
Wetlands		2	20	1	0	<b>23</b>
Water	0	1	1	1	0	<b>3</b>
<b>Total</b>	<b>21</b>	<b>543</b>	<b>467</b>	<b>17</b>	<b>3</b>	<b>1 051</b>

**Table 15 Most frequent misclassification types in CLC2000**

Database code — validation code	Number of cases	Explanation
211-231	69	Pasture misclassified as arable land
312-313	60	Mixed forest misclassified as coniferous forest
211-242	46	Complex cultivation patterns misclassified as arable land
231-242	44	Complex cultivation patterns misclassified as pasture
242-231	41	Pasture misclassified as complex cultivation patterns
242-211	40	Arable land misclassified as complex cultivation patterns
311-313	32	Mixed forest misclassified as broadleaved forest
231-211	29	Arable land misclassified as pasture
243-231	26	Pasture misclassified as land principally occupied by agriculture with significant areas of natural vegetation
324-313	26	Mixed forest misclassified as 'transitional woodland shrub'
312-324	26	Transitional woodland shrub misclassified as coniferous forest
313-312	24	Coniferous forest misclassified as mixed forest
All others	< 20	

**Table 16 Most frequent level 1 misclassification types in CLC2000**

Database code — validation code	Number of cases	Explanation
242-324	13	Transitional woodland shrub misclassified as complex cultivation patterns
323-243	9	Principally occupied by agriculture with significant areas of natural vegetation misclassified as sclerophyllous vegetation
211-321	9	Natural grassland misclassified as arable land
243-311	8	Broadleaved forest misclassified as principally occupied by agriculture with significant areas of natural vegetation
243-323	7	Sclerophyllous vegetation misclassified as principally occupied by agriculture with significant areas of natural vegetation
321-231	7	Pasture misclassified as natural grassland
322-231	7	Pasture misclassified as moors and heathland
324-243	6	Principally occupied by agriculture with significant areas of natural vegetation misclassified as transitional woodland shrub
324-412	6	Peatland misclassified as transitional woodland shrub
311-243	6	Principally occupied by agriculture with significant areas of natural vegetation misclassified as broadleaved forest
All others	< 5	

The special cases mean that although the CLC code was agreed, the extent of the polygon (as confirmed by the other SSUs) was incorrect (see Illustration 30 in Annex 9).

The fact that there were a larger number of commission errors than omission errors confirms that there are more problems with improper codes than with insufficient details (see Figure 6).

Tables 14–16 show the following.

- Seventy-eight per cent of misclassification cases occur at levels 2 and 3. All the most frequent misclassification types listed in Table 14 are level 3 or level 2 mistakes.
- The most frequent mistakes occur in the 'Agricultural' and the 'Forest and semi-natural' groups (77 % of all errors).
- The most frequent level 1 mistakes (Table 16) occur mostly between classes 2 and 3 (15 % of all errors).

### 5.3 Subjectivity of photo interpretation

It is widely recognised that photo interpretation always has some level of subjectivity. In the course of validating CLC2000 with LUCAS, this subjectivity was characterised by the following cases.

- All cases of 'clear agreement', when val-cod1 was not equal to act-cod, or val-cod2 was not equal to act-cod. This means that although one of the codes given by the validating expert was different from the actual CLC2000 code, the decision was 'clear agreement'.
- All cases of 'another interpretation is possible' except 'LC boundary'. Although the codes given by the validating expert were different from the actual CLC2000 code, the interpretation was still acceptable.

By considering the above cases, a 'subjectivity matrix' has been prepared (Annex 8) as follows:

- In the case of 'clear agreement': if val-cod1 was not equal to act-cod, val-cod1 and act-code were input to the matrix; if val-cod2 was not equal to

act-cod, val-cod2 and act-code were input to the matrix.

- In the case of 'another interpretation is possible': val-cod2 and act-code were input to the matrix.

In total, 1 501 cases became subjective. This represents 18.2 % of the total number of valid samples. By comparing the number of subjective cases in a class with the total number of cases of this class, a subjectivity of the class can be calculated. The analysis of the subjectivity matrix (Annex 8) reveals the following.

- The most subjective classes (in descending order; only classes which were validated with less than 50 % relative error are considered):
  - agriculture with significant amount of natural vegetation (243, subjectivity index: 42.3 %, i.e. 157 of the total of 371 occurrences were subjective). Most frequent intermixing classes are (in order of importance): 242, 231, 211, 311, 323, 313, 324;
  - transitional woodland shrub (324, subjectivity index: 36.1 %). Most frequent intermixing classes are (in order of importance): 312, 313, 311, 323;
  - complex cultivation patterns (242, subjectivity index: 34 %). Most frequent intermixing classes are (in order of importance): 211, 243, 231;
  - mixed forest (313, subjectivity index: 33.4 %). Most frequent intermixing classes are (in order of importance): 312, 311, 324;
  - permanently irrigated land (212, subjectivity index: 30.4 %). Most frequent intermixing class: 211.
- The least subjective classes (subjectivity is below 15 %):
  - Water bodies (512, subjectivity index: 4.8 %)
  - Arable land (211, 8.4 %)
  - Agro-forestry (244, 9.4 %)
  - Rivers (511, 11.8 %)
  - Pastures (231, 12.1 %)
  - Olives (223, 13.0 %)
  - Coniferous forest (312, 13.2 %)
  - Peatland (412, 13.9 %).

## 6 Remarks on LUCAS data quality

---

In general, LUCAS data supported the validation process very well. The lack of photos in a large number of PSUs (not only in mountains) made validation less representative. Geometrical match with Image2000 was usually rather good (within one Landsat ETM pixel, i.e. 25 m, see Illustrations 34 and 35 in Annex 9). However, some inconsistencies were found which could be interesting to note. The following remarks might be useful for the preparation of the new LUCAS survey in 2006.

- In some cases, Image2000 and LUCAS field photographs do not match (see Illustrations 36 and 37 in Annex 9).
- Bad photo quality was the most important problem (see Illustration 38, Annex 9).
- Sometimes photos refer completely to another site, while LLC and LLU codes fit to Image2000 (see Illustration 39).
- In a few cases, the same photo is attached to two different view directions (see Illustration 40).
- In many cases, a date appears on the photo, but the date in the camera has not been set. Therefore, it is invalid.
- Sometimes, LUCAS LC codes are definitely bad, as confirmed by Image2000 (see Illustrations 41, 43 and 44).
- In many cases, LUCAS LU codes are difficult to understand (see Illustration 42).
- The LUCAS LC class name: Poplar, Eucalypt (C30) is misleading, as in many cases (e.g. Portugal) this category included coniferous plantations.

## 7 Conclusions

CLC2000 data covering 18 countries of Europe (3.4 million km<sup>2</sup>) were validated by means of LUCAS data. Two kinds of method were applied:

- automatic comparison of CLC2000 codes and LUCAS LU and LC codes from more than 100 000 SSUs;
- reinterpretation of Image2000 data from more than 8 200 LUCAS PSUs based on ground photographs and LUCAS LU and LC codes.

Accuracy/reliability figures for the total amount of samples and also at class level were determined. Representativeness (error values) of the estimates were also derived using statistical principles as a basis.

Main results of the reinterpretation approach were as follows:

- The total reliability of CLC2000 is  $87.0 \pm 0.8$  %. This value is based on an independent CLC interpretation performed on LUCAS PSUs. Therefore, it can be concluded that the 85 % accuracy requirement specified in the technical guidelines (EEA-ETC/TE, 2002) was fulfilled.
- Looking at the class-level values, 22 of the 44 CLC classes could be validated.
- The other 22 CLC classes could not be validated because of low representativeness (or intentional omissions by LUCAS, e.g. CLC class glaciers or sea and ocean). Classes which could not be validated belong mostly to artificial surfaces, wetlands and water.
- The highest class-level reliability (> 95 %) was obtained for rivers (511), lakes (512), industrial and commercial units (121) and discontinuous urban fabric (112).
- The two largest CLC classes in the 18 countries considered were arable land (211) and coniferous forest (312). These were estimated to have a high level of reliability (percentage of correct values between 90 and 95). Two other agricultural classes also enjoyed a high level of

reliability: agro-forestry (244) and permanently irrigated land (212).

- The lowest class-level reliability (below 70 %) was obtained for sparse vegetation (333) class. This highlights the difficulties in interpreting that category.
- Main sources of mistakes included using the wrong code (commission errors) and not interpreting enough details (omission errors). Geometrical inaccuracies played a less important role.
- The majority (78 %) of classification errors occurred at levels 3 and 2.
- Level 1 misclassifications mostly occur between 'agriculture' and 'forest and semi-natural' classes.
- The analysis revealed that subjectivity of photo interpretation could be noticed in 18.2 % of the samples. The most subjective CLC classes are, as follows: agriculture with significant amount of natural vegetation (243), transitional woodland, shrub (324), complex cultivation patterns (242) and mixed forest (313).

The main results of the automatic comparison are as follows:

- Percentage of total agreement (PTA) between CLC2000 and LUCAS LU and LC is  $74.8 \pm 0.6$  %. This value can be interpreted as follows: CLC2000 approximates LUCAS thematic data with a 74.8 % average accuracy (i.e. out of four randomly selected locations three samples are expected to be correctly classified by CLC2000). Considering the different sizes of the observation units (25 ha versus circle of 3 m diameter) these results should be considered to be rather positive.
- Looking at the class-level values, 13 of the 44 CLC classes could not be tested because of low representativeness. By removing as well from the analysis the 2.4.x classes (complex agriculture) (as having no meaning at the SSU level), a total of 27 classes were validated.



- 
- Highest PTA (> 85 %) values were obtained for water classes: lakes (512), estuaries (522) and coastal lagoons (521).
  - Lowest PTA values (< 50 %) were obtained for sparse vegetation (333) and mineral extraction sites (131). PTA for class 333 definitely refers to classification mistakes (as shown also by the reinterpretation), while the case of 131 might be explained by the small characteristic size of the representatives of this class (result of generalisation).
  - Positive examples of simultaneous display of Image2000, CLC2000 and LUCAS data could be used (if Eurostat/LUCAS copyright would allow) in CLC2000-related training/demonstration activities.
  - The new European LUCAS survey in 2006 could benefit also from the results, for instance, by improving consistency of LU/LC codification and by producing better field photographs.

Some potential applications of the results are described below.

- In case of the next update of CLC, special attention should be paid to the less accurate classes. This could refer to the improvement of the definition of mapping rules and the use of multi-temporal satellite data during interpretation.

Another type of application is the decomposition of CLC mixed classes (e.g. 242, 243) into pure land cover classes based on LUCAS LC statistics. These are derived from 'clear agreement' cases of the given CLC class. Similarly, all classes could be decomposed to LUCAS LC classes (to account for the generalisation in CLC). This way, the CLC-based statistics would be more easily comparable with national statistics (at present this comparison is not possible for two reasons: mixed classes and generalisation).

# References

- APAT: The project Image and Corine land cover 2000 in Italy, April 2005.
- Avikainen, J., Delincé, J., Croi, W., Kayadjanian, W., Bettio, M., Mariano, A., 2003. LUCAS land use/Cover area frame statistical survey. Technical document, No 1: Sampling plan (Version 2.4). EUROSTAT/LAND/LUCAS1
- Bossard, M., Feranec, J., Otahel, J., 2000. 'Corine land cover technical guide — Addendum 2000'. Technical report, No 40. Copenhagen (EEA). <http://terrestrial.eionet.eu.int>
- Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G. and Soukup, T., 2004. 'The Corine land cover 2000 project'. EARSel eProceedings 3(3), pp. 331–346.
- Caetano, M., Mata, F., 2005. 'Accuracy assessment of CLC2000 for Portugal'. 25th EARSel Symposium, Porto, 6–9 June 2005 (to be published in the proceedings).
- Centre for Ecology and Hydrology (NERC), 2005. 'Corine land cover 2000: semi-automated updating of Corine land cover in the UK', Final report, May 2005.
- Duhamel, C., Eiden, G., Aifantopoulou, D., Croi, W., 2003. LUCAS land use/Cover area frame statistical survey. Technical document, No 2: The nomenclature (Version 1.5), EUROSTAT/LAND/LUCAS2
- EEA-ETC/TE, 2002. Corine land cover update, Image 2000 and CLC2000 project, Technical guidelines, <http://terrestrial.eionet.eu.int>
- EPA, 2003. 'Corine land cover 2000 update (Ireland)', Final report, Environment Protection Agency, Ireland.
- Eurostat, 2005. 'The LUCAS project (Land use/Land cover area frame survey)', Call for tender, 22 April 2005.
- Finnish Environment Institute, 2003. 'Corine land cover 2000 and Image 2000 in Finland', Interim report, Helsinki, Finland. Final report.
- Hazeau, G. W., 2003. CLC2000. Land cover database of the Netherlands, Alterra Rapport, Wageningen, 775 ISSN 1566–7197.
- Heymann, Y., Steenmans, C., Croissille, G., Bossard, M., 1994. 'Corine land cover'. Technical guide. EUR 12585, Luxembourg, Office for Official Publications of the European Communities.
- Kayadjanian, M., 2001. LUCAS land use/Cover area frame statistical survey. Technical document, No 6: Data transfer and control procedures (Version 2.2), EUROSTAT/LAND/LUCAS6
- Lantmäteriet: Swedish CLC2000, Final report, March 2005.
- Maucha, G., Büttner, G., Kosztra, B., 2003. 'Applying LUCAS data for verification/validation of CLC2000'. A pilot study for Hungary (ETC-TE internal report).

# Annexes

---

- 1 Corine land cover nomenclature**
- 2 LUCAS land cover nomenclature**
- 3 LUCAS land use nomenclature**
- 4 Results of cross-tabulating CLC2000 and LUCAS LC**
- 5 Results of cross-tabulating CLC2000 and LUCAS LU**
- 6 Correspondence table between CLC classes and LUCAS land cover and land use classes**
- 7 Reclassification of Image2000 using LUCAS: error matrix**
- 8 Reclassification of Image2000 using LUCAS: subjectivity matrix**
- 9 Illustrations (screen shots)**

# Annex 1 Corine land cover nomenclature

Level 1	Level 2	Level 3	
1. Artificial surfaces	1.1 Urban fabric	1.1.1 Continuous urban fabric	
		1.1.2 Discontinuous urban fabric	
	1.2 Industrial, commercial and transport units	1.2.1 Industrial or commercial units	
		1.2.2 Road and rail networks and associated land	
		1.2.3 Port areas	
		1.2.4 Airports	
	1.3 Mine, dump and construction sites	1.3.1 Mineral extraction sites	
		1.3.2 Dump sites	
		1.3.3 Construction sites	
	1.4 Artificial, non-agricultural vegetated areas	1.4.1 Green urban areas	
		1.4.2 Sport and leisure facilities	
	2. Agricultural areas	2.1 Arable land	2.1.1 Non-irrigated arable land
			2.1.2 Permanently irrigated land
2.1.3 Rice fields			
2.2 Permanent crops		2.2.1 Vineyards	
		2.2.2 Fruit trees and berry plantations	
		2.2.3 Olive groves	
2.3 Pastures		2.3.1 Pastures	
2.4 Heterogeneous agricultural areas		2.4.1 Annual crops associated with permanent crops	
		2.4.2 Complex cultivation patterns	
		2.4.3 Land principally occupied by agriculture, with significant areas of natural vegetation	
		2.4.4 Agro-forestry areas	
3. Forests and semi-natural areas		3.1 Forests	3.1.1 Broadleaved forest
			3.1.2 Coniferous forest
	3.1.3 Mixed forest		
	3.2 Scrub and/or herbaceous vegetation associations	3.2.1 Natural grasslands	
		3.2.2 Moors and heathland	
		3.2.3 Sclerophyllous vegetation	
		3.2.4 Transitional woodland shrub	
	3.3 Open spaces with little or no vegetation	3.3.1 Beaches, dunes, sands	
		3.3.2 Bare rocks	
		3.3.3 Sparsely vegetated areas	
		3.3.4 Burnt areas	
		3.3.5 Glaciers and perpetual snow	
	4. Wetlands	4.1 Inland wetlands	4.1.1 Inland marshes
4.1.2 Peat bogs			
4.2 Maritime wetlands		4.2.1 Salt marshes	
		4.2.2 Salines	
		4.2.3 Intertidal flats	
		4.2.4 Intertidal flats	
5. Water bodies	5.1 Inland waters	5.1.1 Water courses	
		5.1.2 Water bodies	
	5.2 Marine waters	5.2.1 Coastal lagoons	
		5.2.2 Estuaries	
		5.2.3 Sea and ocean	
		5.2.3 Sea and ocean	

## Annex 2 LUCAS land cover nomenclature

<b>CODE</b>	<b>Meaning</b>	<b>CODE</b>	<b>Meaning</b>
A11	Buildings with one to three floors	B71	Apple fruit
A12	Buildings with more than three floors	B72	Pear fruit
A13	Greenhouses	B73	Cherry fruit
A21	Non built-up area features	B74	Nuts trees
A22	Non built-up linear features	B75	Other fruit trees and berries
B11	Common wheat	B76	Oranges
B12	Durum wheat	B77	Other citrus fruit
B13	Barley	B81	Olive groves
B14	Rye	B82	Vineyards
B15	Oats	B83	Nurseries
B16	Maize	B84	Permanent industrial crops
B17	Rice	C11	Broadleaved forest
B18	Other cereals	C12	Coniferous forest
B21	Potatoes	C13	Mixed forest
B22	Sugar beet	C21	Other broadleaved wooded area
B23	Other root crops	C22	Other coniferous wooded land
B31	Sunflower	C23	Other mixed wooded land
B32	Rape seeds	C30	Poplars, eucalyptus
B33	Soya	D01	Shrubland with sparse tree cover
B34	Cotton	D02	Shrubland without tree cover
B35	Other fibre and oleaginous crops	E01	Permanent grassland with sparse tree/shrub cover
B36	Tobacco	E02	Permanent grassland without tree/shrub cover
B37	Other non-permanent industrial crops	F00	Bare land
B41	Dry pulses	G01	Inland water bodies
B42	Tomatoes	G02	Inland running water
B43	Other fresh vegetables	G03	Coastal water bodies
B44	Floriculture and ornamental plants	G04	Wetland
B50	Temporary, artificial pastures	G05	Glaciers, permanent snow
B60	Fallow land		

## Annex 3 LUCAS land use nomenclature

---

<b>CODE</b>	<b>Meaning</b>
U11	Agriculture
U12	Forestry
U13	Fishing
U14	Mining and quarrying
U21	Energy production
U22	Industry and manufacturing
U31	Transport, communication, storage, protective works
U32	Water and waste treatment
U33	Construction
U34	Commerce, finance, business
U35	Community services
U36	Recreation, leisure, sport
U37	Residential
U40	Unused

# Annex 4 Results of cross-tabulating CLC2000 and LUCAS LC

%	CLC2000																																																				All						
LUCAS LC	111	112	121	122	123	124	131	132	133	141	142	211	212	213	221	222	223	231	241	242	243	244	311	312	313	321	322	323	324	331	332	333	334	335	411	412	421	422	423	511	512	521	522	523	All														
A11	34.1	22.5	18.9	7.3					8.9	5.8	2.7	0.9	1.8	1.2	1.6	0.8	1.0	3.3	2.2	1.2	0.3	0.3	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.2												0.2					2.9	1.3											
A12	24.0	2.3	7.5	12.5					2.2			0.0				0.1	0.0		0.1	0.1																									0.5	2.1	1.5	0.2											
A13		0.2	0.5									0.1	1.2			0.1			0.1							0.0																							0.0										
A21	12.4	12.1	25.2	12.2	25.0	2.0	13.3		8.9	3.8	8.1	0.8	1.6	0.7	0.5	0.5	1.0	0.4	1.5	0.7	0.1	0.2	0.3	0.2	0.1	0.1	0.2	0.2		0.2											0.4			0.0	2.1	2.9	1.0												
A22	11.6	12.1	13.2	36.6	25.0	22.4	5.6	8.3	4.4	11.5	7.2	2.1	4.1	1.9	3.0	2.3	1.8	2.0	5.2	2.9	2.6	1.1	1.1	1.3	1.3	1.0	0.4	1.1	1.2																1.6	0.1		2.0											
B11		0.5	0.2						2.2	0.5	17.7	3.9	0.9	1.0	0.7	1.6	1.1	4.6	2.3	1.4	0.4	0.0	0.0	0.1																										4.4									
B12		0.1	0.5	2.4								3.9	5.3	3.8	1.9	0.7	2.3	0.2	8.1	1.6	1.4	2.2	0.1	0.0	0.0	0.3	0.0	0.5	0.1	1.3																				1.2									
B13		0.3	0.2									12.5	8.0	2.2	0.7	0.7	1.6	0.4	4.2	3.1			0.2	0.1	0.2	0.3		0.4	0.2																						3.3								
B14		0.0	0.2									0.5	1.9			0.2	0.5		0.1	0.4	0.5	0.6	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.0																					0.5								
B15		0.3										1.9	2.1	1.0	0.1		0.6	0.2	1.5	0.9	1.3	1.7	0.0	0.1	0.1	0.3		0.1	0.1																							0.7							
B16		0.9	0.9				1.1		1.9	9.1	17.3	7.7	0.7	3.1	0.1	2.5	5.5	6.2	1.6				0.3	0.0	0.1																												2.8						
B17				2.4								0.1	0.8	70.2																																							0.1						
B18		0.1										1.0	1.1				0.3		0.5	0.7	0.8	0.5	0.1	0.1	0.0	0.0																											0.3						
B21		0.1										1.2	1.7			0.2	0.2	0.1	0.4	0.7	0.3			0.0	0.0	0.0																											0.3						
B22		0.1	0.2									2.0	2.1			0.3		0.1		0.4	0.0	0.1	0.0	0.0																														0.5					
B23		0.0										0.2	0.2					0.1		0.1	0.0																																	0.1					
B31		0.0	0.2				6.1					2.2	3.4		0.9	0.3	0.4	0.0	0.4	1.0	0.3			0.0	0.0	0.0	0.0																										0.6						
B32		0.3										0.5	3.7			0.2			0.2		0.4	0.4			0.1																													0.9					
B33												0.4	0.2						0.0		0.1	0.1																																	0.1				
B34												0.3	12.0								0.1																																	0.2					
B35													0.3						0.0	0.4	0.1				0.0																														0.1				
B36												0.2	0.5					0.1			0.1	0.0																																		0.0			
B37		0.0	0.2									0.2		0.2						0.2																																			0.1				
B41		0.1										1.7	1.0		0.3		0.3	0.2		0.3	0.3	0.1	0.0	0.0																															0.4				
B42												0.1	0.8							0.1																																				0.0			
B43		1.7	0.5						1.9	0.8	2.1	0.6	1.6	0.3	0.2	1.1	1.2	0.3					0.0	0.0	0.0																															0.4			
B44	0.8	1.0	0.7						1.9			0.1								0.1		0.0	0.1																																	0.1			
B50	0.8	1.4	0.5					8.3	1.9	0.9	6.6	5.8	1.7	1.4	1.0	5.6	5.5	7.4	7.3	2.9	0.6	0.5	0.7	0.7	0.2	0.5	0.5		0.2																												3.0		
B60	1.6	0.7	0.7			2.0	1.1	8.9			6.2	8.4	5.8	4.3	3.5	3.0	1.0	6.3	4.3	5.0	6.1	0.4	0.1	0.2	1.0	0.1	1.8	0.5																													2.3		
B71		0.6										0.2	0.2		0.2	11.0	0.3	0.2	0.4	1.0	0.5			0.1	0.0	0.0																															0.2		
B72		0.1										0.1	0.2				2.6		0.0	0.4	0.3	0.1																																				0.1	
B73		0.3	0.2									0.1				0.4	1.0	0.2	0.0	0.4	0.2	0.3			0.0	0.0	0.0																															0.1	
B74		0.1										0.2	0.1			0.8	9.6	0.8	0.0	2.6	1.5	0.3			0.0		0.0	0.2																														0.3	
B75		0.7										0.3	1.7			3.1	13.6	2.4	0.1	1.1	2.0	0.3			0.1	0.0																																0.4	
B76		0.0	0.2	2.4								0.0	1.0			13.1	0.2		1.8	0.3																																						0.1	
B77	0.8	0.1										0.0	0.1			0.2	3.8	0.2		0.4	0.1	0.1																																				0.0	
B81		0.5	0.2						2.2			0.4	0.4			2.4	2.6	65.1	0.0	12.2	4.0	2.9	1.0	0.2	0.1	0.1	0.2																																1.4
B82		0.4	0.5						4.4			0.6	2.1			54.5	2.1	1.8	0.1	6.3	4.3	1.3	0.1	0.2	0.1	0.1	0.1	0.0	0.2	0.1																													1.1
B83		0.1										0.1	0.1								0.0				0.0	0.0	0.0																																0.0
B84													0.0								0.1	0.1			0.0																																		0.0
C11		2.4	1.7					6.7	4.4	17.3	4.5	2.1	0.5		1.2	2.3	1.8	3.3	2.6	3.2	11.7	41.7	56.5	4.1	14.1	4.2	3.8	10.5	9.9	2.7	0																												





# Annex 6 Correspondence table between CLC classes and LUCAS land cover and land use classes

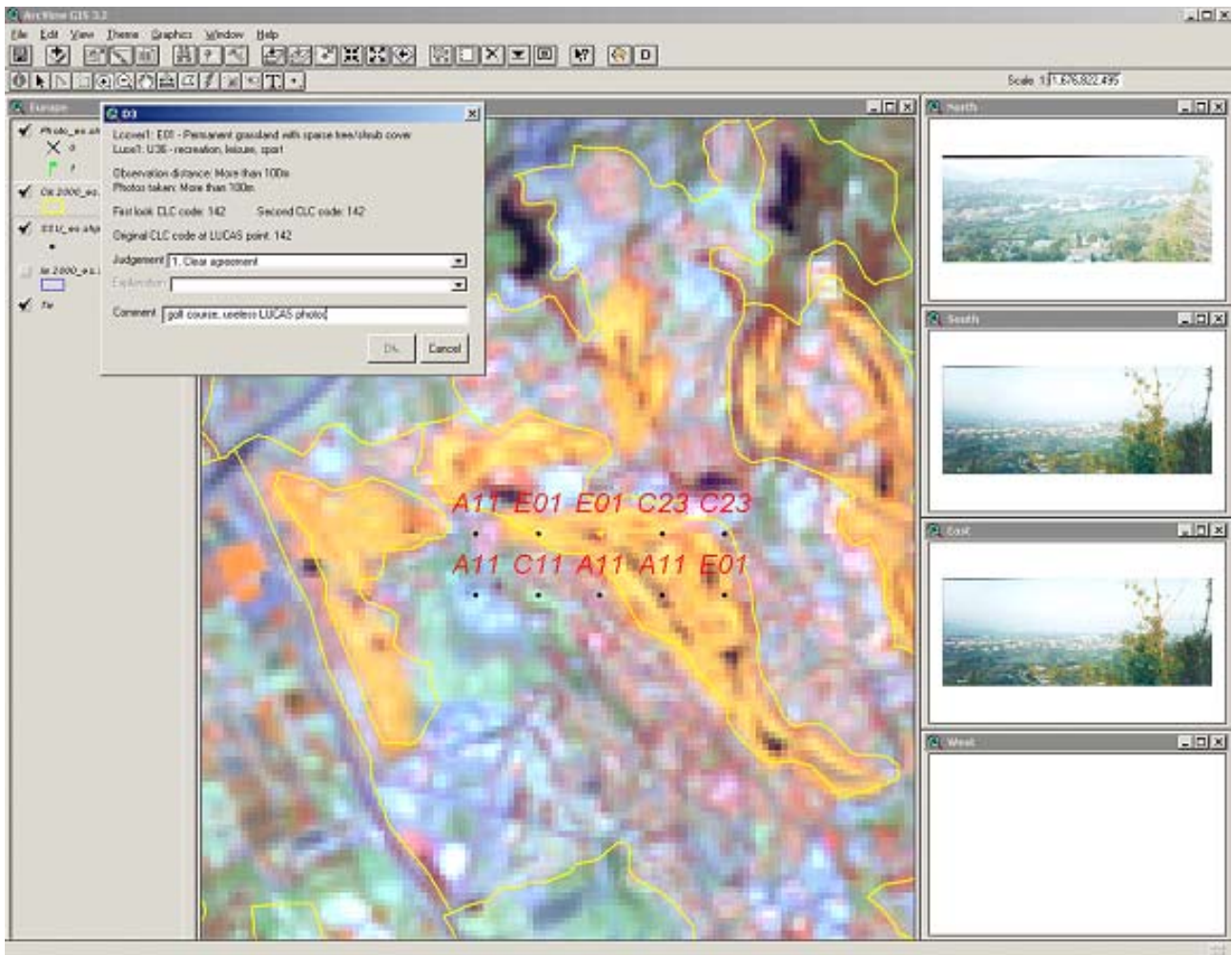
CLC code	CLC category	LUCAS land cover	LUCAS land use	OTHER
111	Continuous urban fabric	A12 A11 A2	U37 U35 U34 U31	
112	Discontinuous urban fabric	A11 A12 A2 B43 B44 C1 C2 E	U35 U37 U34 U31 U11 U36	
121	Industrial or commercial units	A12 A11 A2 E	U22 U21 U34 U35 U31 U40	
122	Road and rail network and associated land	A2 E	U31	
123	Port areas	A2 G01 G02 G03	U31 U36	
124	Airports	A2 E02	U31	
131	Mineral extraction sites	A2 F00 G01	U14 U40	
132	Dump sites	A2 F00	U32	
133	Construction sites	A11 A12 A2 F00	U33 U31	
141	Green urban areas	A21 A22 C1 C2 E G01	U36 U31	
142	Sport and leisure facilities	A11 A2 C1 C2 E G01	U36 U31	
211	Non-irrigated arable land	A13 A22 B11 B12 B13 B14 B15 B16 B18 B2 B3 B4 B50 B60 B83	U11	
212	Permanently irrigated land	A13 A22 B12 B16 B2 B31 B33 B34 B35 B36 B37 B4 B50 G02	U11	
213	Rice fields	B17 B60 A22 G02	U11	
221	Vineyards	B82 E02 A22 A11	U11 U40	
222	Fruit trees and berry plantations	B7 B84 A22 E02	U11 U40	
223	Olive groves	B81 A22 E	U11 U40	
231	Pastures	B50 B60 A22 E G04	U11 U40	
241	Annual crops associated with permanent crops	A11 A22 B11 B12 B13 B14 B15 B16 B18 B2 B3 B4 B50 B60 B7 B81 B82	U11	
242	Complex cultivation patterns	A11 A13 A22 B11 B12 B13 B14 B15 B16 B18 B2 B3 B4 B50 B60 B7 B81 B82 E	U11 U37	
243	Land principally occupied by agriculture, with significant areas of natural vegetation	A11 A13 A22 B11 B12 B13 B14 B15 B16 B18 B2 B3 B4 B50 B60 B7 B81 B82 C D E G01 G02 G04	U11 U12 U40	
244	Agro-forestry areas	A22 B11 B12 B13 B14 B15 B16 B18 B50 B60 C11 C21 E	U11 U12	
311	Broad-leaved forest	A22 C11 C30 C21 C13 C23 D01 E	U12 U36 U40	
312	Coniferous forest	A22 C12 C22 C13 C23 D01 E	U12 U36 U40	
313	Mixed forest	A22 C1 C2 D01 E	U12 U36 U40	
321	Natural grasslands	E D02	U40 U11	
322	Moors and heathland	D E	U40 U11	
323	Sclerophyllous vegetation	D E	U11 U40	
324	Transitional woodland shrub	B83 C1 C2 C30 D01 E	U12 U40	
331	Beaches, dunes and sand plains	F00	U40 U36	
332	Bare rocks	F00	U40	
333	Sparsely vegetated areas	F00 E	U40	
334	Burnt areas	C D	U12 U40	NATHAZARD = 1
335	Glaciers and perpetual snow	G05 F00	U40	
411	Inland marshes	G04 G01 E	U40 U11	
412	Peat bogs	G04 C1 C2 D	U14 U40 U12	
421	Salt marshes	G04 G03 E	U40	
422	Salines	G03	U14	
423	Intertidal flats	F00 G04	U40	
511	Water courses	G02	U31 U40 U13 U36	
512	Water bodies	G01	U13 U31 U40 U36 U21	
521	Coastal lagoons	G03 G04	U40 U31 U13 U36	
522	Estuaries	G03 G02	U40 U31 U13	
523	Sea and ocean	G03	U40 U13 U36 U31	-





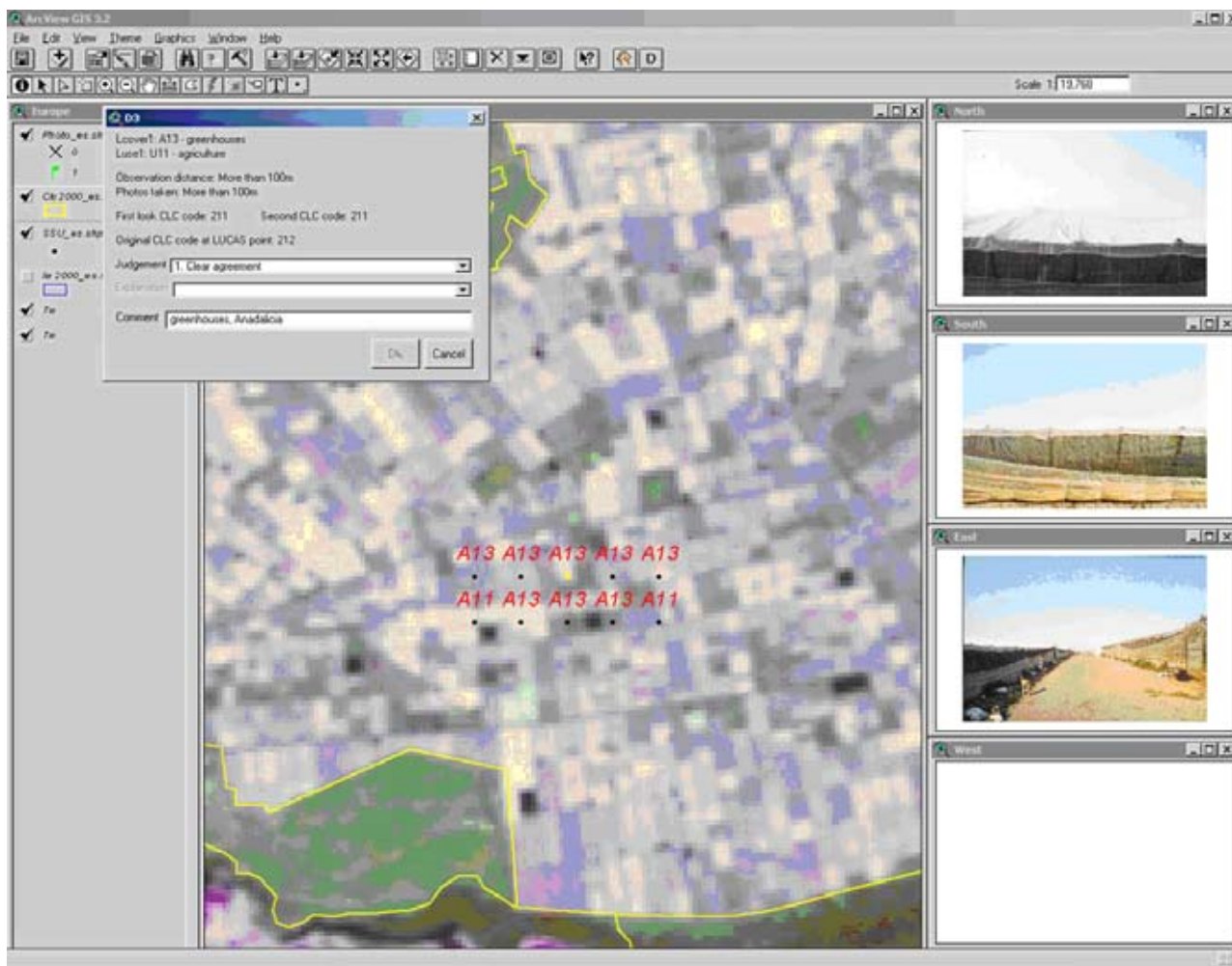
## Annex 9 Illustrations (screen shots taken during reinterpretation – see Section 3.2)

Illustration 1 Clear agreement (142, golf course in Andalusia, Spain)



**Note:** LLC: permanent grassland;  
LLU: recreation, leisure, sport;  
CLC2000: 142.  
Bad quality of LUCAS photos.

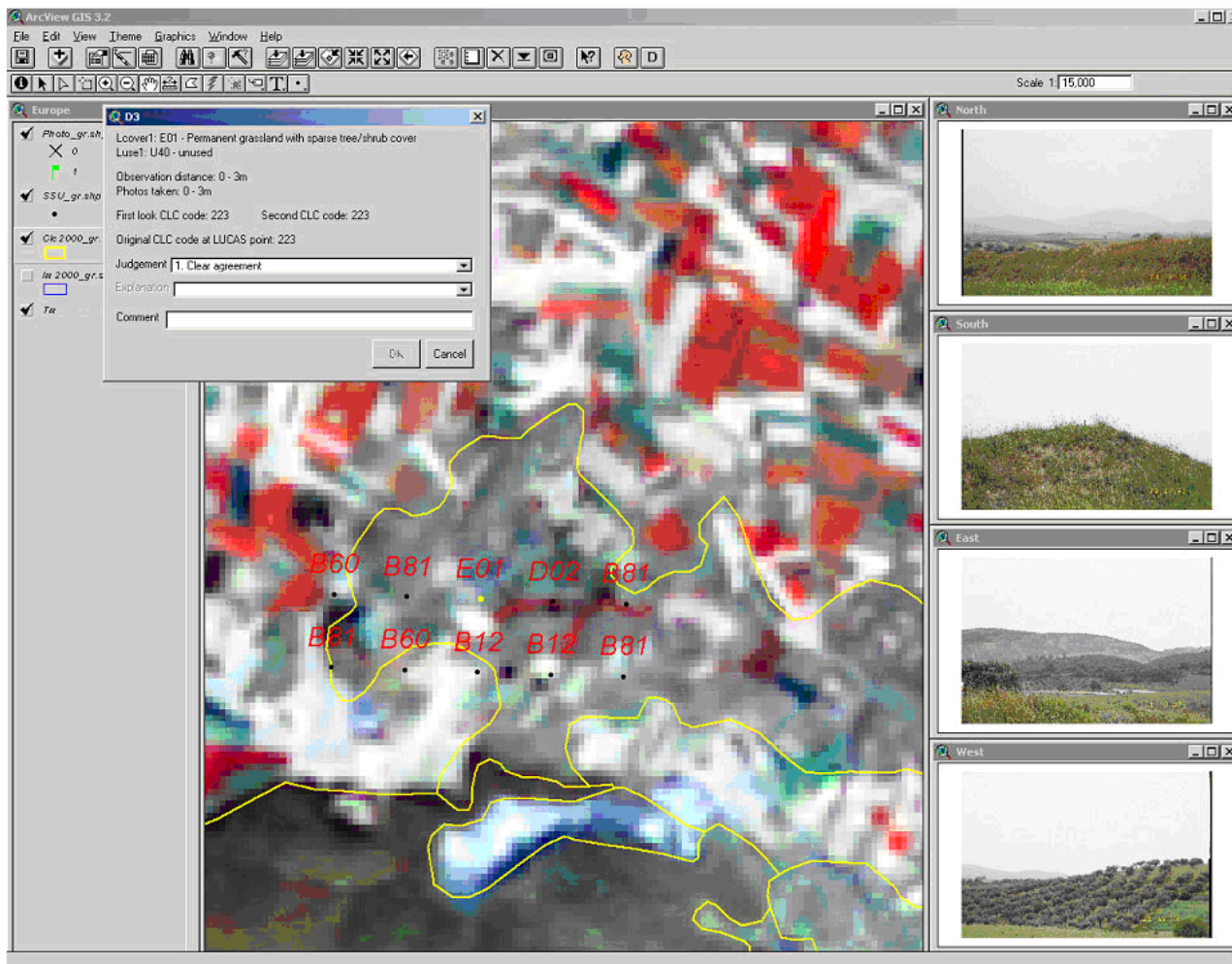
**Illustration 2 Clear agreement (212, Andaluca, Spain)**



**Note:** LLC: greenhouses;  
 LLU: agriculture;  
 CLC2000: 212.

'Clear agreement', although both validation codes were 211 (self-correction).

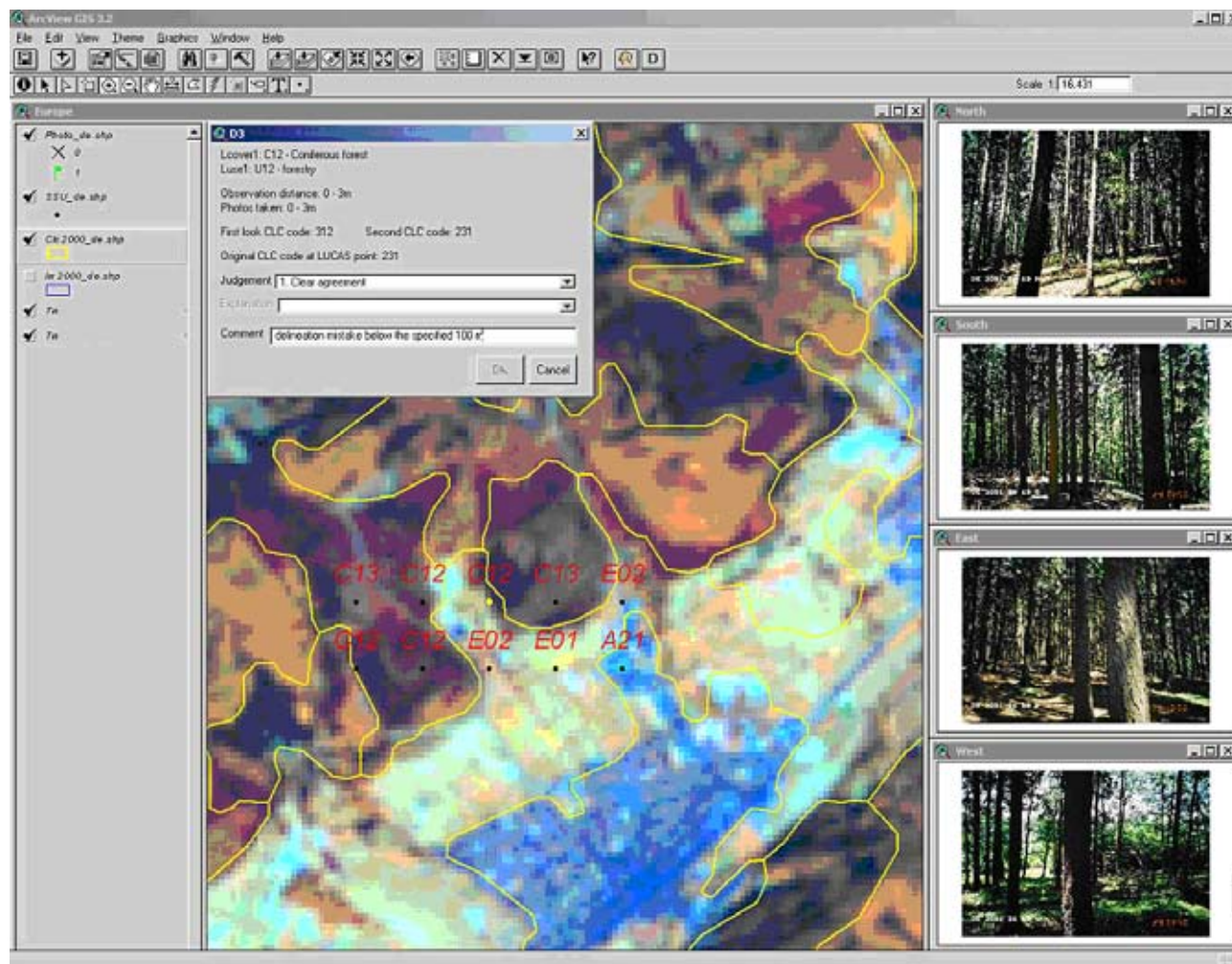
**Illustration 3 Clear agreement (223, Greece)**



**Note:** LLC: permanent grassland with sparse tree/shrub cover;  
 LLU: unused;  
 CLC2000: 223.

'Clear agreement' because grassland area is small, and photos of the 4 SSUs show olives.

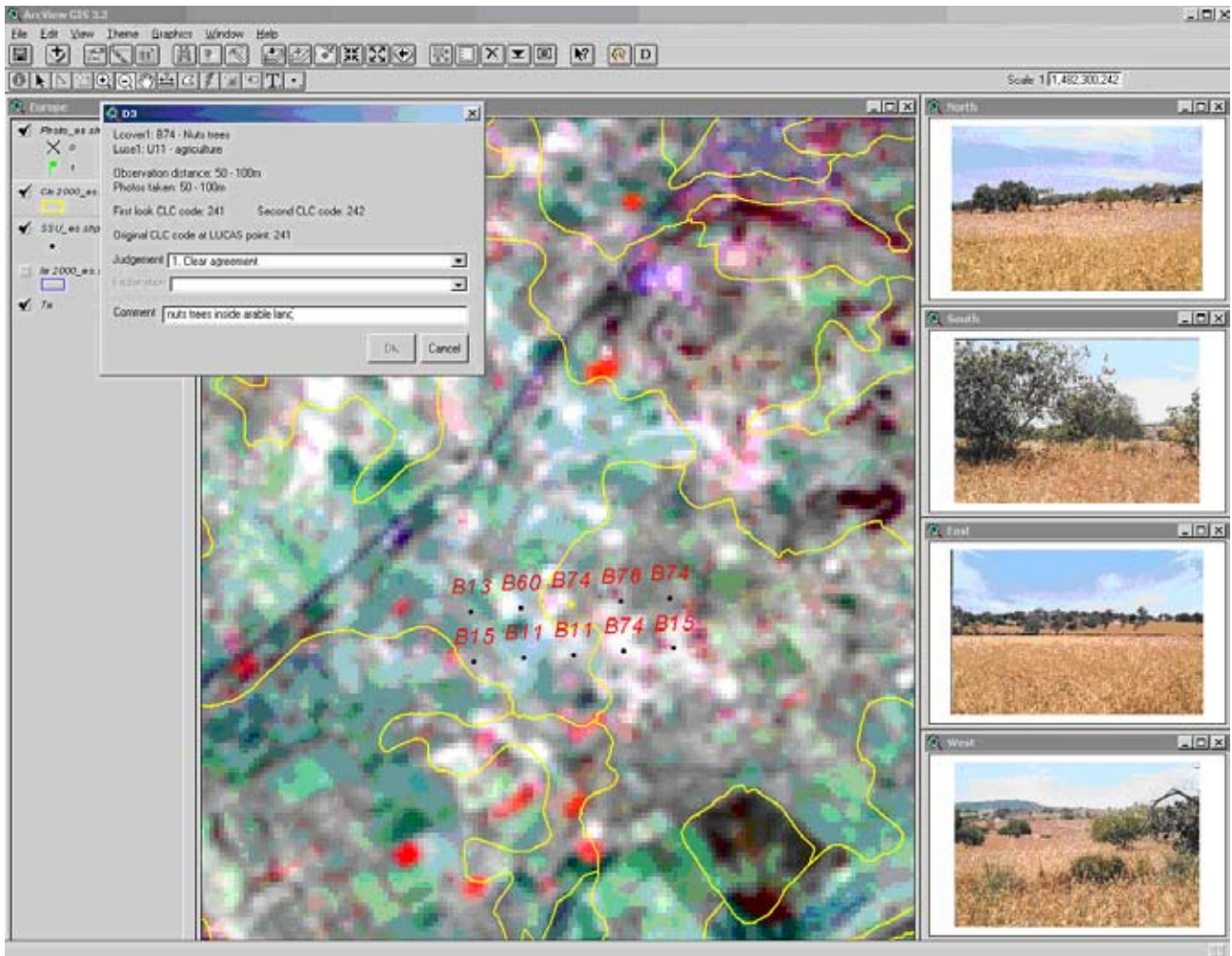
Illustration 4 Clear agreement (231/312, Germany)



**Note:** LLC: coniferous forest;  
 LLU: forestry;  
 CLC2000: 231 (neighbour polygon is 312).

'Clear agreement' because the delineation mistake was found to be below the specified 100 m.

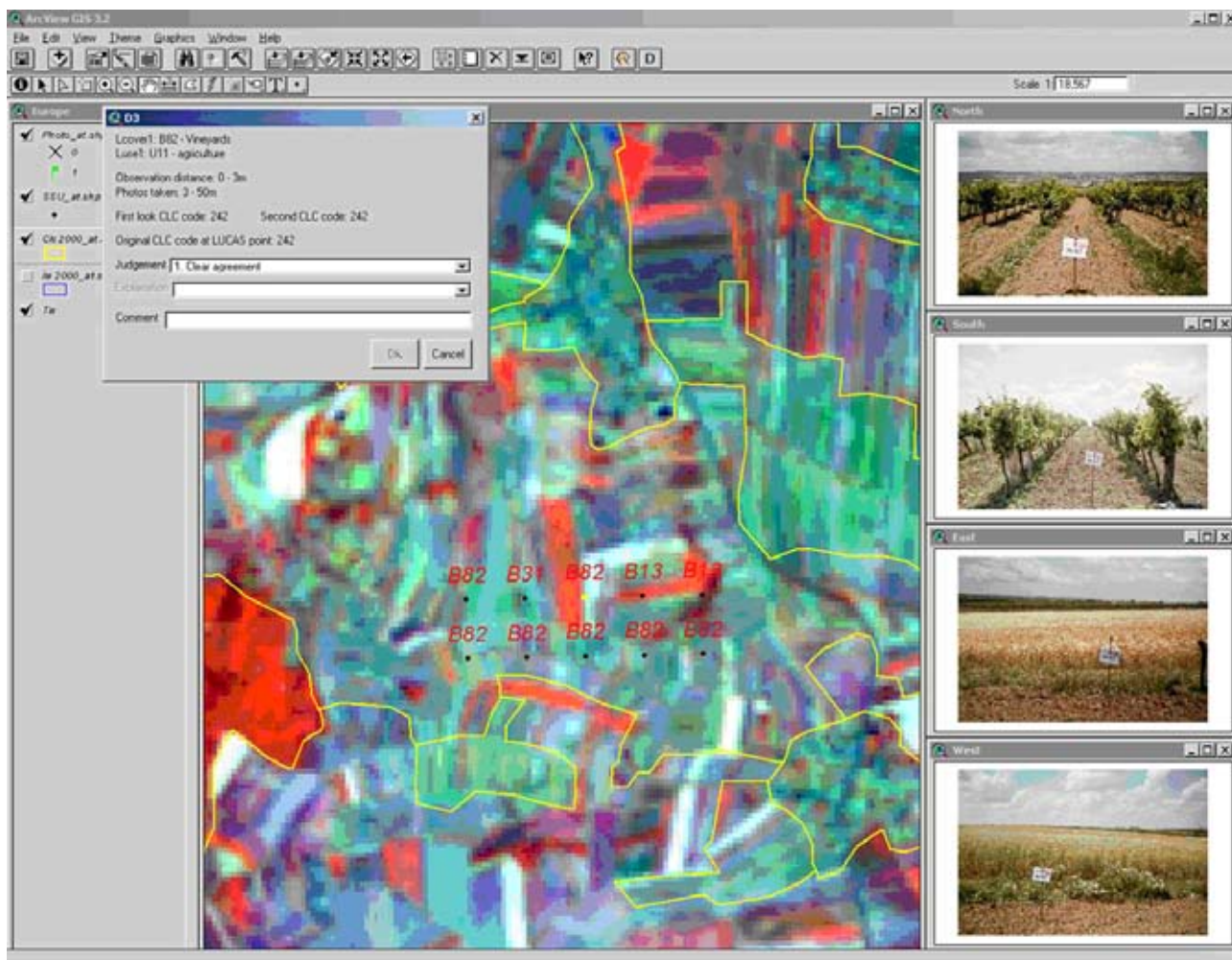
**Illustration 5 Clear agreement (241, Balears, Spain – nuts trees inside arable land)**



**Note:** LLC: nuts trees;  
LLU: agriculture;  
CLC2000: 241.  
'Clear agreement' although the 2nd validation code was 242.

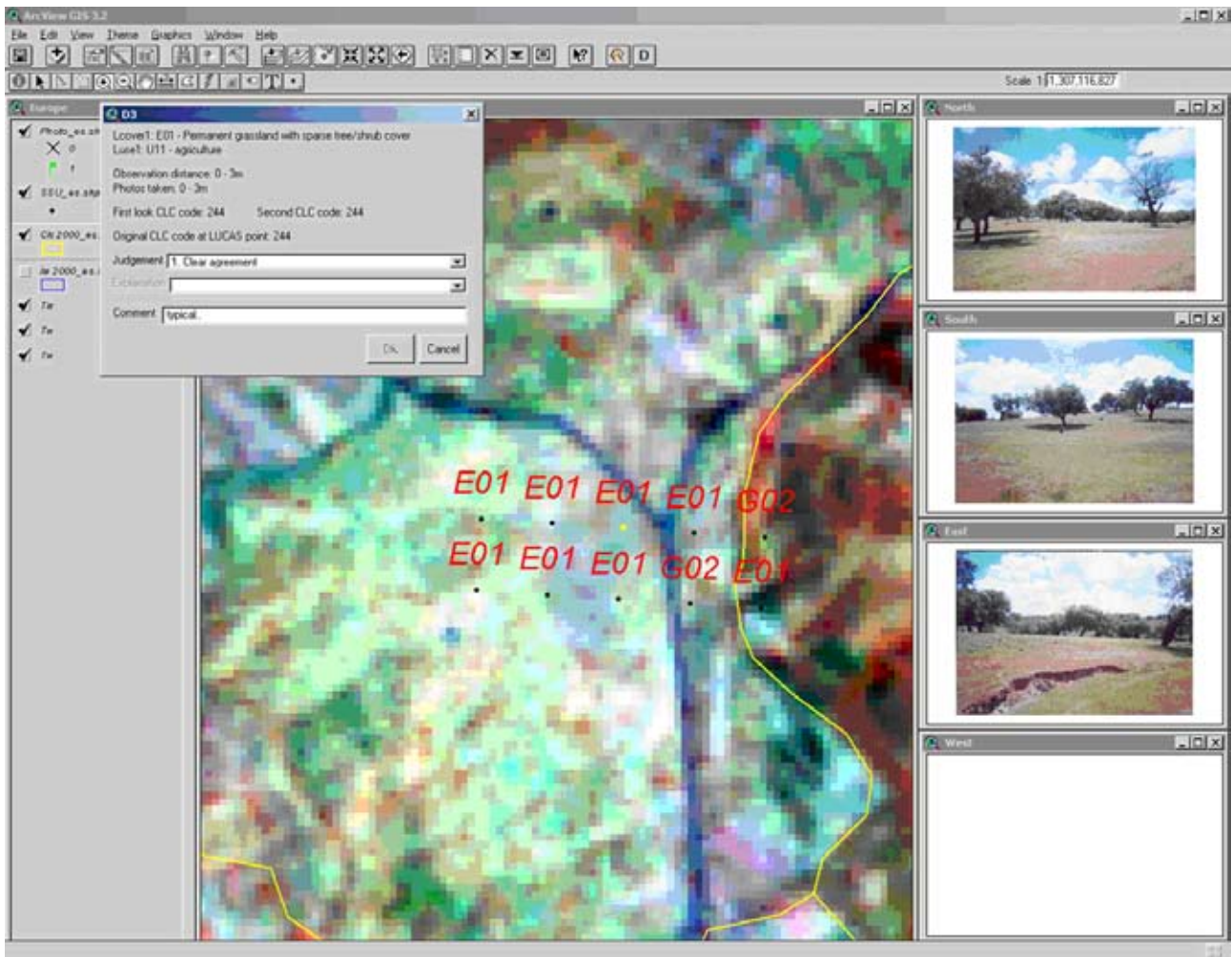


**Illustration 6 Clear agreement (242, Austria – mixed small parcels of annual and permanent crops)**



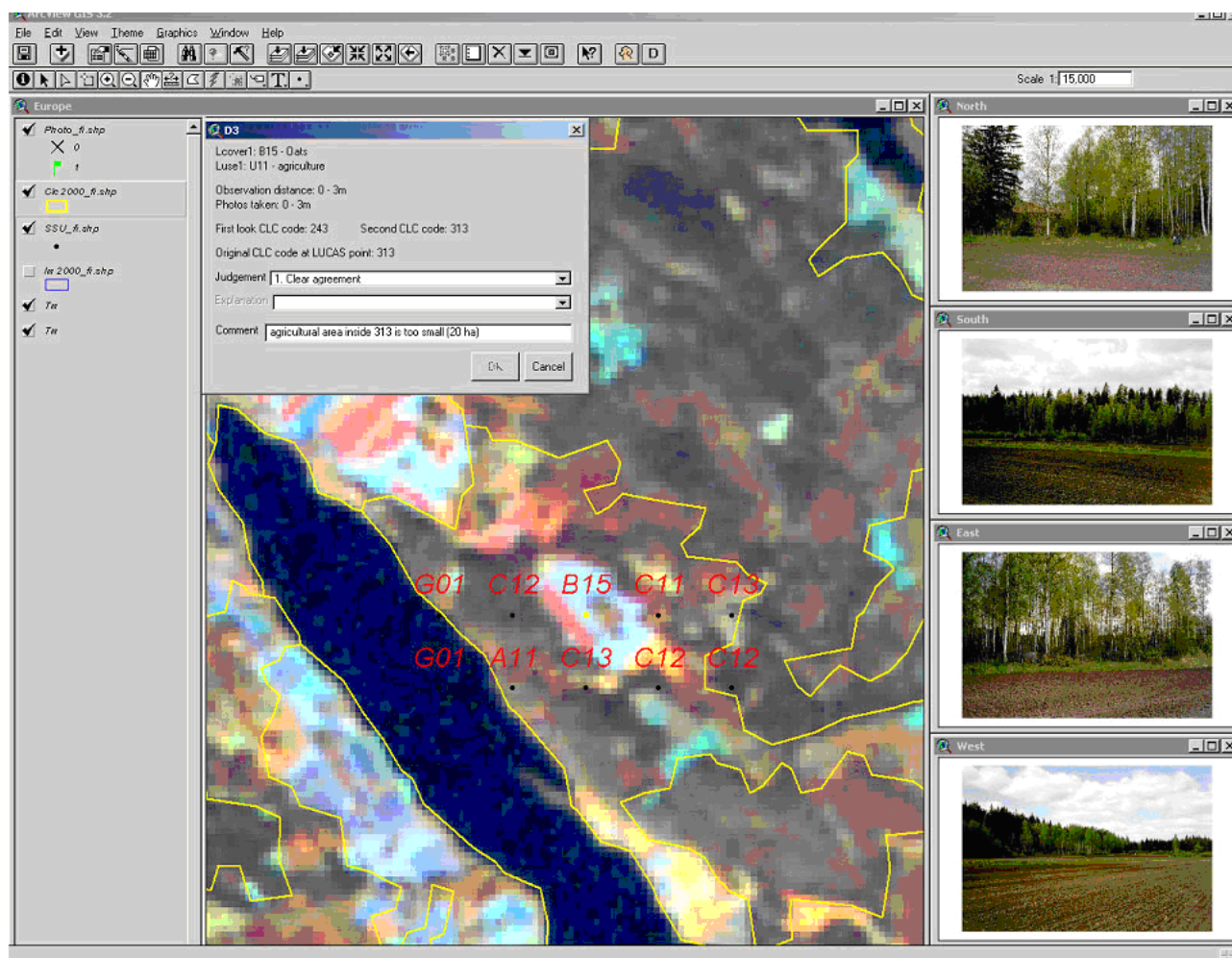
**Note:** LLC: vineyards;  
 LLU: agriculture;  
 CLC2000: 242.

**Illustration 7 Clear agreement (244, Andalusia, Spain – grassland with scattered forest trees (oaks))**



**Note:** LLC: permanent grassland with sparse tree/shrub cover;  
LLU: agriculture;  
CLC2000: 244.

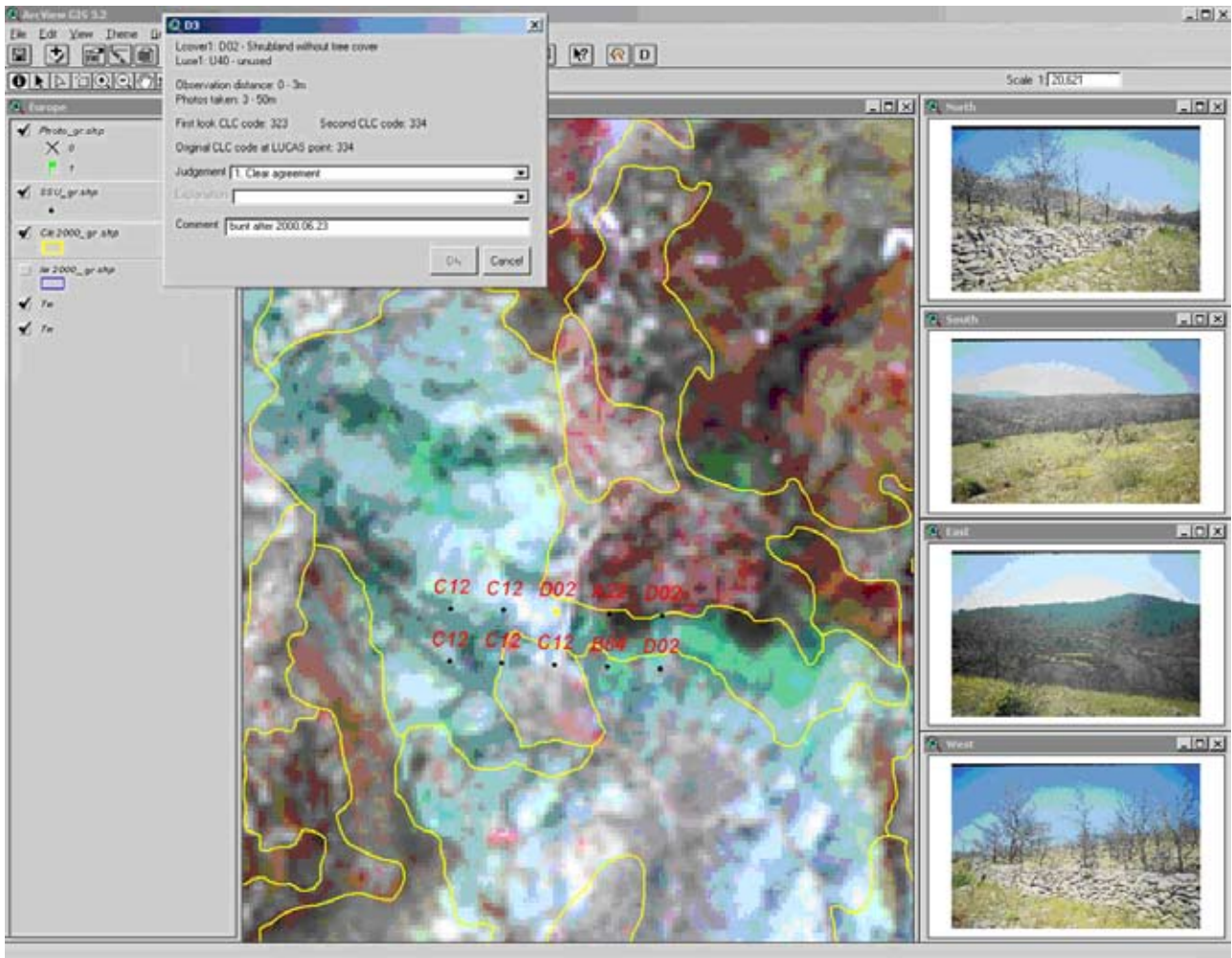
**Illustration 8 Clear agreement (313, Finland)**



**Note:** LLC: oats;  
 LLU: agriculture;  
 CLC2000: 313.

'Clear agreement' because arable land area is small (realised when entering the 2nd validation code).

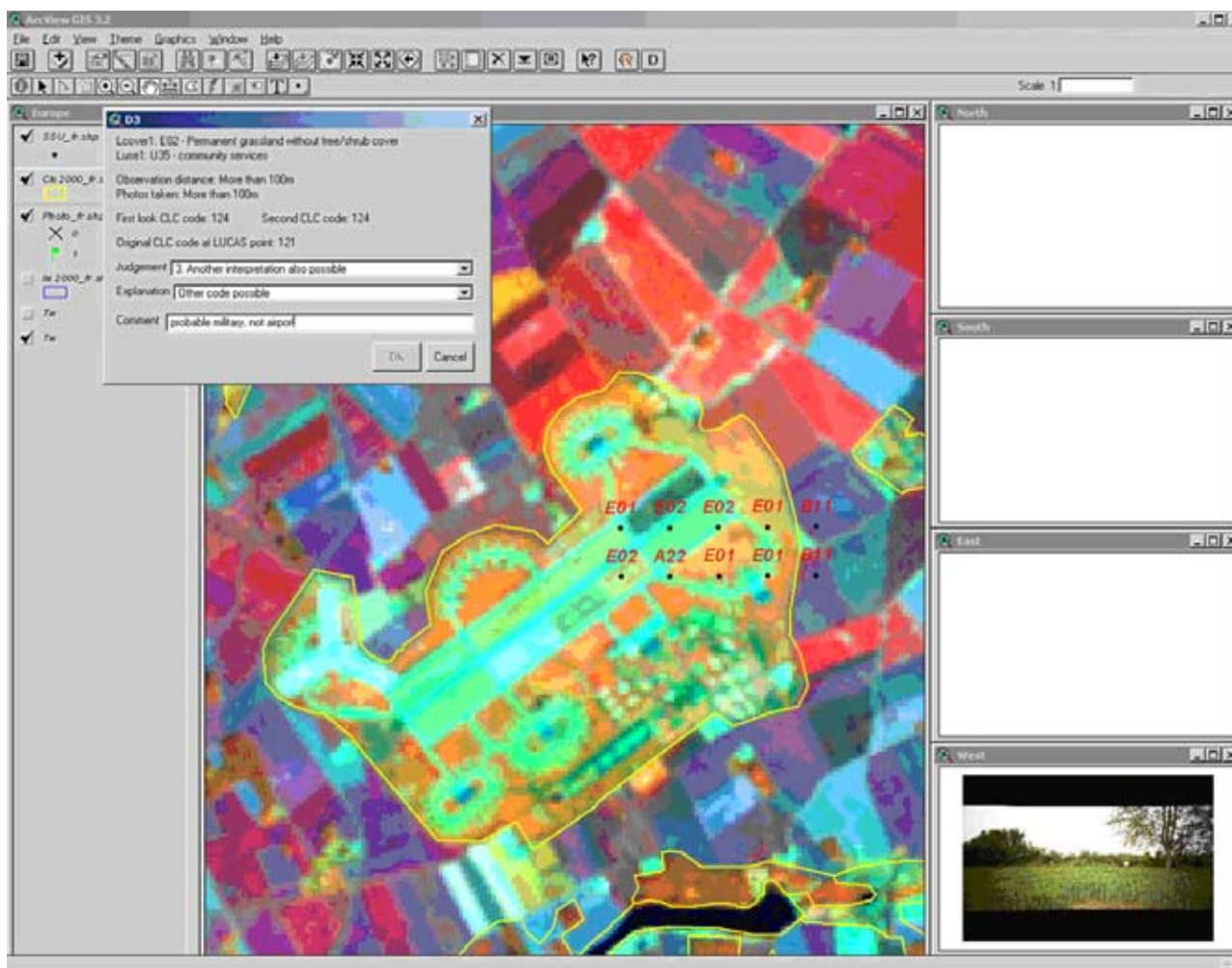
**Illustration 9 Clear agreement (334, Greece)**



**Note:** LLC: shrubland without sparse tree cover;  
 LLU: unused;  
 CLC2000: 334.

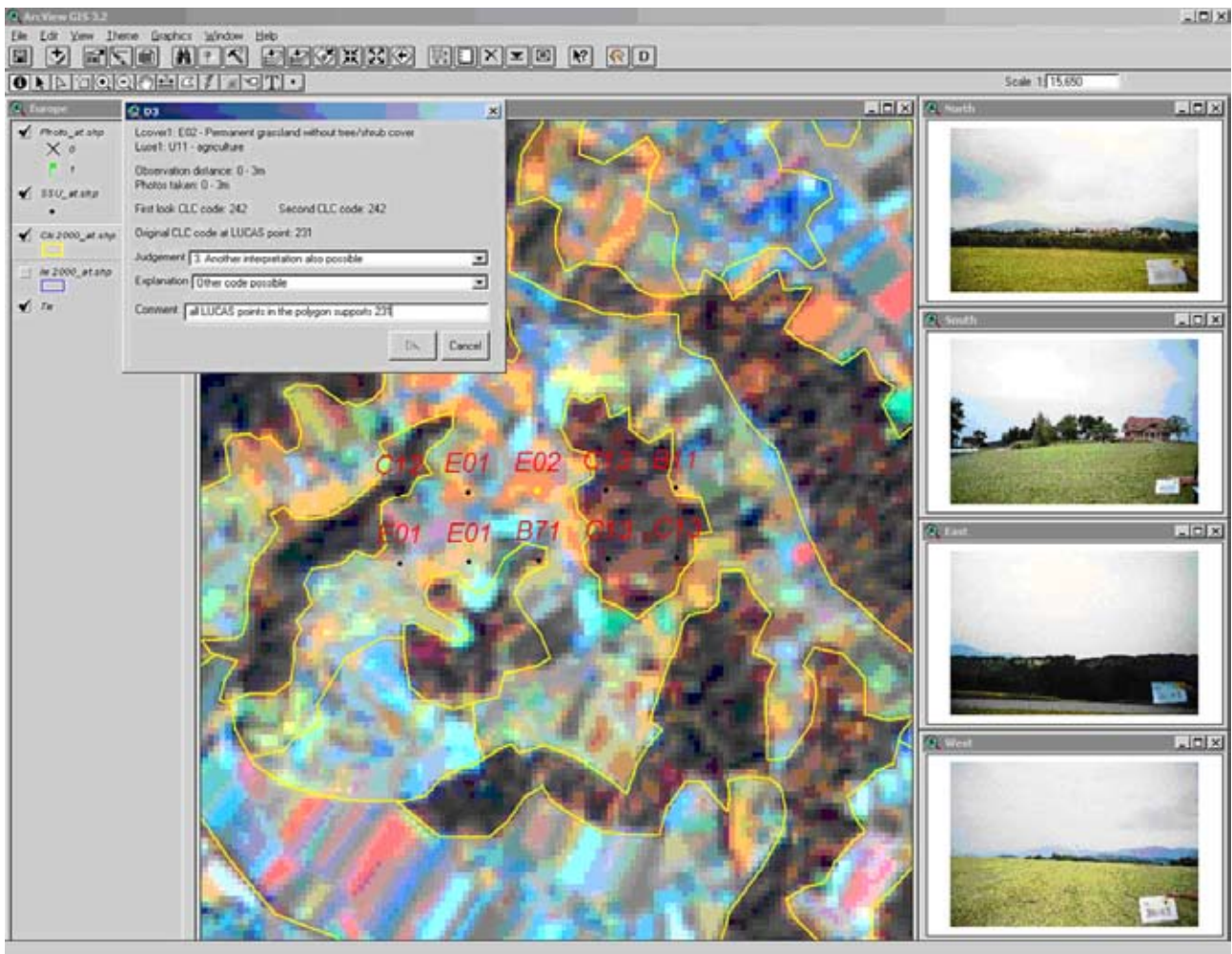
'Clear agreement' although the 1st validation code was 323. This decision was supported by the additional LUCAS variable: NATHAZARD = 1.

**Illustration 10 Another interpretation is also possible/other code possible (121/124)**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: community services;  
 CLC2000: 121.  
 Could be a military site with airport (test facility?).

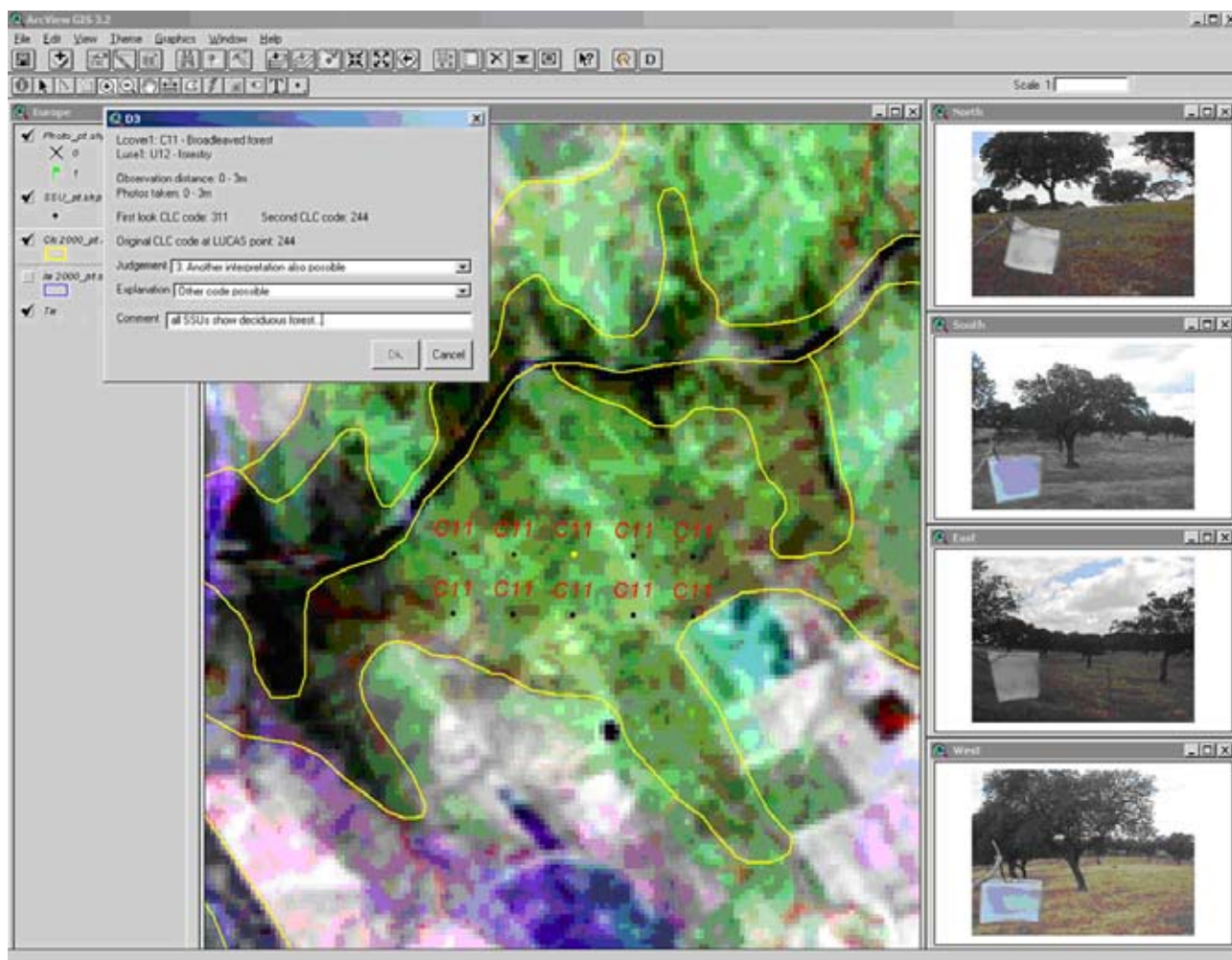
**Illustration 11 Another interpretation is also possible/other code possible (231/242, Austria)**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: agriculture;  
 CLC2000: 231.

Although both validation codes were 242, no 'wrong interpretation' was given. This was due to the presence of four SSUs coded as grassland in the polygon.

**Illustration 12 Another interpretation is also possible/other code possible (244/311, Spain)**

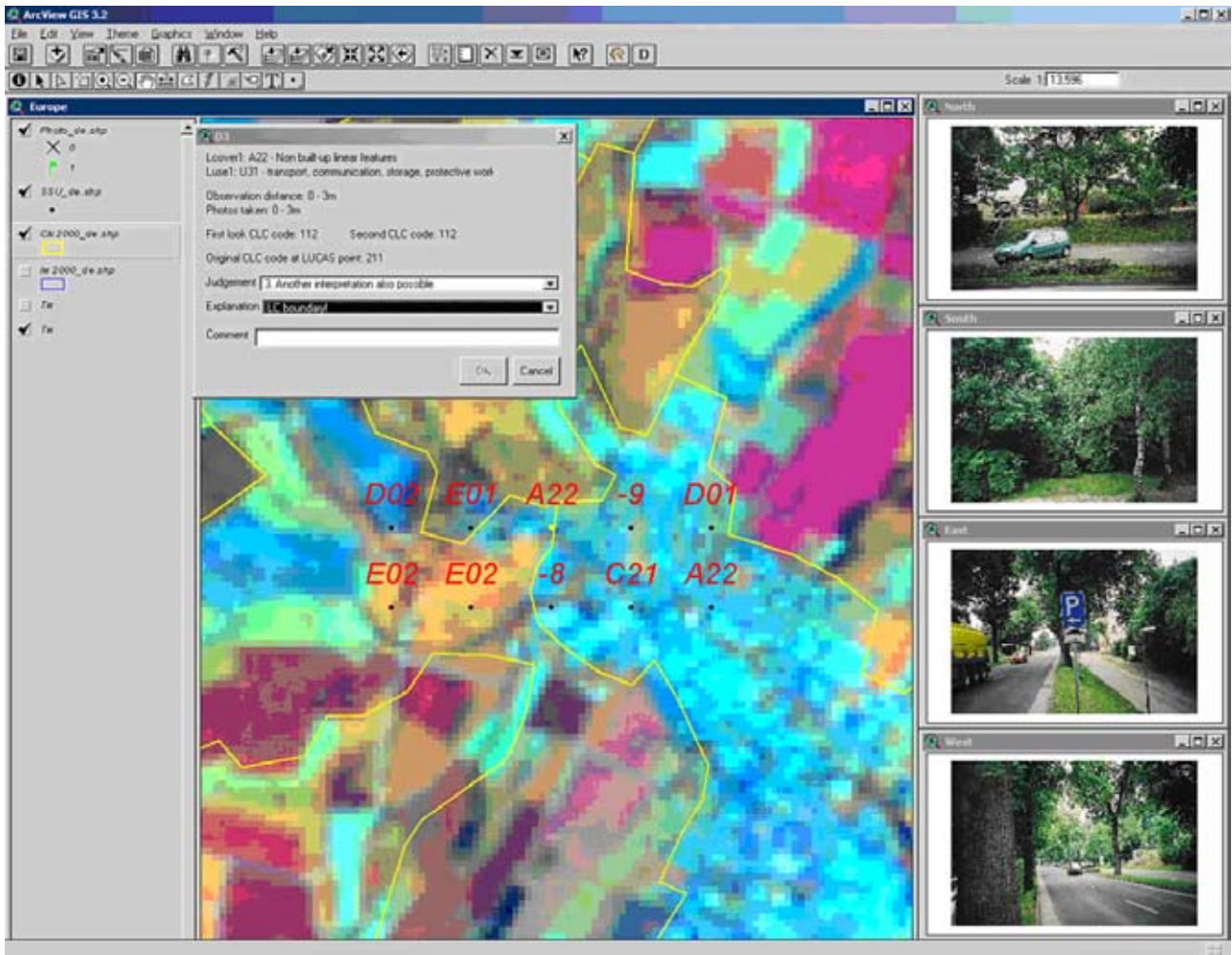


**Note:** LLC: broadleaved forest;  
 LLU: forestry;  
 CLC2000: 244.

Although both validation codes were 311, no 'wrong interpretation' was given, because photographs support the typical 'Dehesa' (i.e. forest trees over agricultural land).

LUCAS, obviously, cannot handle this typical Mediterranean agricultural landscape type.

**Illustration 13** Another interpretation is also possible/LC boundary (211/112, Germany)

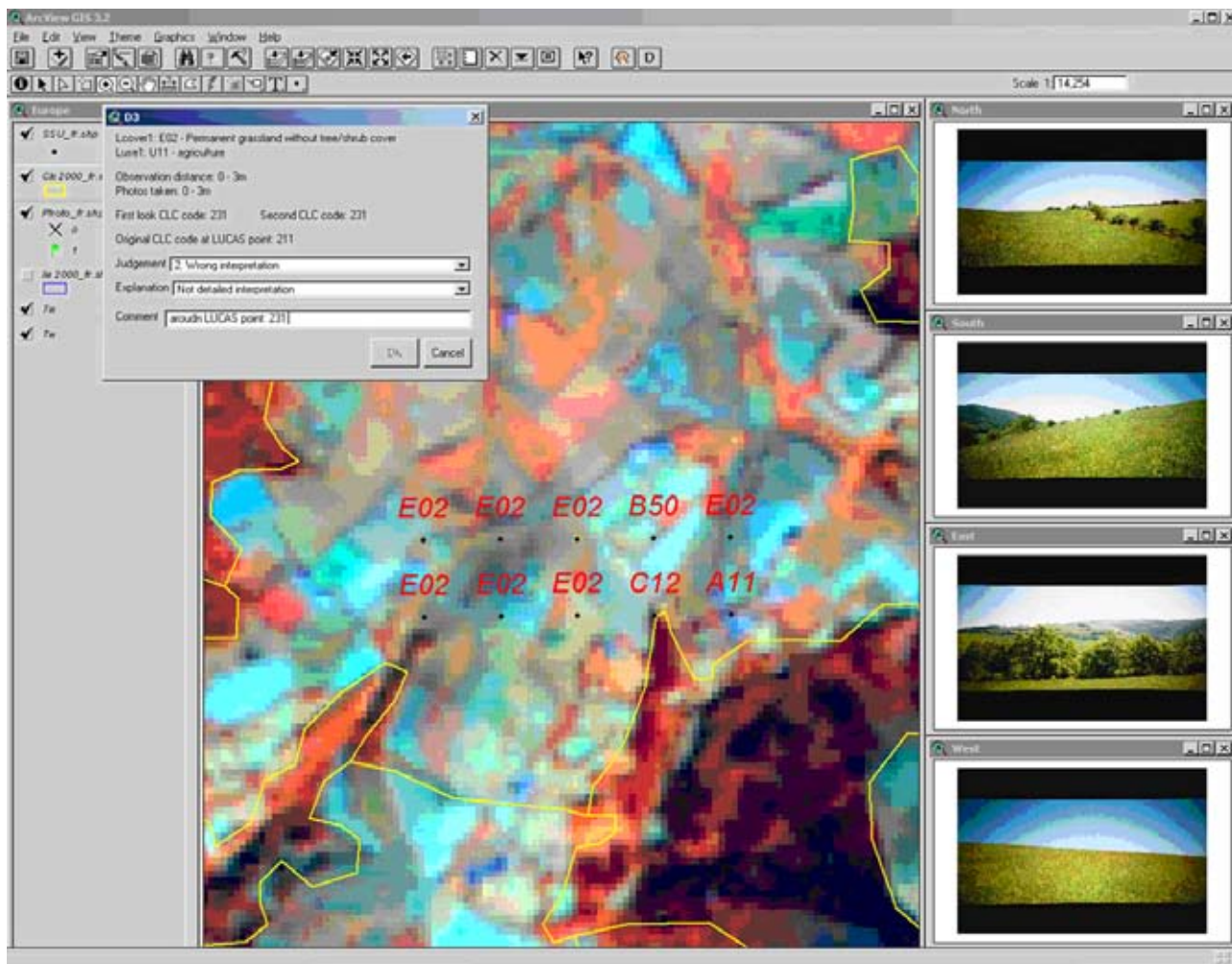


**Note:** LLC: non built-up linear features;  
 LLU: transport, communication, storage, protective works;  
 CLC2000: 211.

Although both validation codes were 112, no 'wrong interpretation' was given because the SSU is on the boundary between settlement and arable land.

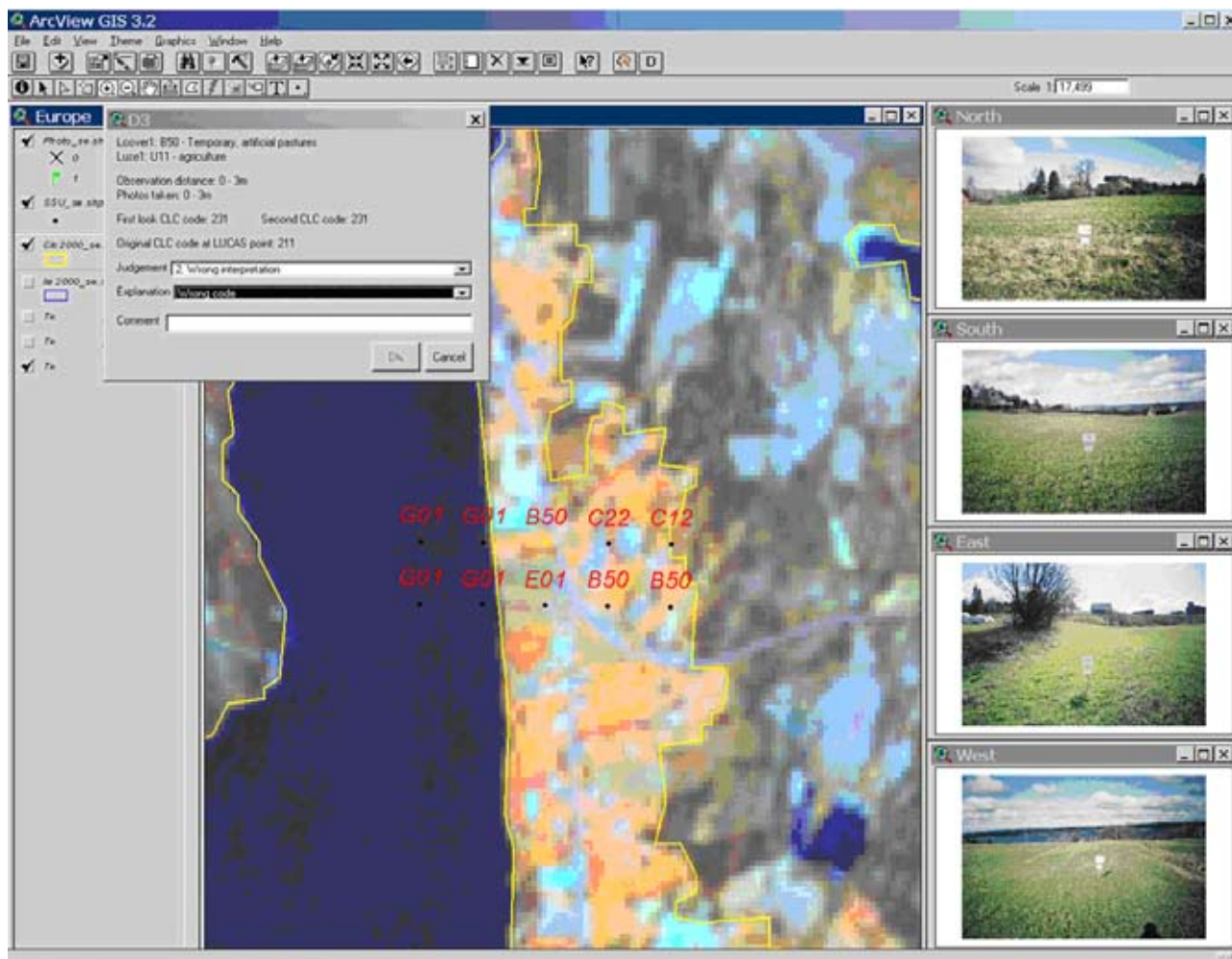


**Illustration 14 Wrong interpretation/not detailed interpretation**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: agriculture;  
 Validation: 231, 231;  
 CLC2000: 211.  
 Around LUCAS photo point: 231, polygon is not detailed enough.

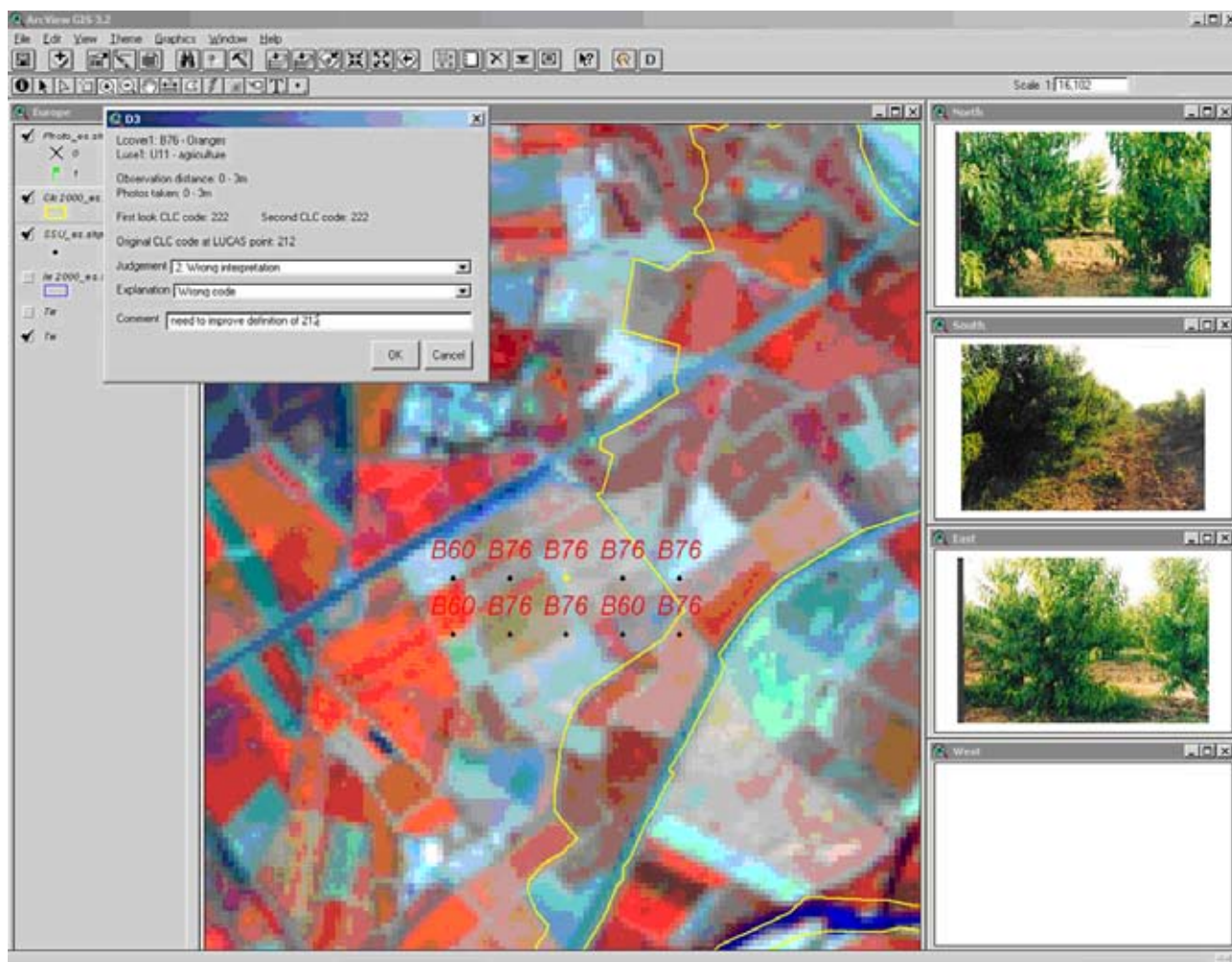
**Illustration 15 Wrong interpretation/wrong code**



**Note:** LLC: temporary, artificial pastures;  
 LLU: agriculture;  
 Validation: 231, 231;  
 CLC2000: 211.

LUCAS photos and 4 SSUs clearly show pasture. Note the two different LLC codes for pastures: B50 and E01.

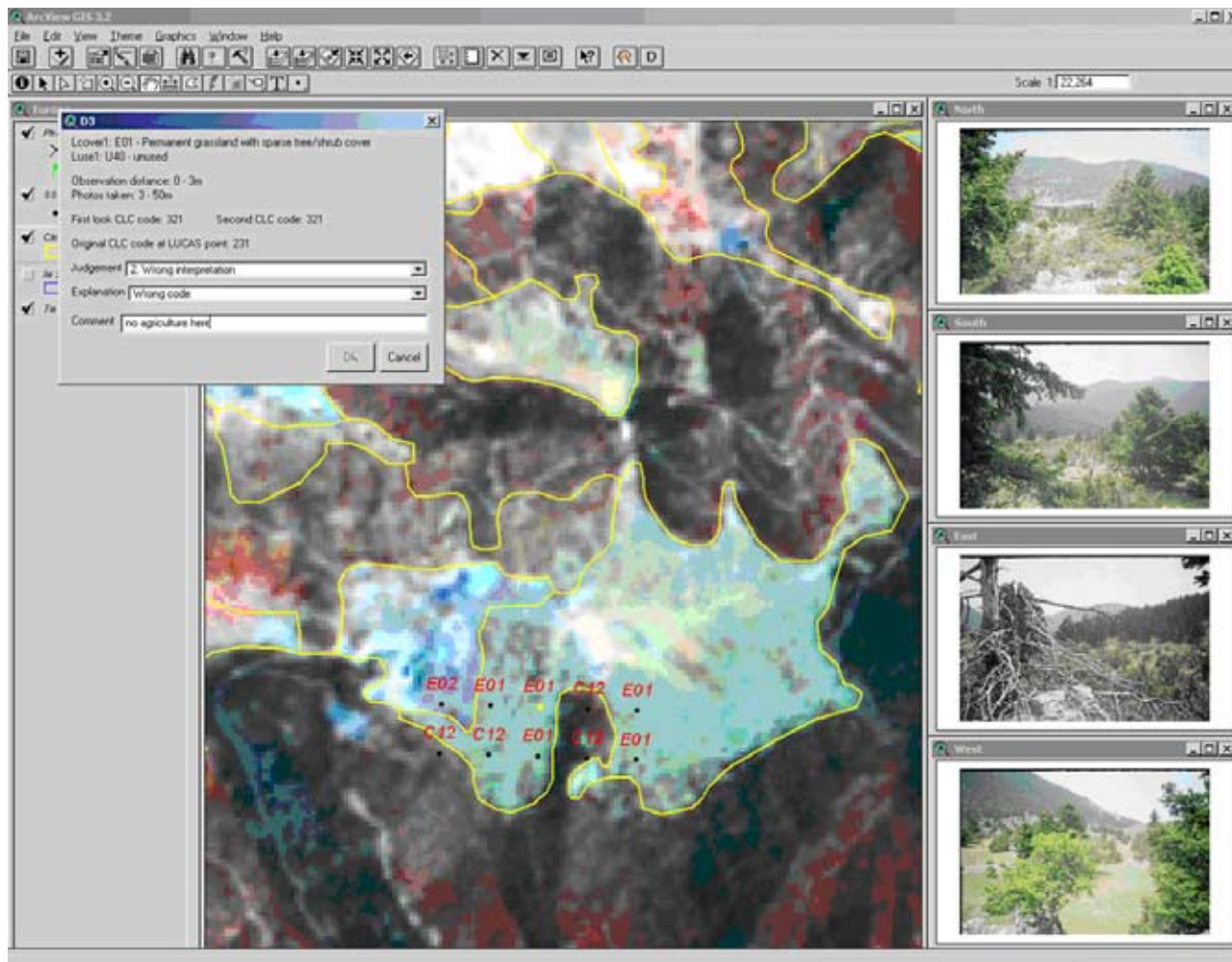
**Illustration 16 Wrong interpretation/wrong code**



**Note:** LLC: oranges;  
 LLU: agriculture;  
 Validation: 222, 222;  
 CLC2000: 212.

Oranges are coded as irrigated land (it is true, but contradicts the definition of 212).

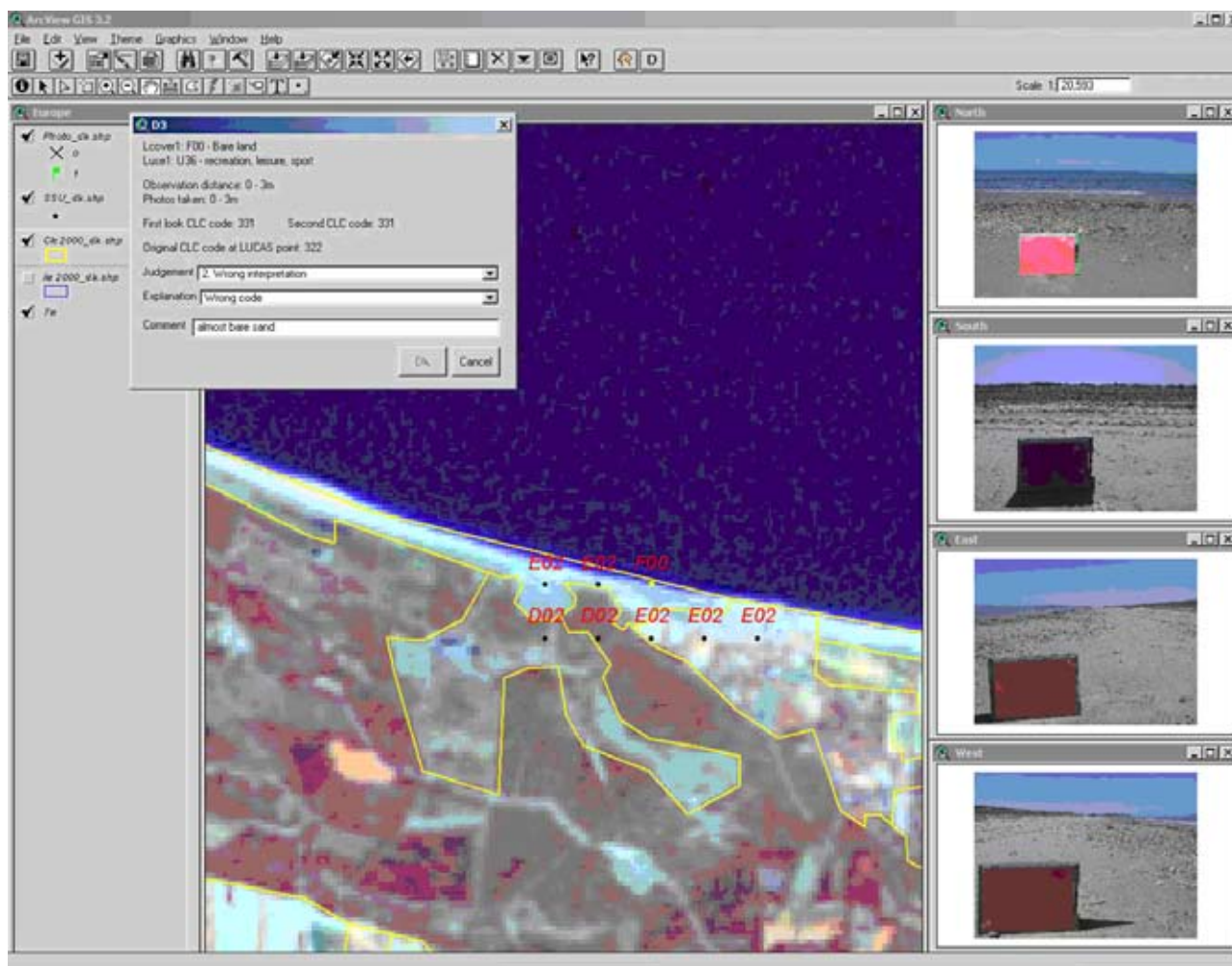
**Illustration 17 Wrong interpretation/wrong code**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: unused;  
 Validation: 321, 321;  
 CLC2000: 231.

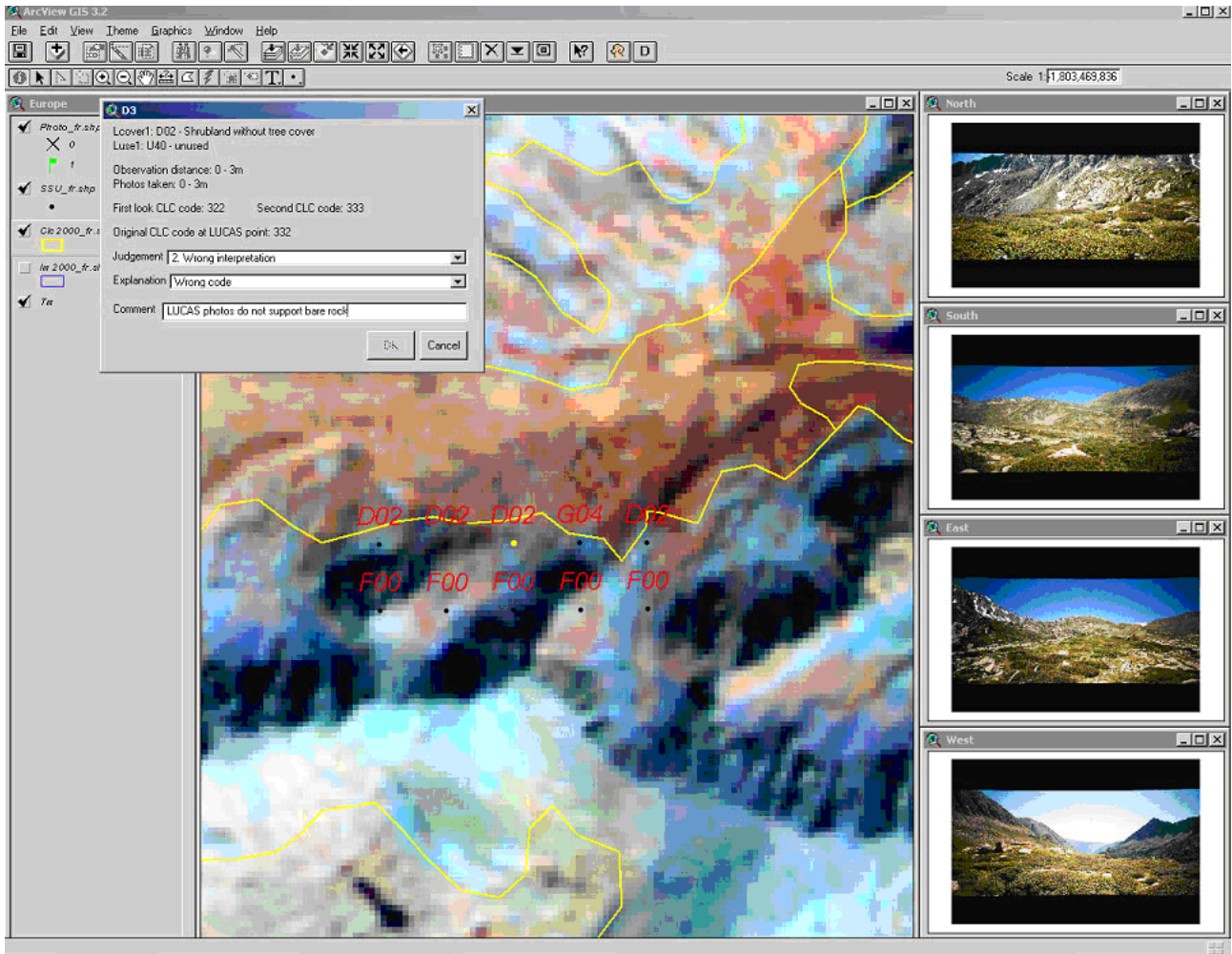
Large polygon, no settlements nearby and no evidence of human impact. It is considered unused according to LUCAS.

Illustration 18 Wrong interpretation/wrong code



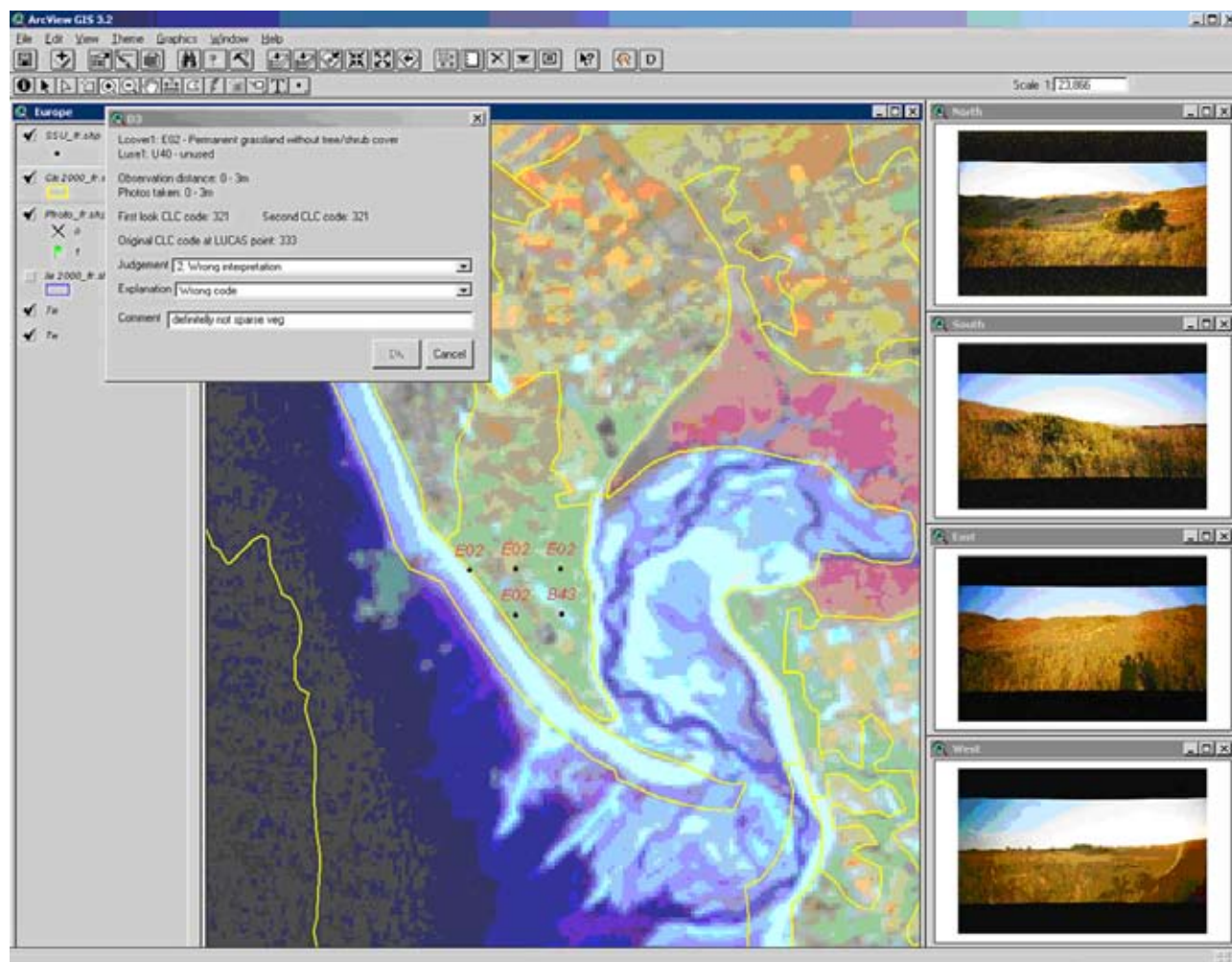
**Note:** LLC: bare land;  
 LLU: recreation, leisure, sport;  
 Validation: 331, 331;  
 CLC2000: 322.  
 No significant vegetation cover.

**Illustration 19 Wrong interpretation/wrong code**



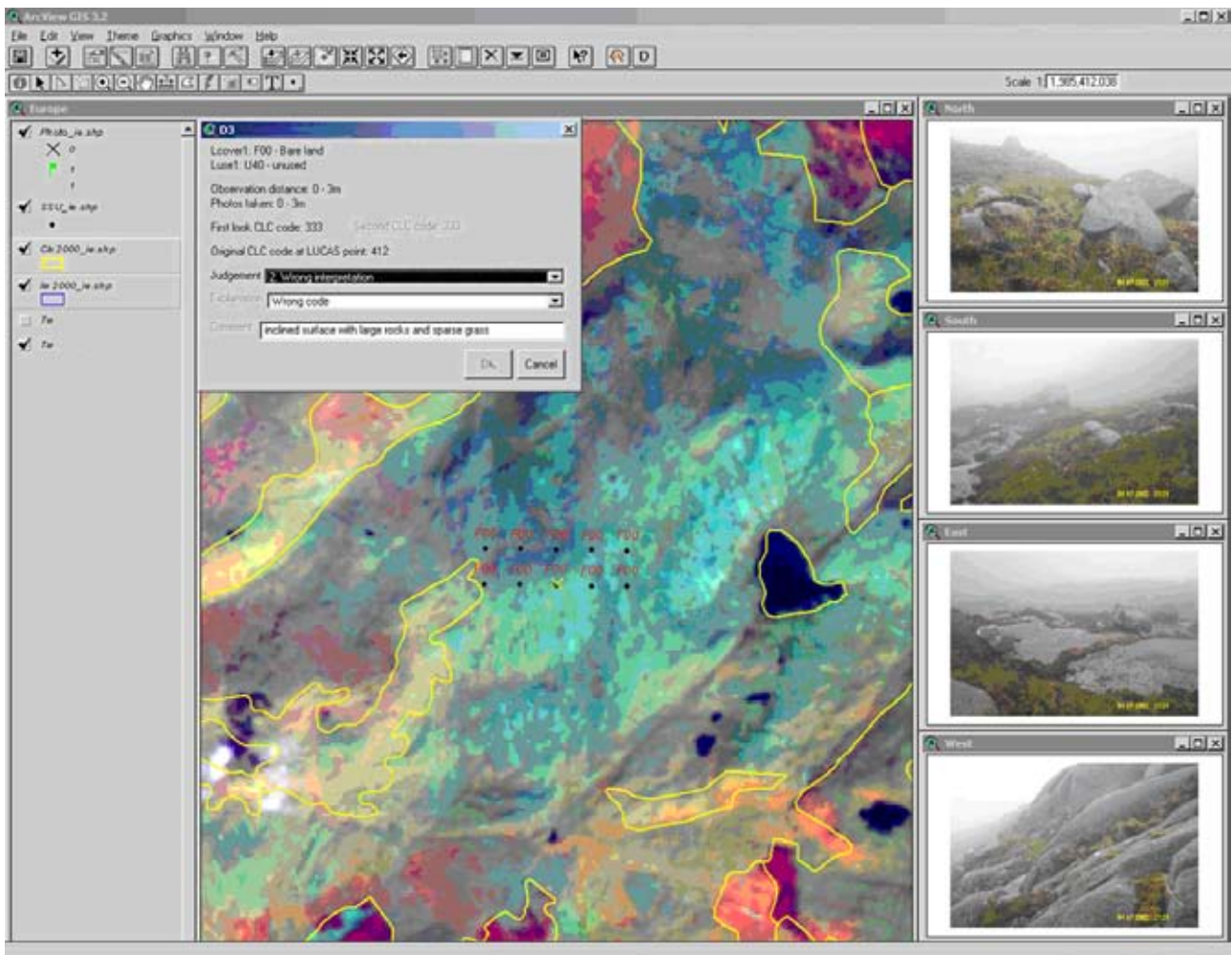
**Note:** LLC: shrubland without tree cover;  
 LLU: unused;  
 Validation: 322, 333;  
 CLC2000: 332.  
 LUCAS photos show some vegetation cover.

**Illustration 20 Wrong interpretation/wrong code**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: unused;  
 Validation: 321, 321;  
 CLC2000: 333.  
 LUCAS photos show no sparse vegetation.

**Illustration 21 Wrong interpretation/wrong code**

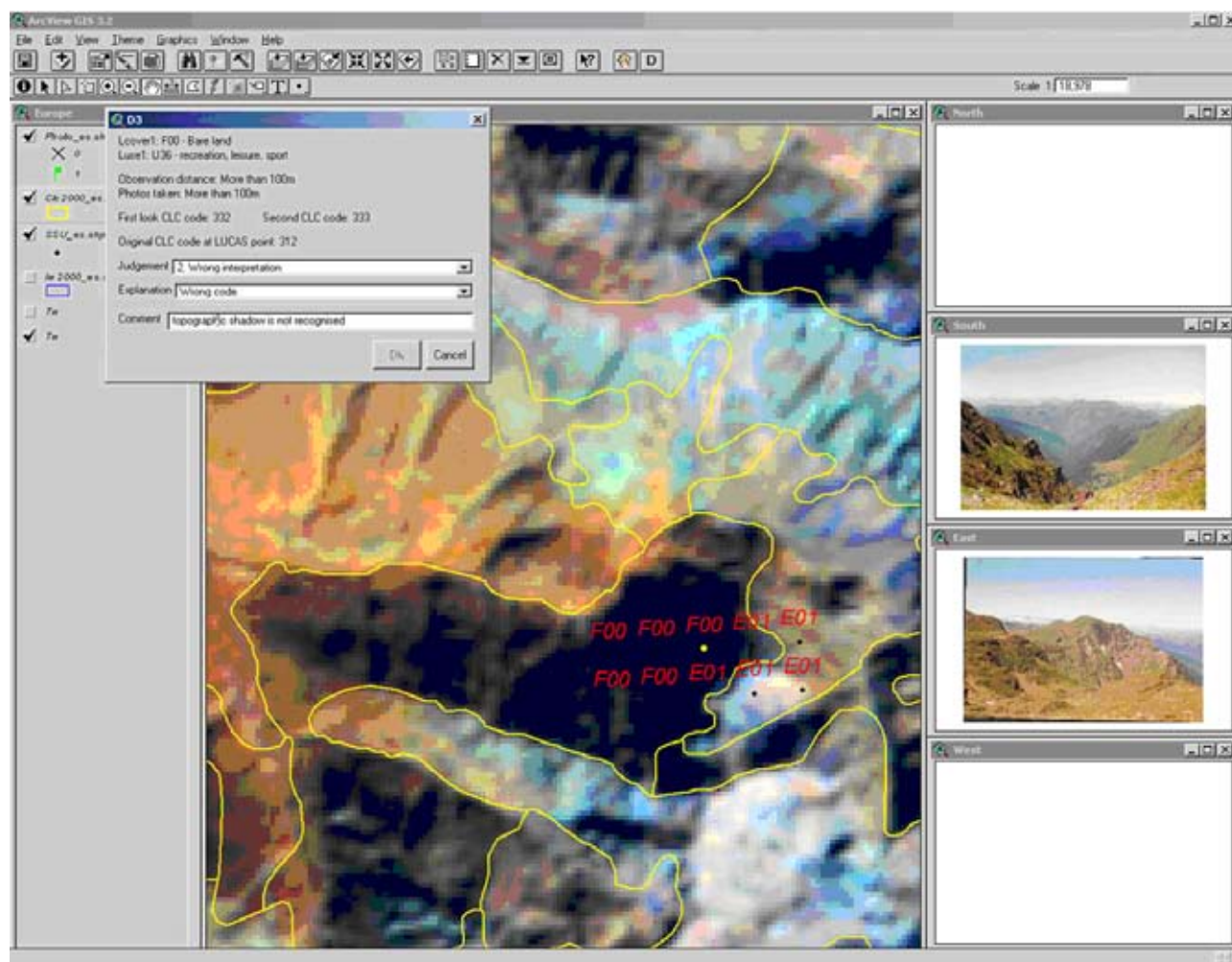


**Note:** LLC: bare land;  
 LLU: unused;  
 Validation: 333, 333;  
 CLC2000: 412.

LUCAS photos show an inclined, rocky surface with sparse grass cover; no peatland.



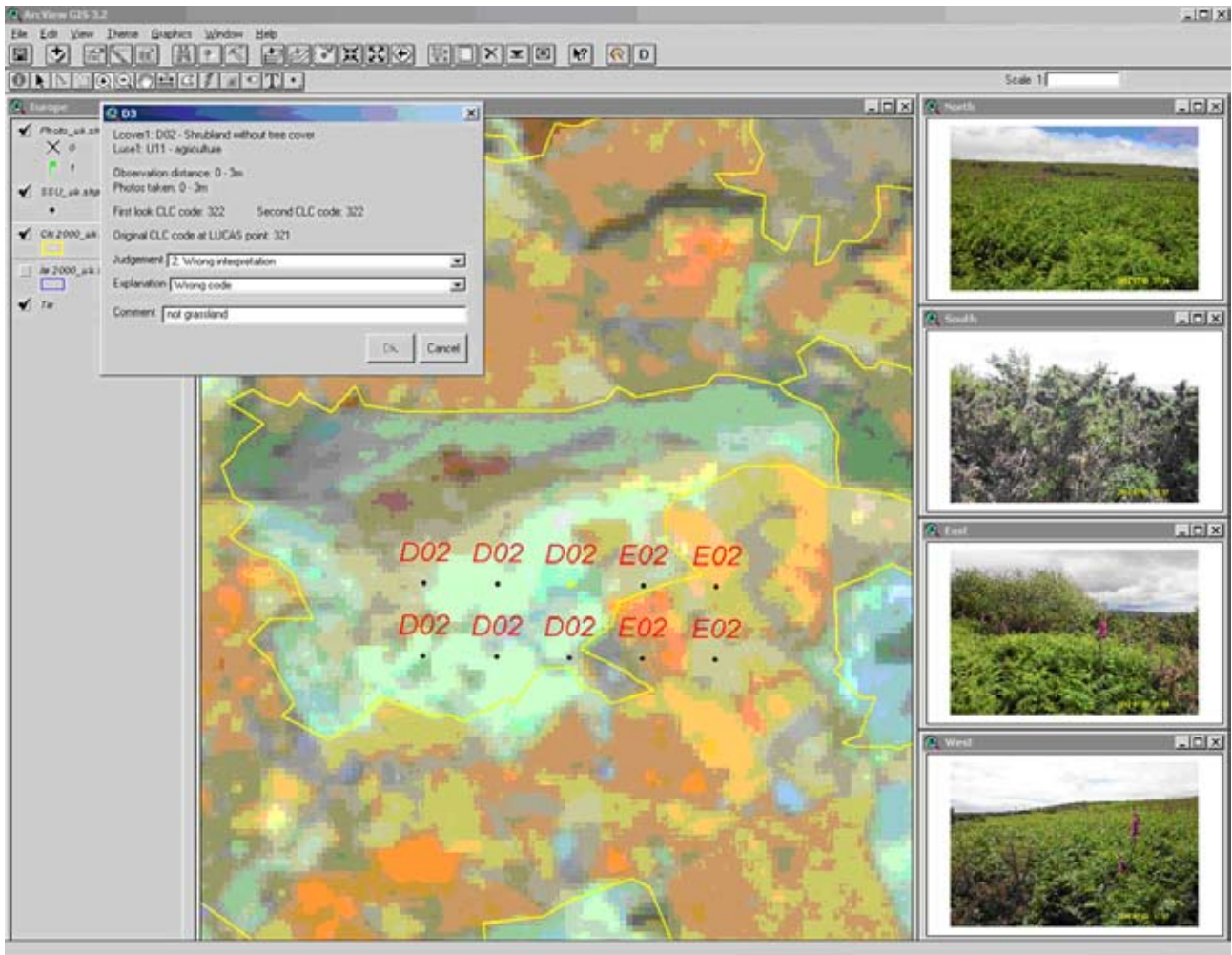
Illustration 22 Wrong interpretation/wrong code



**Note:** LLC: bare land;  
 LLU: recreation, leisure, sport;  
 Validation: 332, 333;  
 CLC2000: 312.

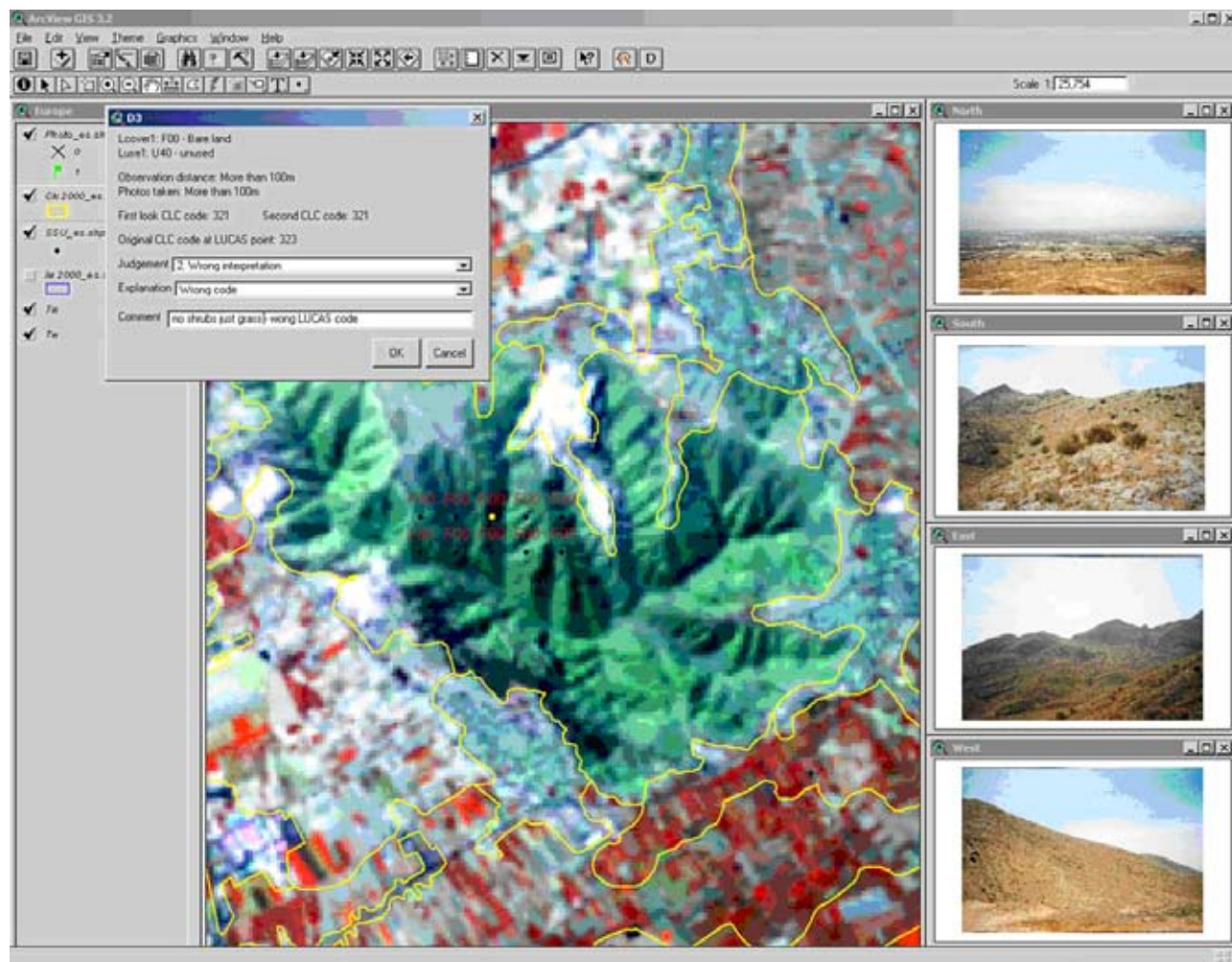
LUCAS photos show a high mountainous area with sparse vegetation. The interpreter was confused by the topographic shadow. Both LUCAS codes are questionable.

**Illustration 23 Wrong interpretation/wrong code**



**Note:** LLC: shrubland without tree cover;  
 LLU: agriculture;  
 Validation: 322, 322;  
 CLC2000: 321.  
 Shrubland misclassified as grassland.

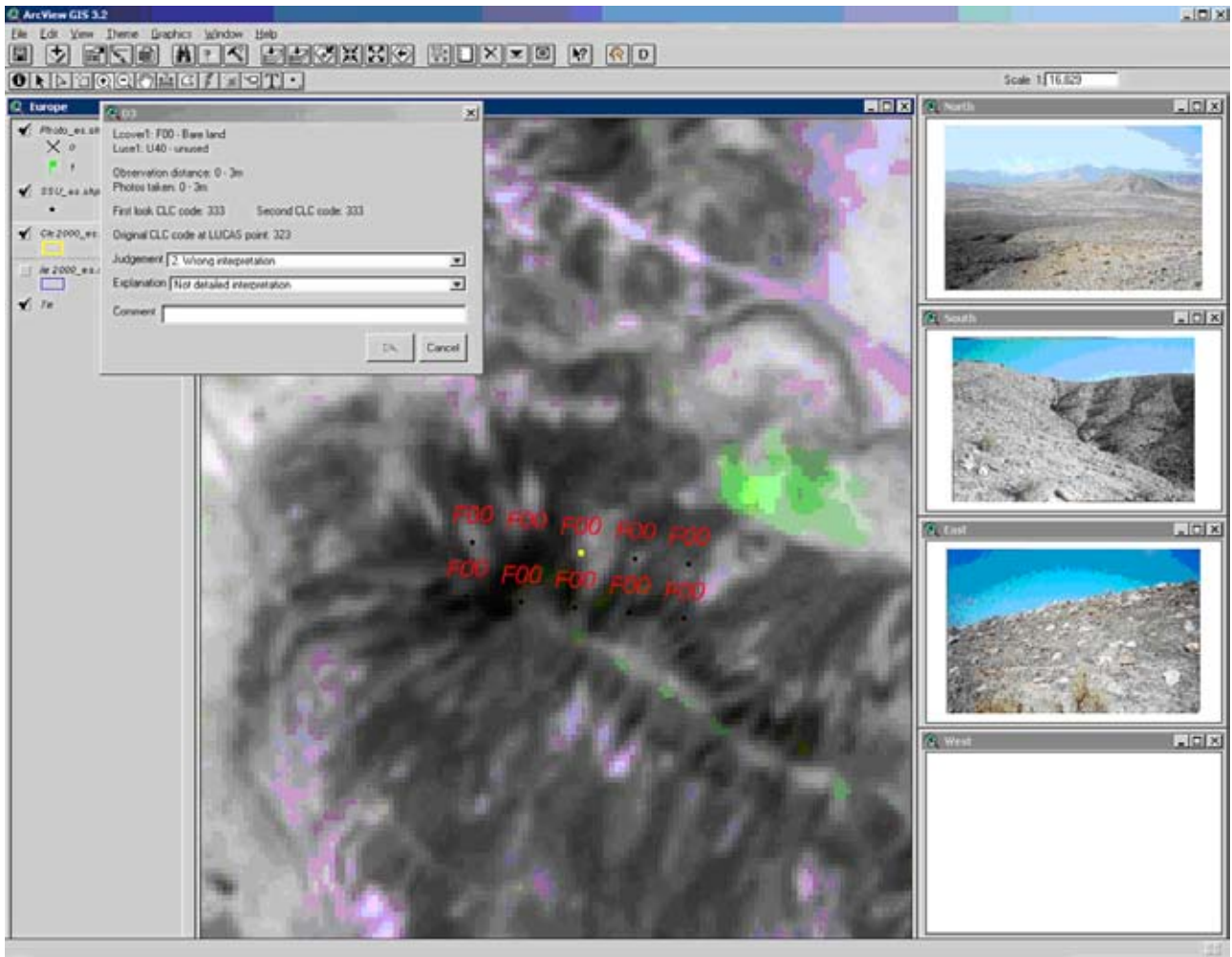
**Illustration 24 Wrong interpretation/wrong code**



**Note:** LLC: bare land;  
 LLU: unused;  
 Validation: 321, 321;  
 CLC2000: 323.

Photos show natural grassland and not shrubland. LUCAS LC code is questionable.

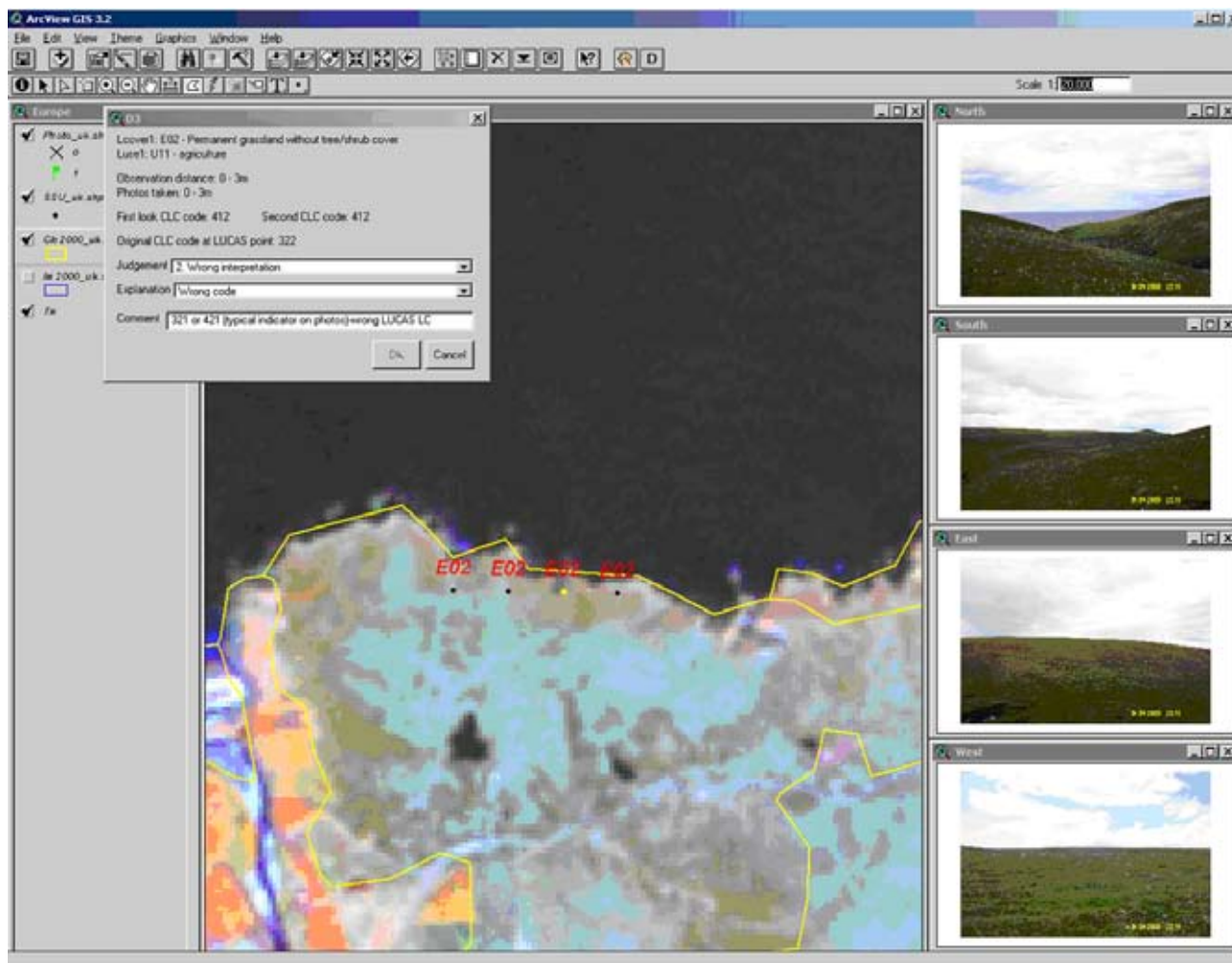
**Illustration 25 Wrong interpretation/wrong code**



**Note:** LLC: bare land;  
LLU: unused;  
Validation: 333, 333;  
CLC2000: 323.

Photos show rather sparse vegetation and not sclerophyllous shrub.

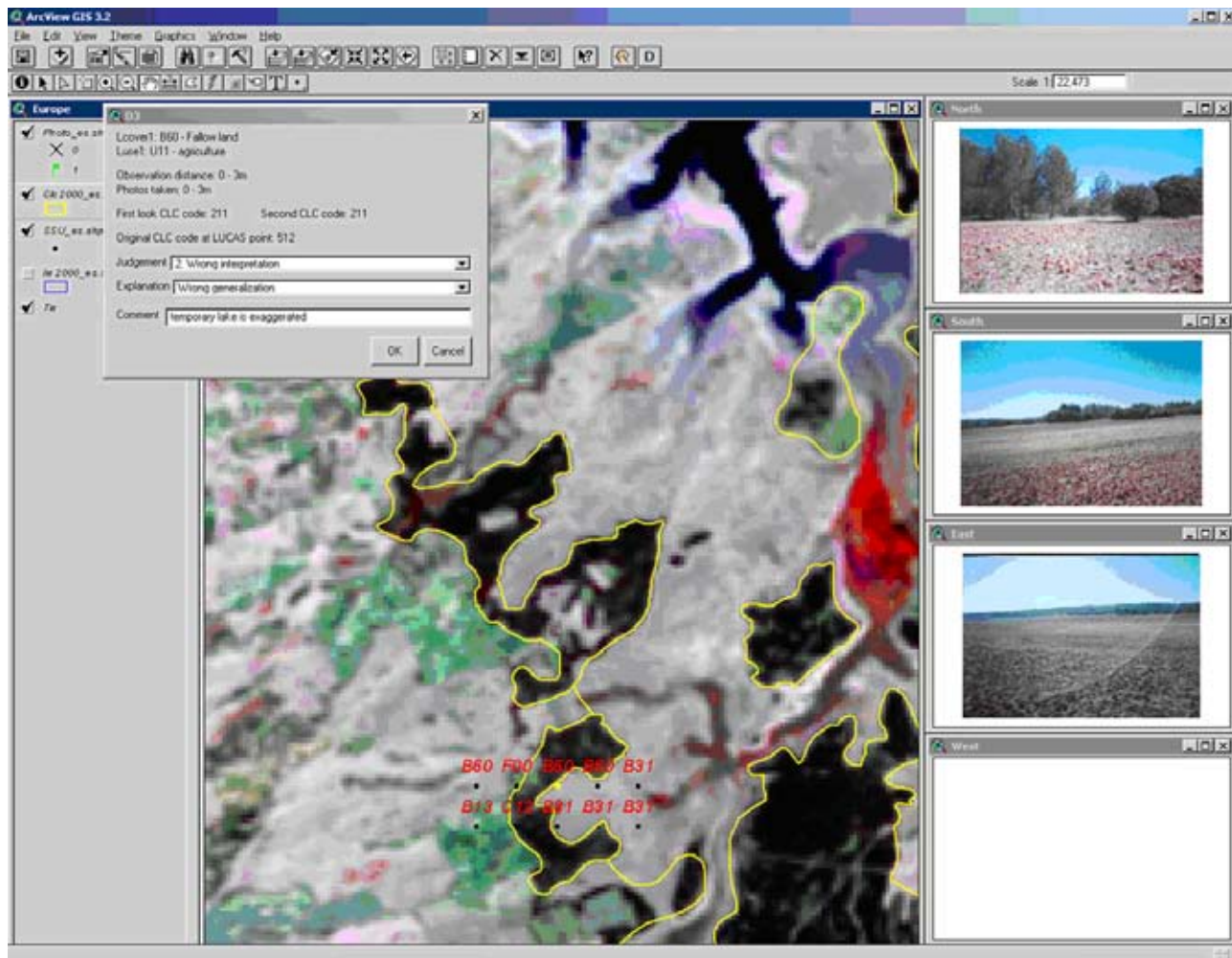
Illustration 26 Wrong interpretation/wrong code



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: agriculture;  
 Validation: 412, 412 (definitely mistyped 421);  
 CLC2000: 322.

Photos show peatland (NB: the typical white indicator plant — Eriophorum sp. — Cotton Grass). Wrong LUCAS LC code.

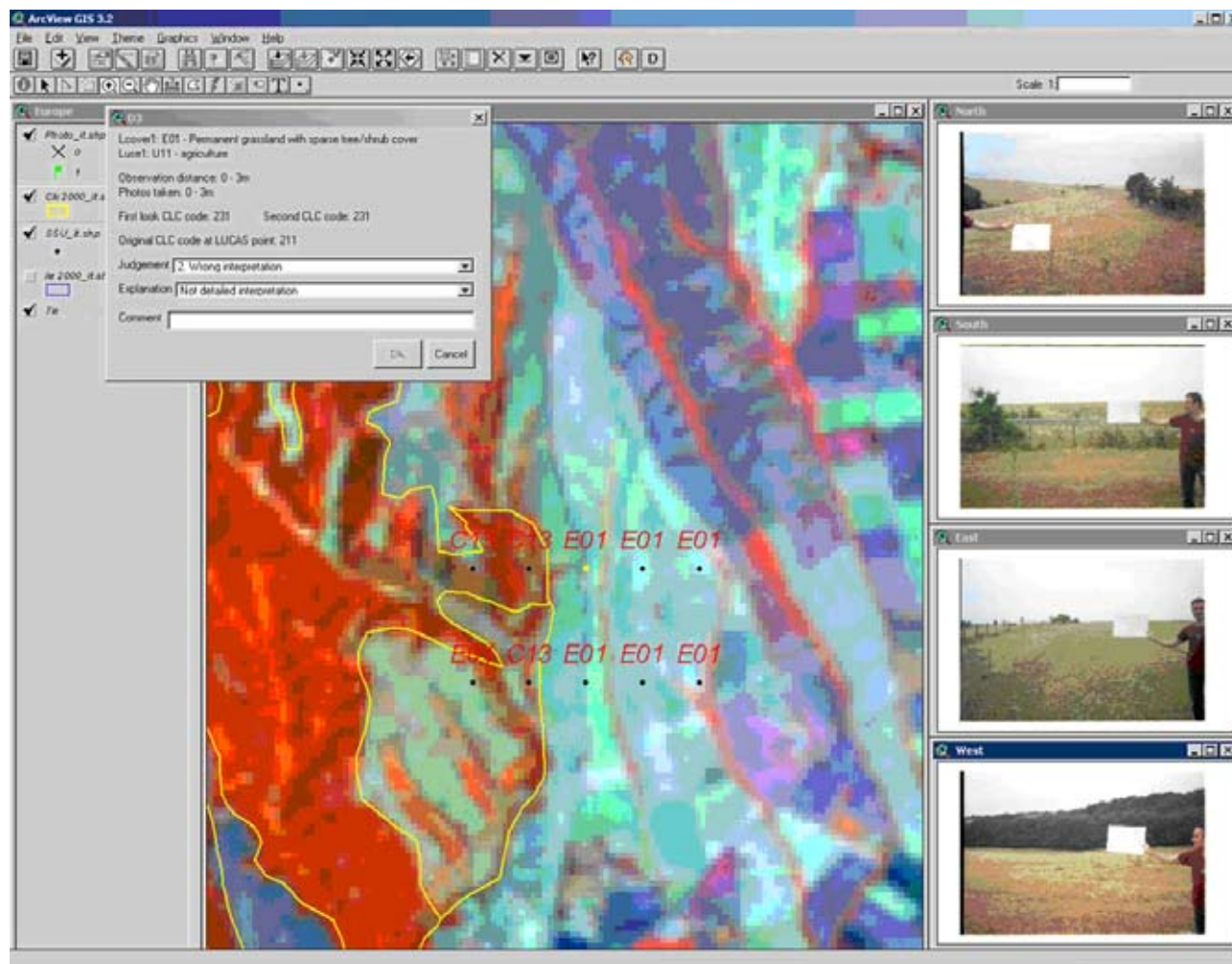
**Illustration 27 Wrong interpretation/wrong generalisation**



**Note:** LLC: fallow land;  
 LLU: agriculture;  
 Validation: 211, 211;  
 CLC2000: 512.

LUCAS photos clearly show agriculture. The reason for misinterpretation was the shallow lake (see NE corner of Image2000 extract). The interpreter considered a low water status for the lake based on the image taken in summer.

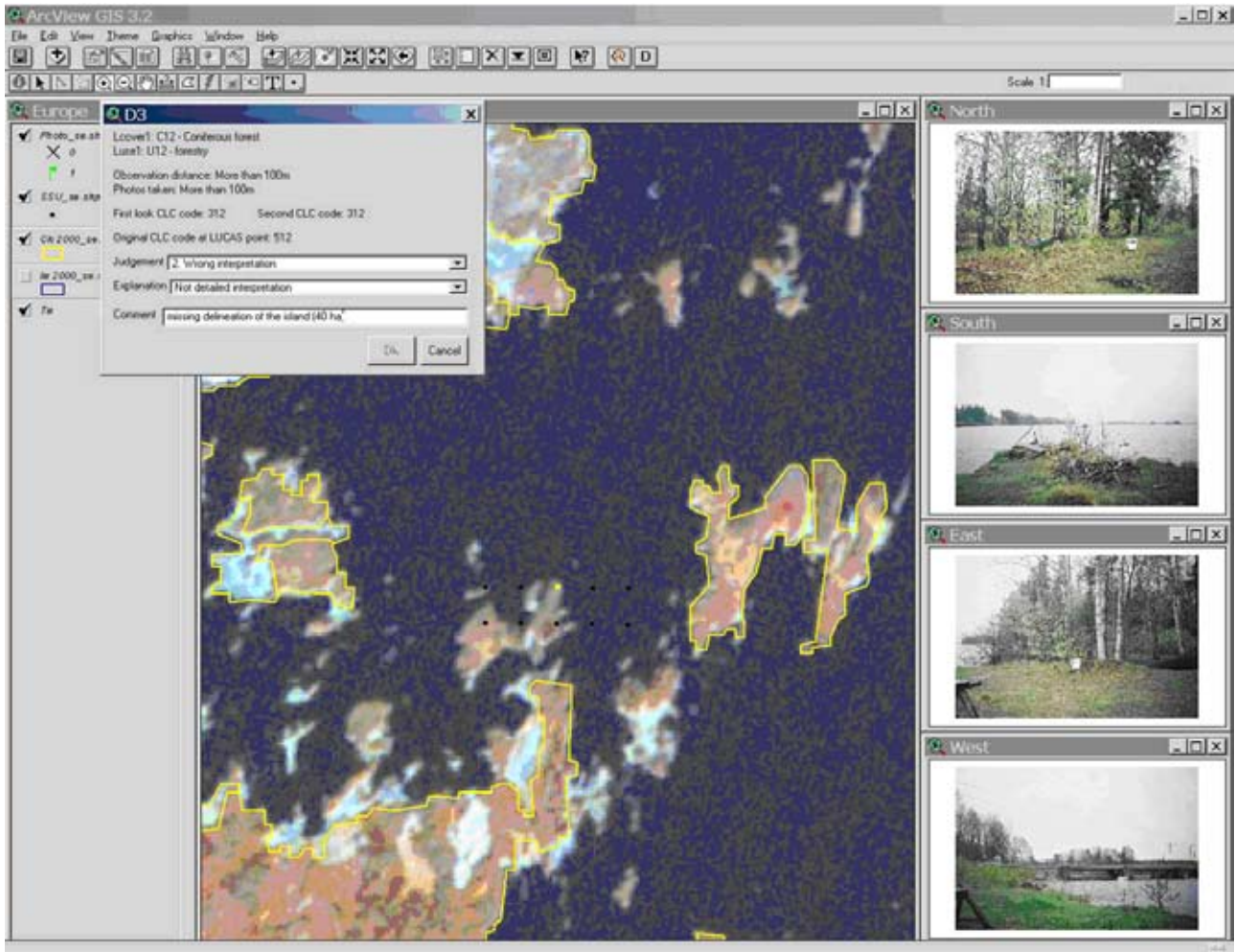
Illustration 28 Wrong interpretation/not detailed interpretation



**Note:** LLC: permanent grassland with sparse tree/shrub cover;  
 LLU: agriculture;  
 Validation: 231, 231;  
 CLC2000: 211.

LUCAS photos and six SSUs clearly show pasture. The interpreter omitted the separation 231 in the large 211.

**Illustration 29 Wrong interpretation/not detailed interpretation**

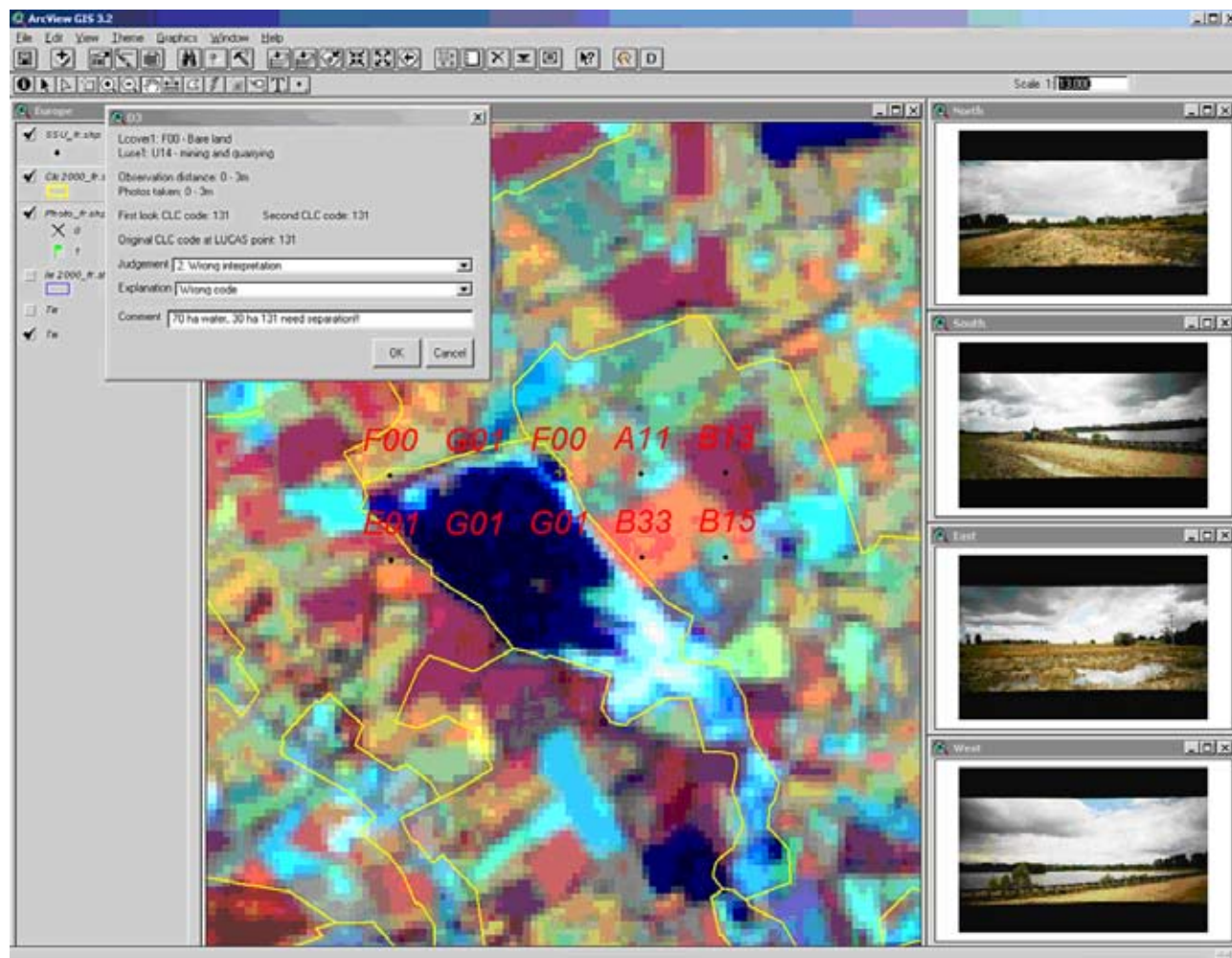


**Note:** LLC: coniferous forest;  
 LLU: forestry;  
 Validation: 312, 312;  
 CLC2000: 512.

Omitted separation of the island (size about 40 ha), covered with forest.



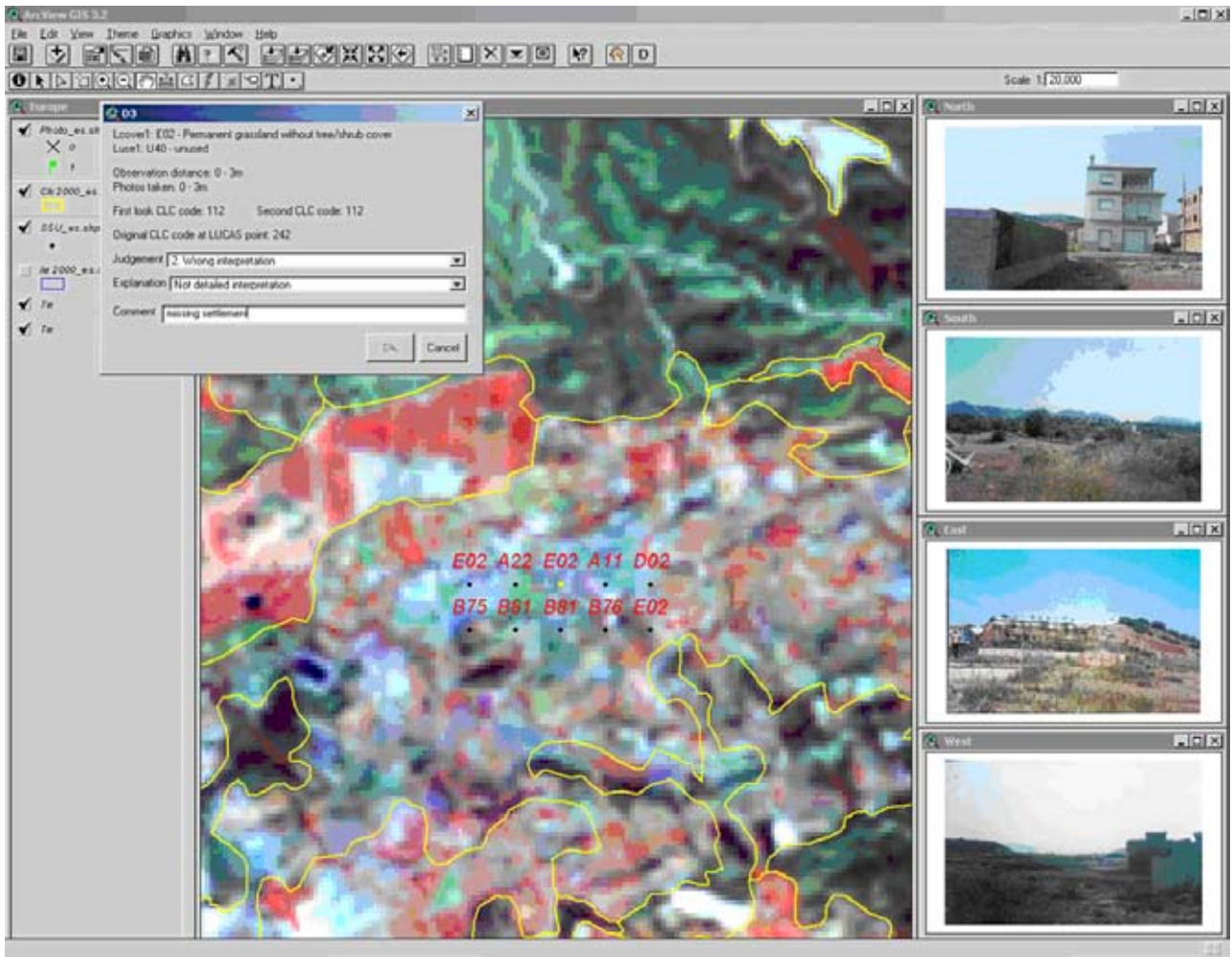
Illustration 30 Wrong interpretation/not detailed interpretation



**Note:** LLC: bare land;  
 LLU: mining and quarrying;  
 Validation: 131, 131;  
 CLC2000: 131.

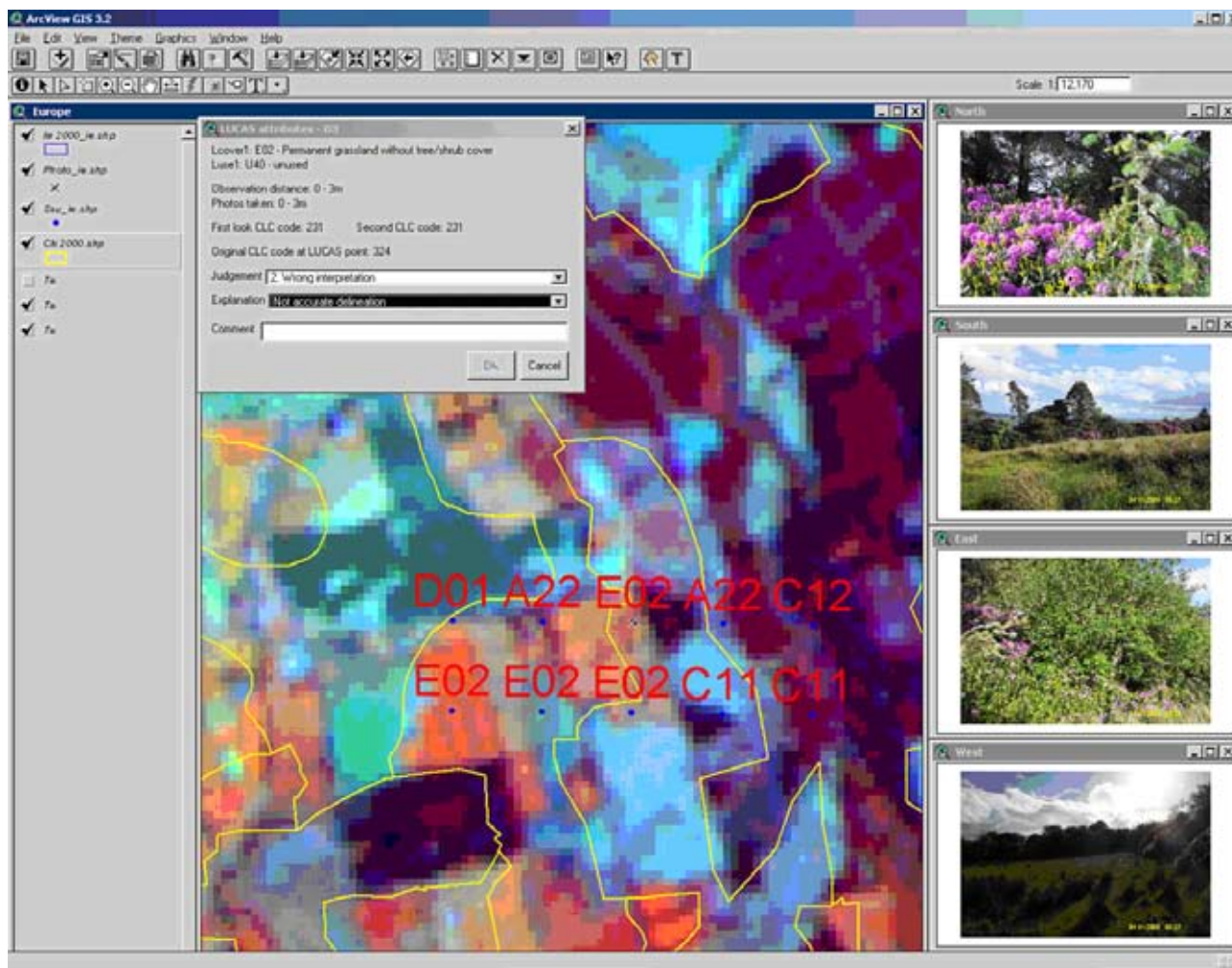
Even though the two validation codes were the same as the CLC2000 code, it was finally recognised that the 70 ha size polygon had to be split in two parts: 512 (the proper code around LUCAS points) and 131 (the rest of the polygon), where active mining was taking place.

**Illustration 31 Wrong interpretation/not detailed interpretation**



**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: unused;  
 Validation: 112, 112;  
 CLC2000: 242.

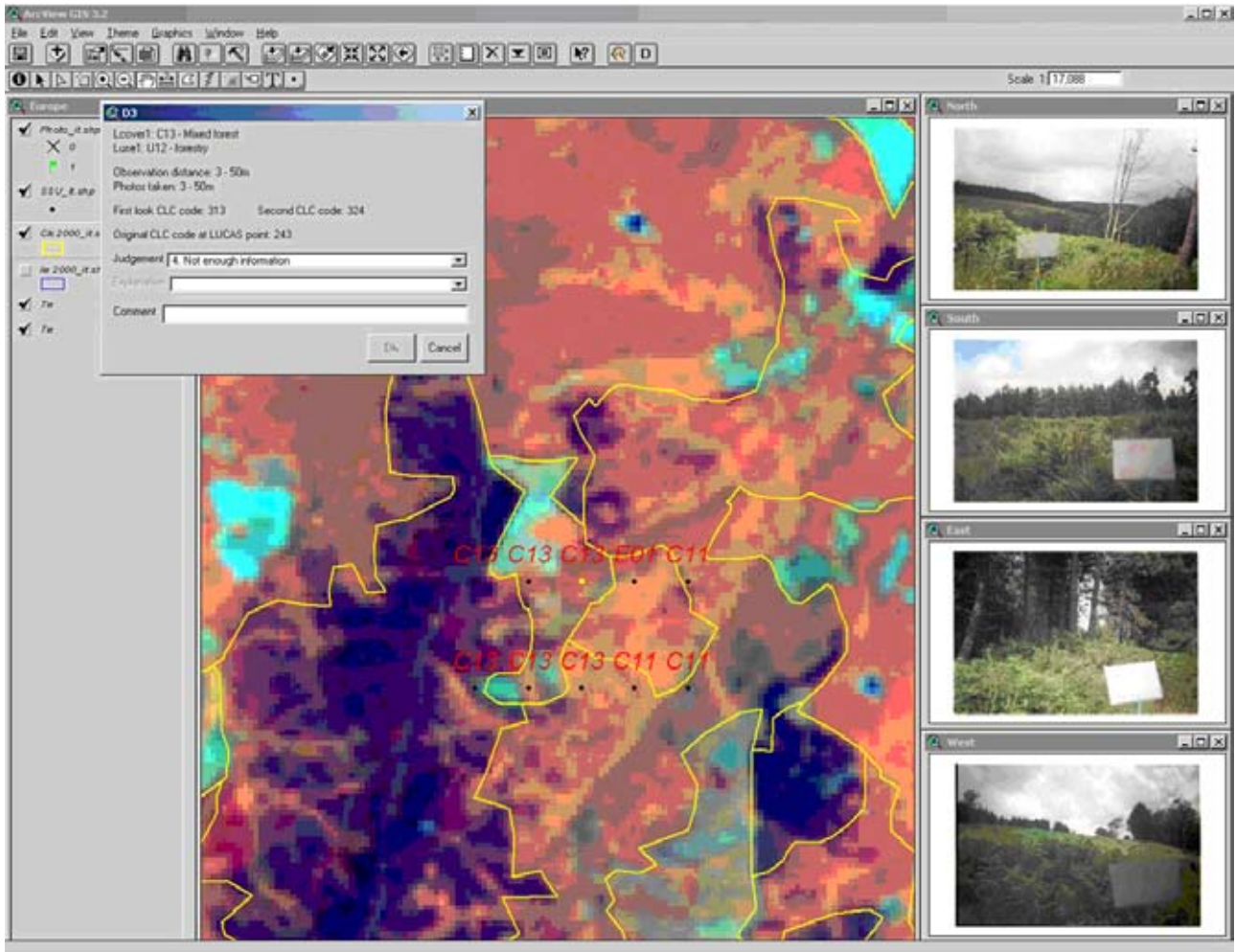
LUCAS photos and 2 of the SSUs (neighbours of the central SSU) show an omitted settlement (size around 25 ha).

**Illustration 32 Wrong interpretation/not accurate delineation**

**Note:** LLC: permanent grassland without tree/shrub cover;  
 LLU: unused;  
 Validation: 231, 231;  
 CLC2000: 324.

The polygon around the central LUCAS SSU is misplaced due to omitted geometric improvement of CLC90. Displacement exceeds 100 meters.

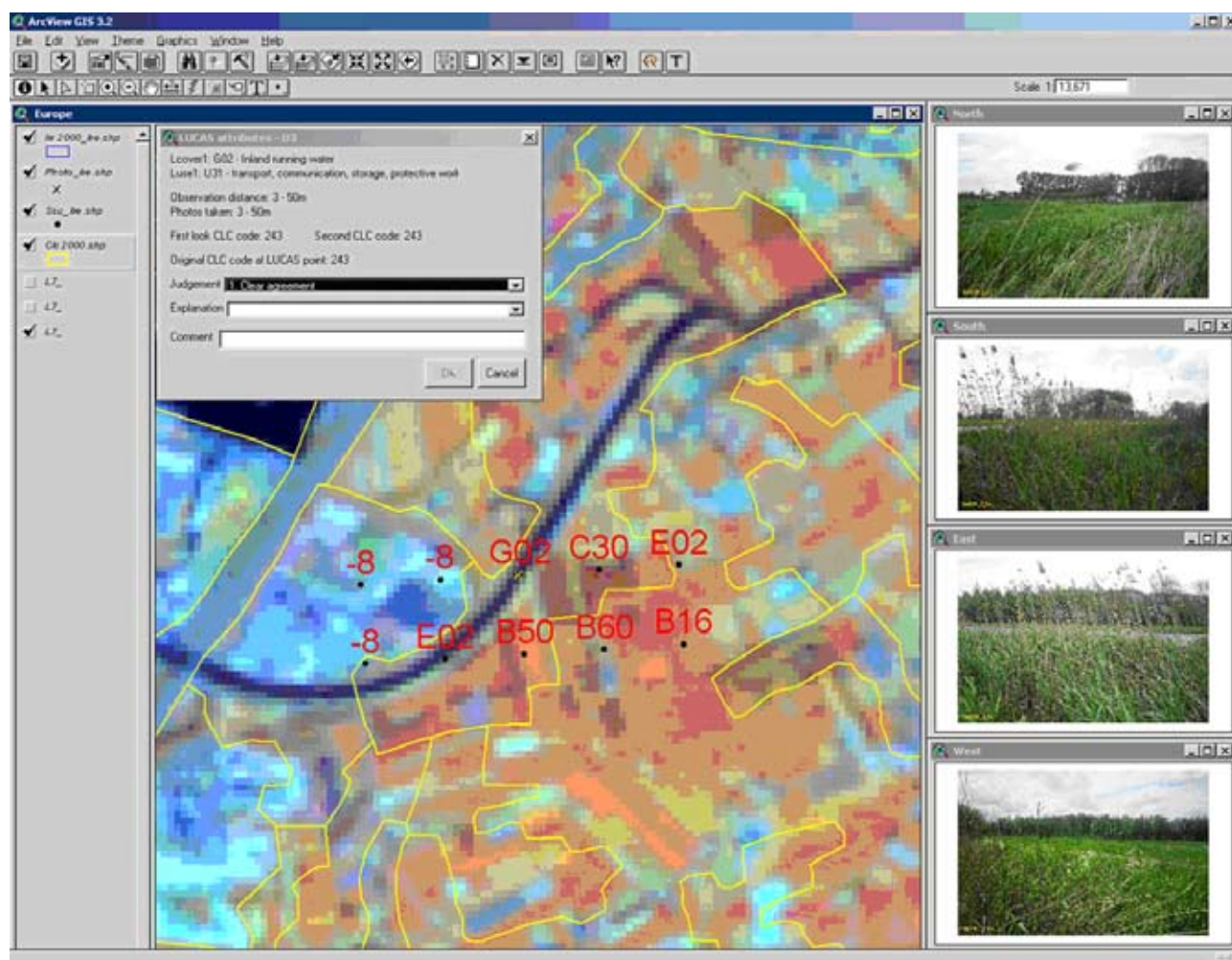
**Illustration 33 Not enough information**



**Note:** LLC: mixed forest;  
LLU: forestry;  
Validation: 313, 324;  
CLC2000: 243.

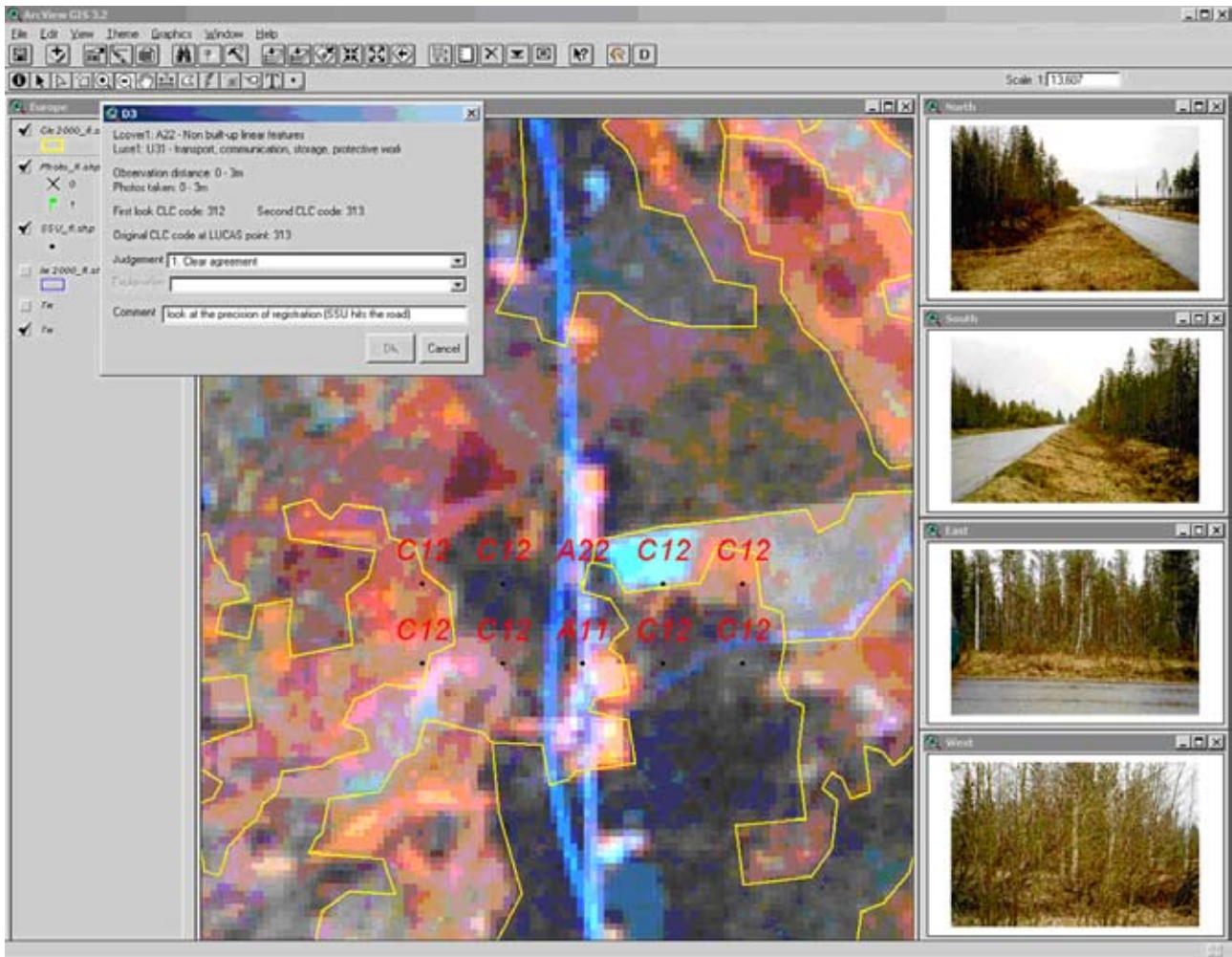
Based on LUCAS, it was not possible to decide whether the non-forested part of the polygon was agricultural land or clear cut.

**Illustration 34** Remarks on LUCAS data quality – Good geometrical match with Image2000 (Netherlands)



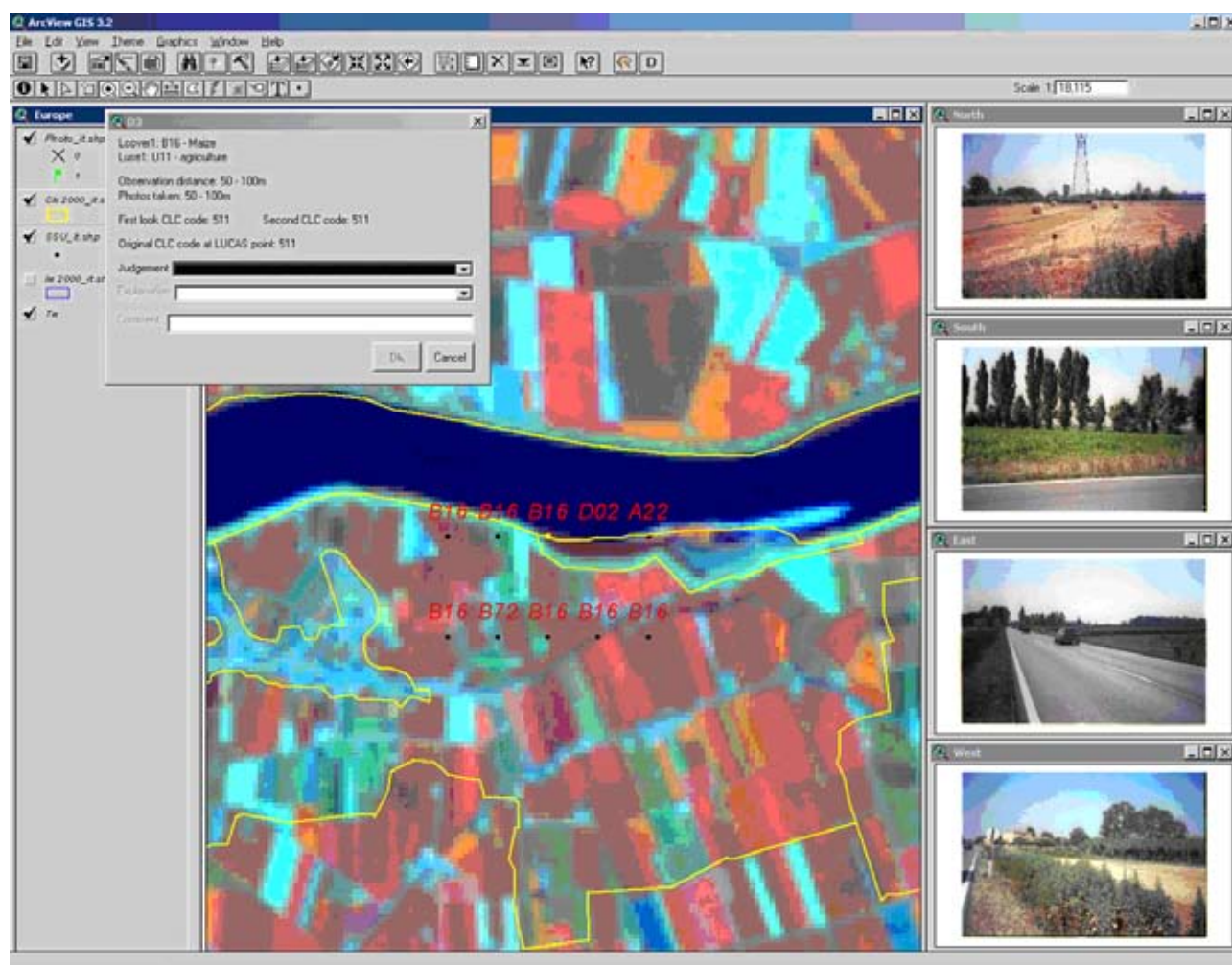
**Note:** Excellent match of LUCAS data (G02 = inland running water) on the small river shown by Image2000.

**Illustration 35** Remarks on LUCAS data quality – Good geometrical match with Image2000 (Sweden)



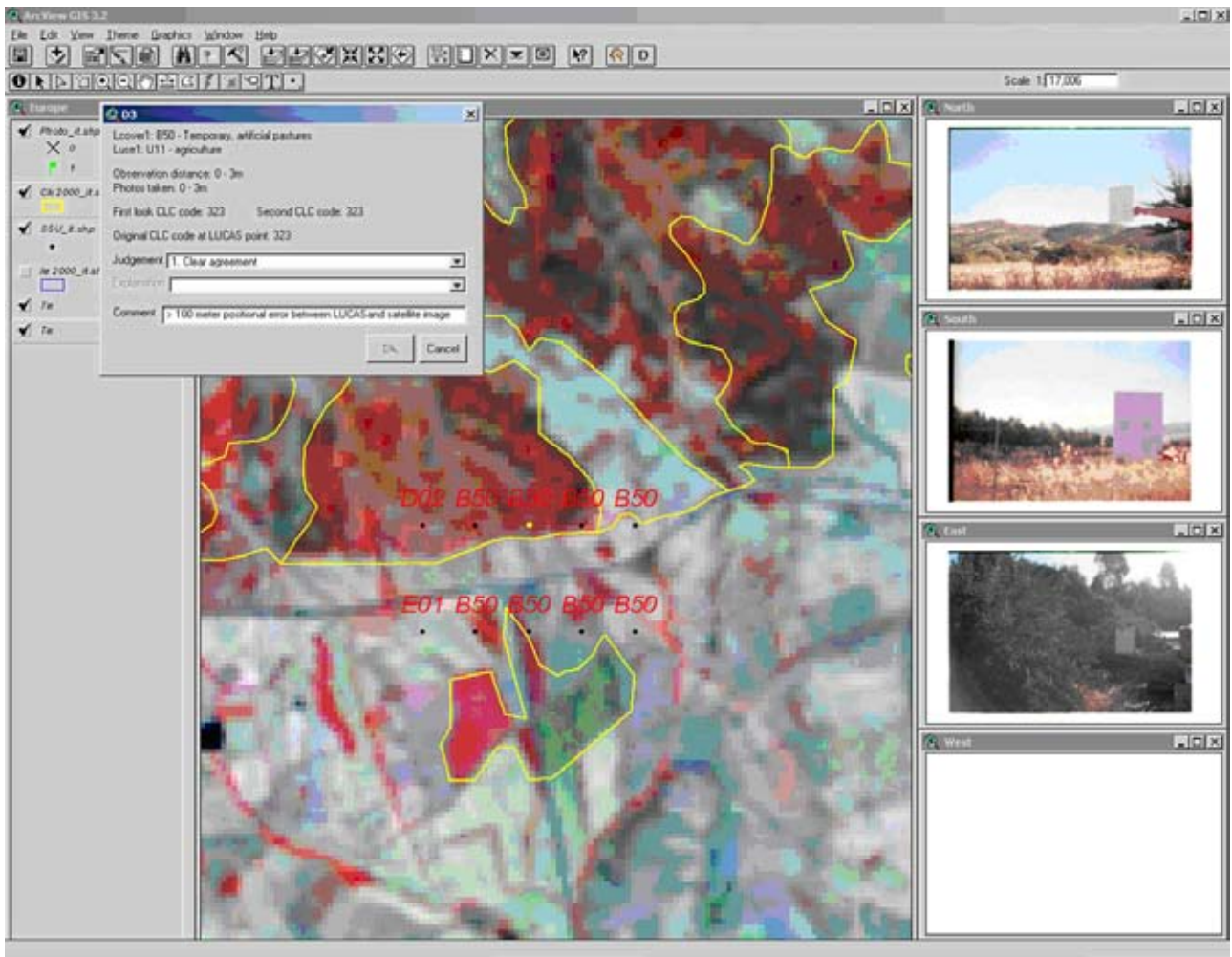
**Note:** Excellent match of LUCAS data (A22 = non built-up linear feature) on the road shown by Image2000.

**Illustration 36** Remarks on LUCAS data quality – bad geometrical match with Image2000 (Italy)



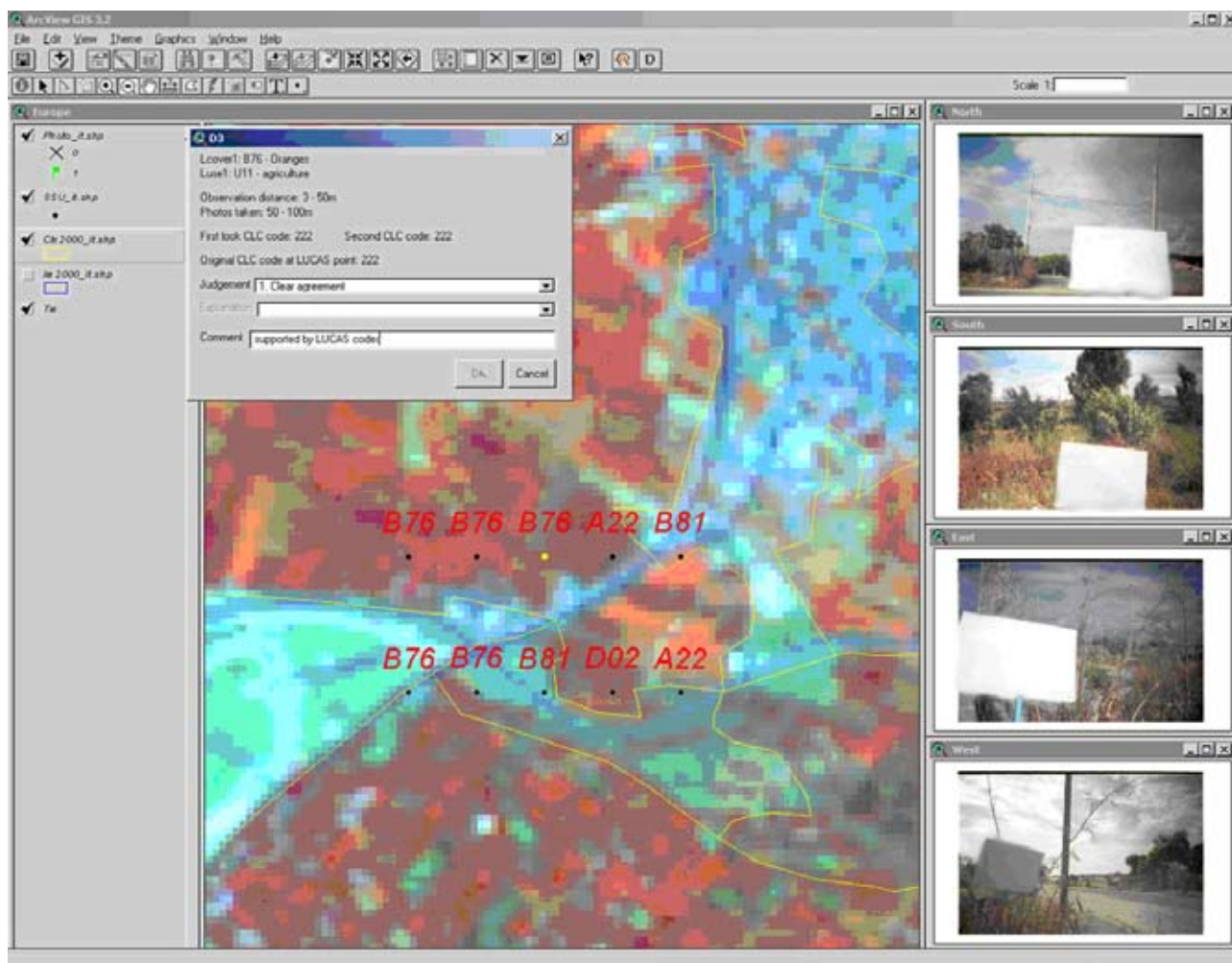
**Note:** Inconsistency in the geometry of the localisation. The river indicated by Image2000 is not shown by LUCAS.

**Illustration 37** Remarks on LUCAS data quality – bad geometrical match with Image2000 (Italy/Sardegna island)



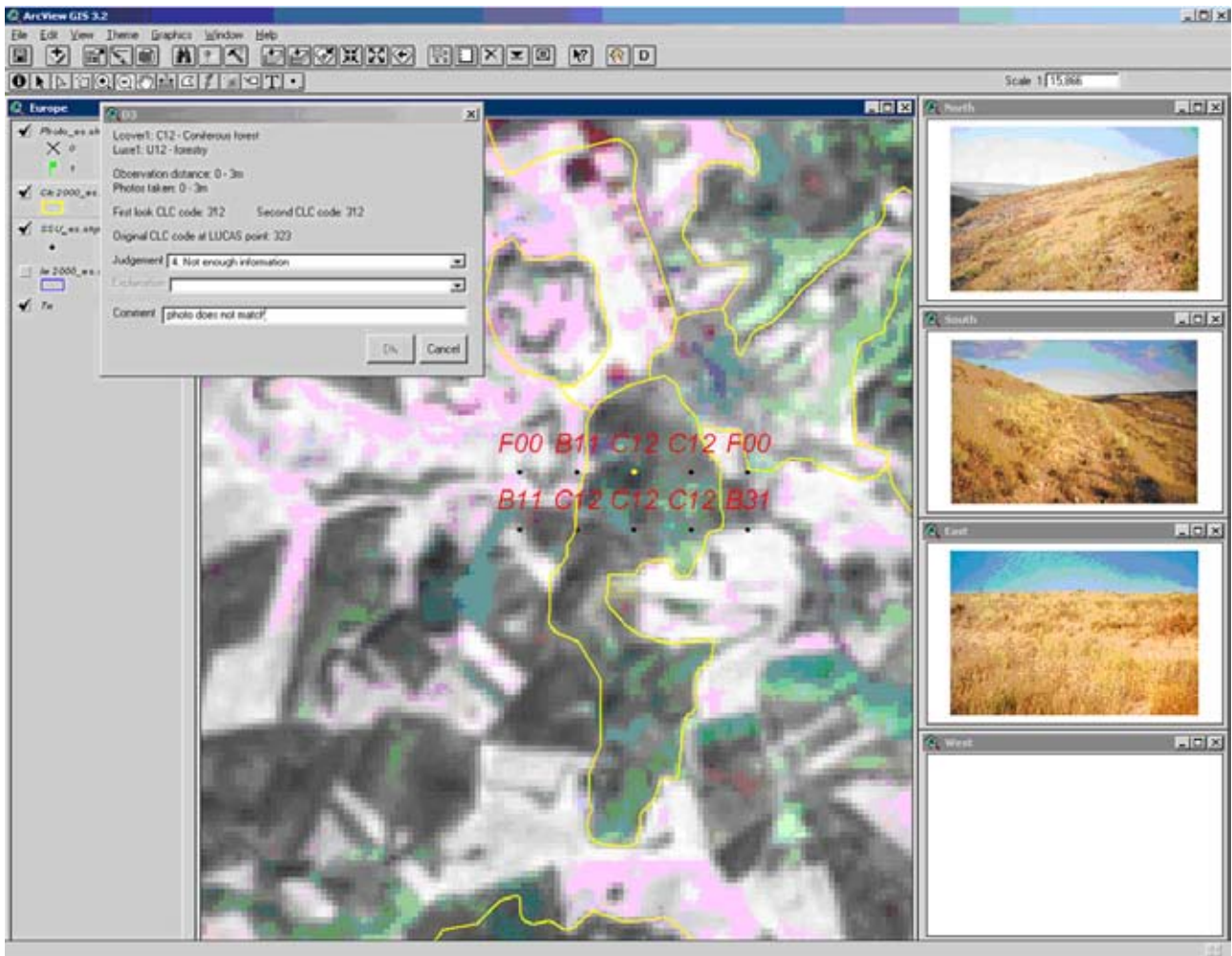
**Note:** Inconsistency in the geometry of the localisation. Temporary artificial pasture is shown by LUCAS inside natural vegetation shown by Image2000.



**Illustration 38** Remarks on LUCAS data quality – bad photo quality

**Note:** This example shows a case when decision is supported by LUCAS LC codes attached to SSUs, rather than bad quality photos.

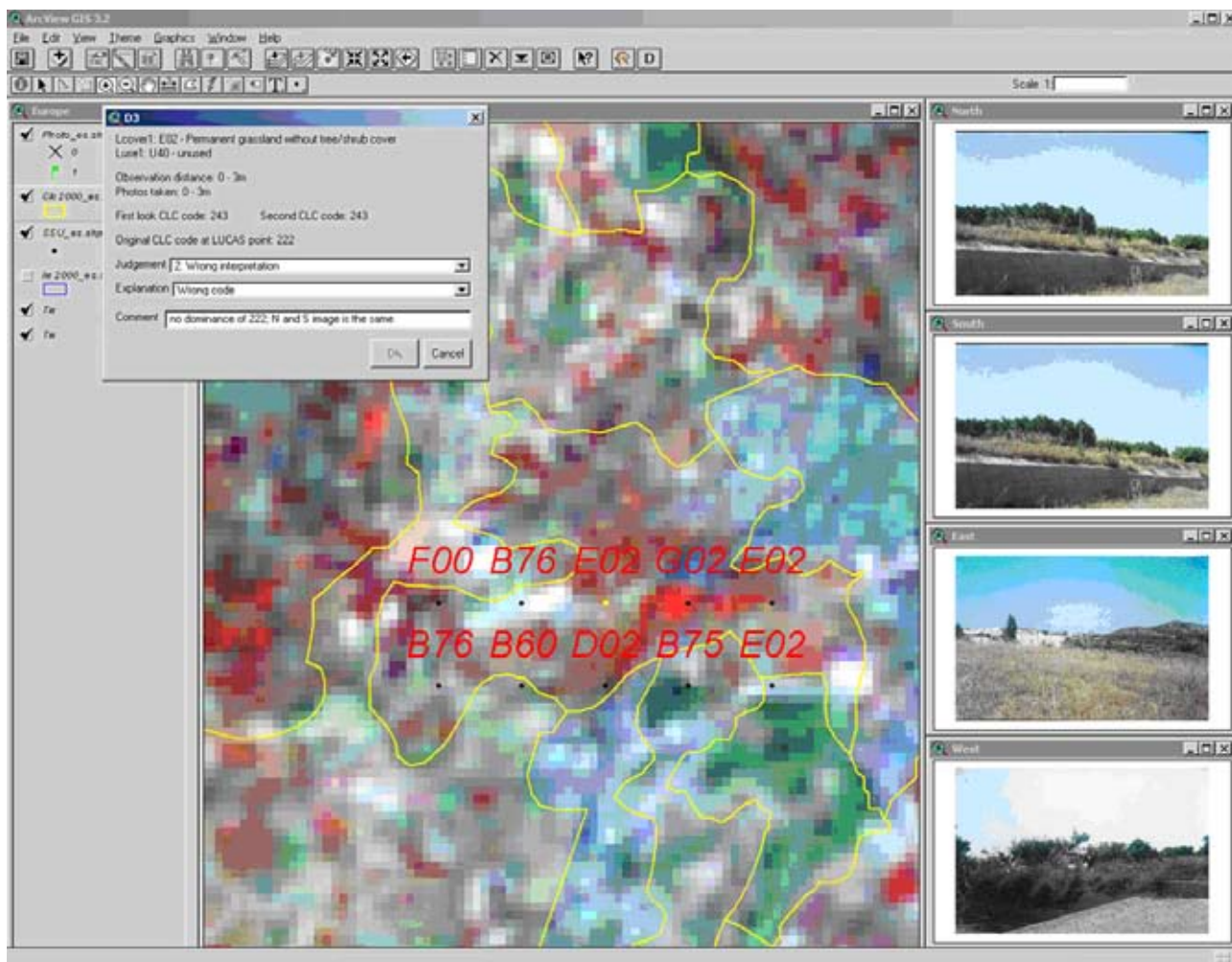
**Illustration 39** Remarks on LUCAS data quality – photos do not refer to the site



**Note:** LLC: coniferous forest;  
 LLU: forestry;  
 Validation: 312, 312;  
 CLC2000: 323.

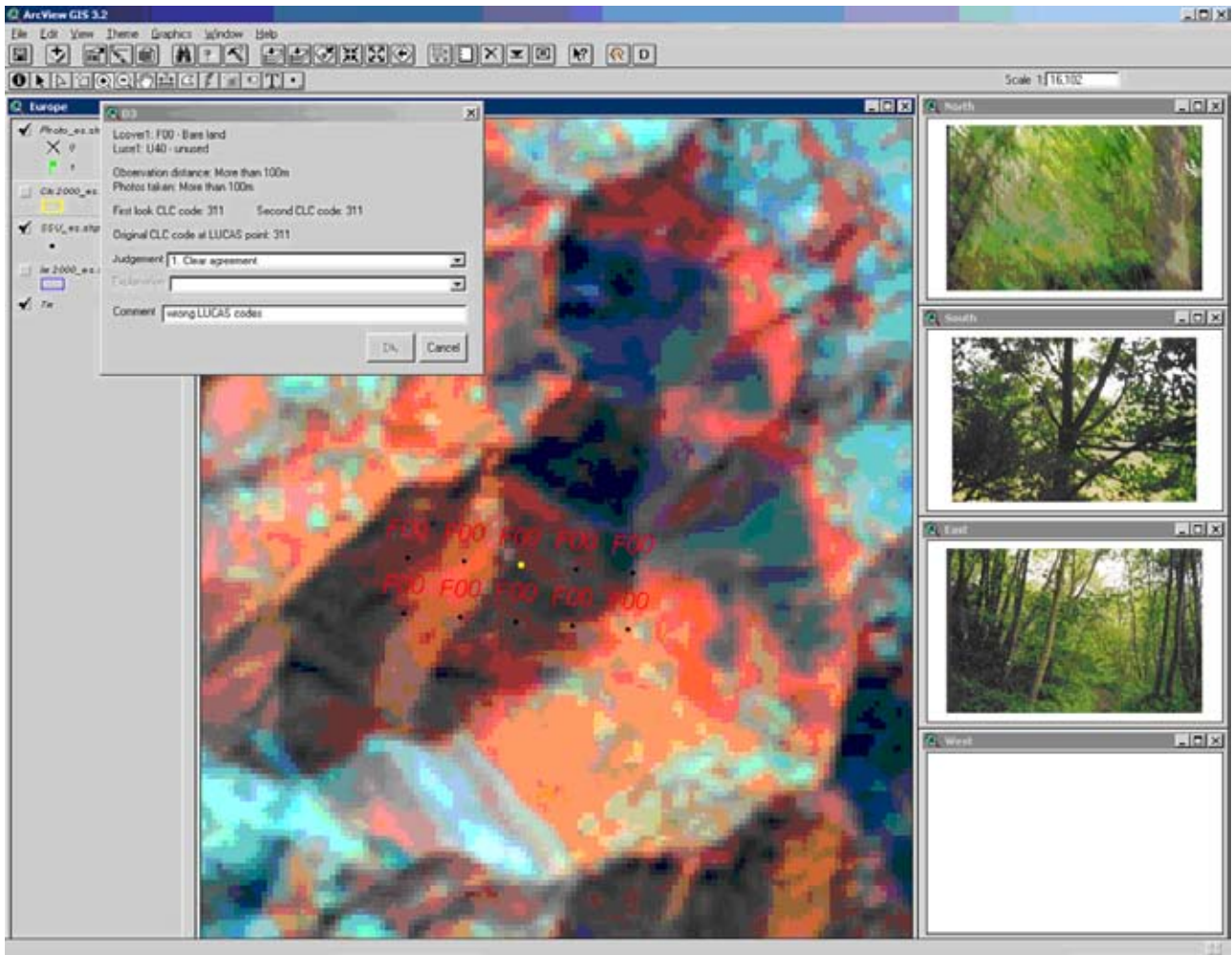
Image2000 and LUCAS LC/LU codes match. LUCAS photographs have no connection to this site.

**Illustration 40** Remarks on LUCAS data quality – direction/orientation of photos



**Note:** North and South photos are the same.

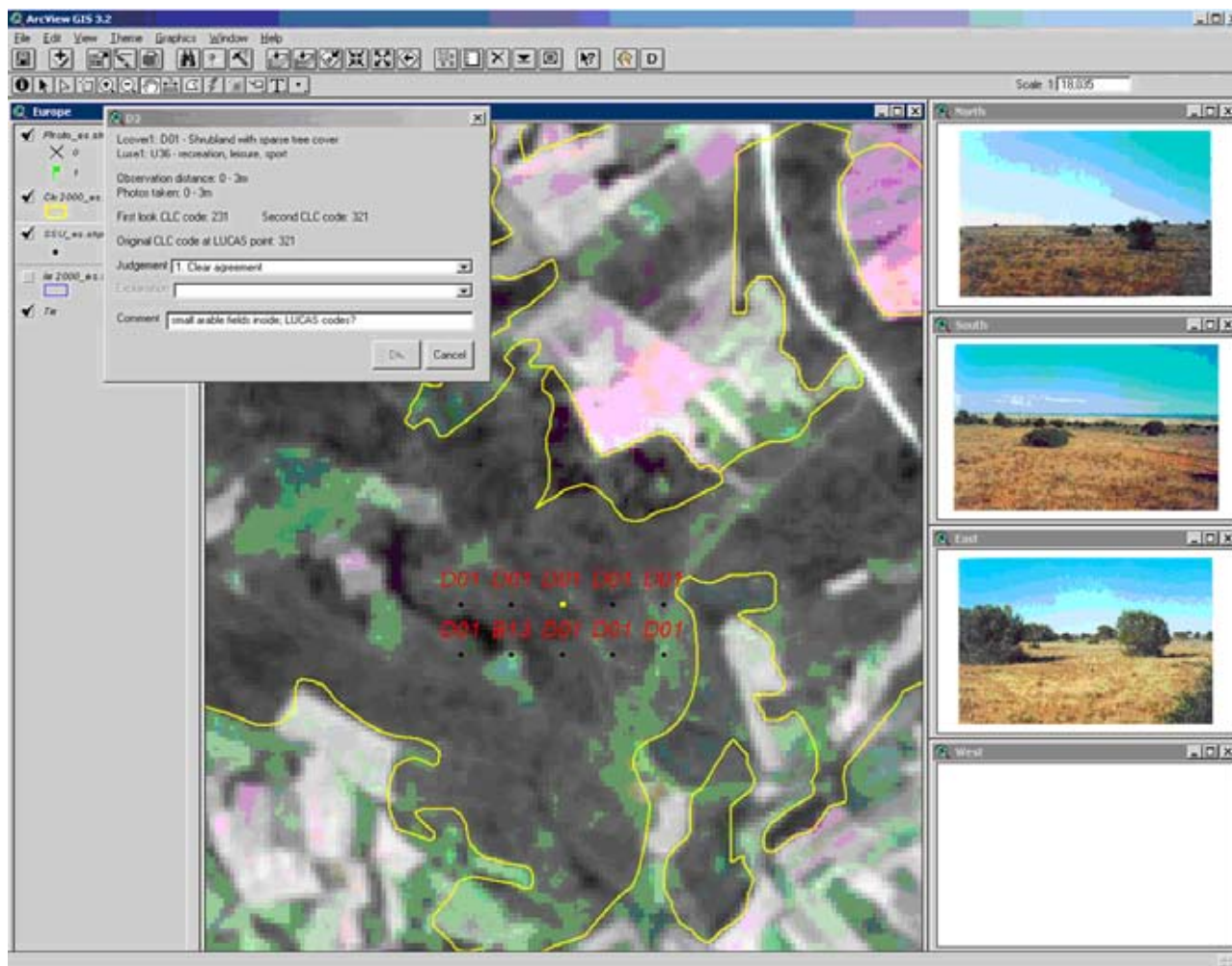
**Illustration 41** Remarks on LUCAS data quality – bad quality photos, improper LUCAS codes



**Note:** LLC: bare land (all SSUs);  
 LLU: unused (all SSUs);  
 Validation: 311, 311;  
 CLC2000: 311 (rather dense vegetation).

LUCAS LU/LC codes prove counter to what could be derived from Image2000. LUCAS LU/LC codes were neglected from the analysis of this example.

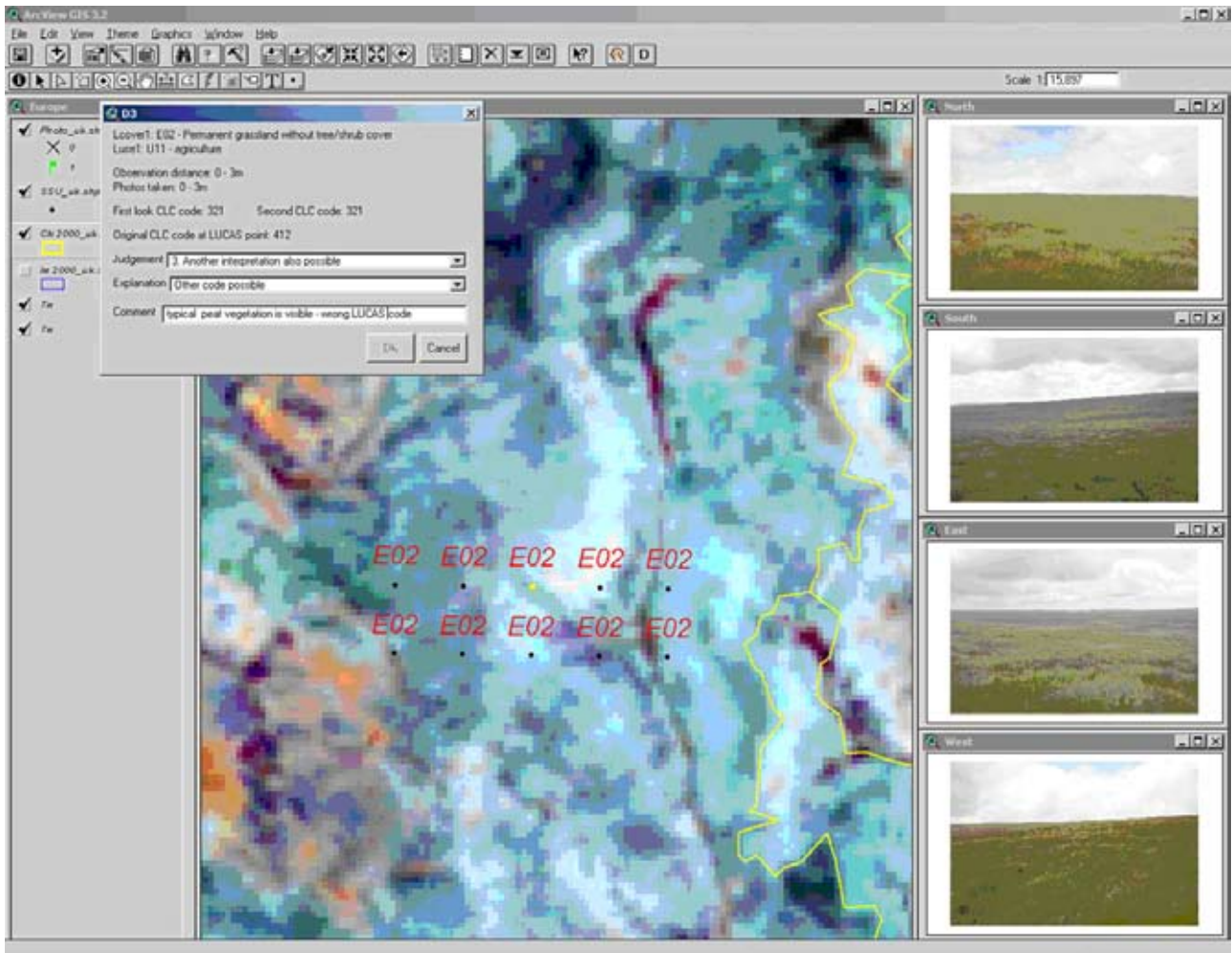
Illustration 42 Remarks on LUCAS data quality – improper LUCAS codes



**Note:** LLC: shrubland with sparse tree cover;  
 LLU: recreation, leisure, sport;  
 Validation: 231, 321;  
 CLC2000: 231.

Grassland is confirmed by photos. LUCAS LU codes are not clear; land use, in particular, is difficult to understand (hunting?).

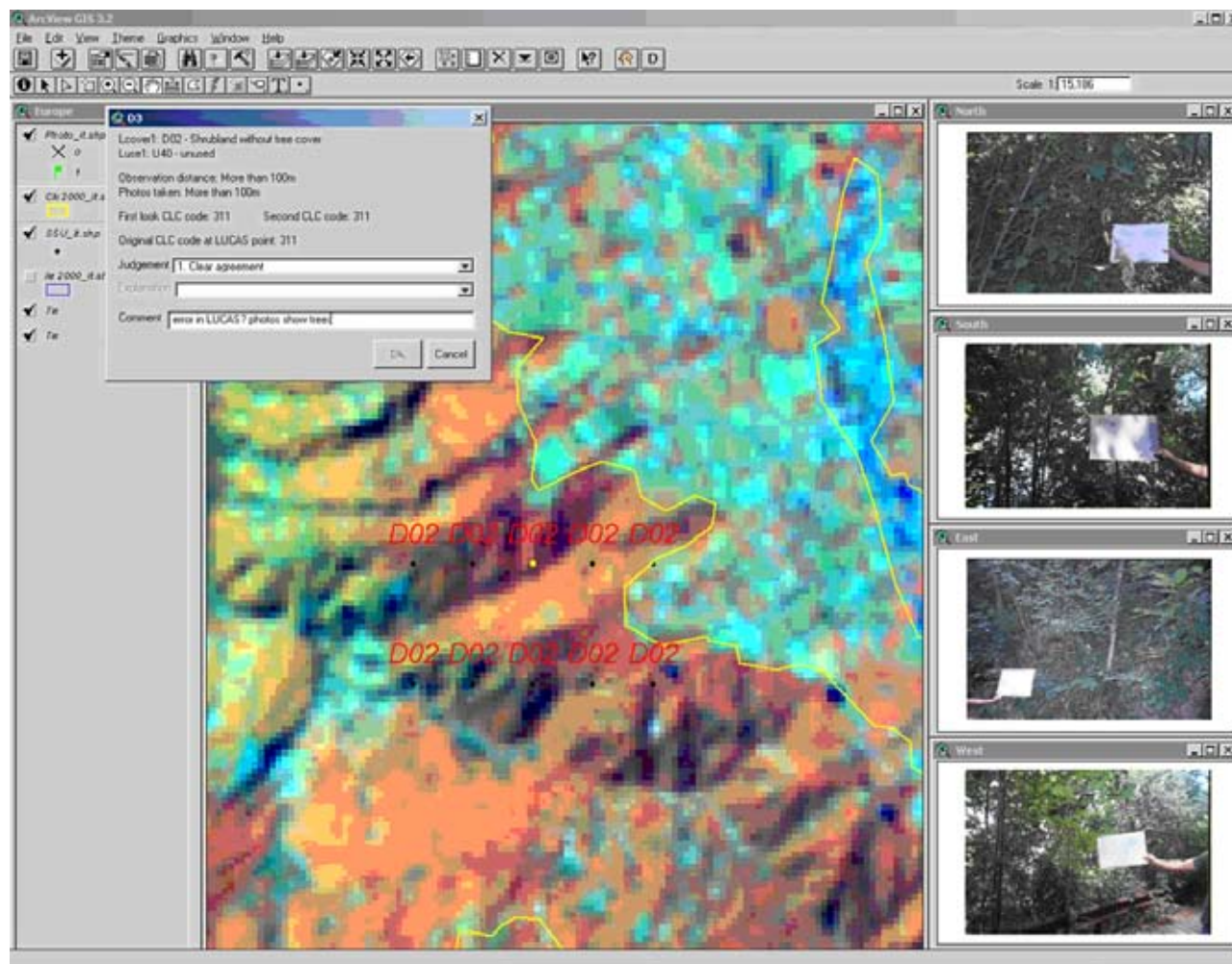
**Illustration 43** Remarks on LUCAS data quality – improper LUCAS codes



**Note:** LLC: permanent grassland without tree/shrub cover (all SSUs);  
 LLU: agriculture;  
 Validation: 321, 321;  
 CLC2000: 412.

Photos show grass-like vegetation with the typical white indicator plant – Cotton Grass – of peatland. LUCAS LC code is wrong as the surveyor did not recognise the peatland.

**Illustration 44** Remarks on LUCAS data quality – improper LUCAS codes



**Note:** LLC: shrubland with sparse tree cover (all SSUs);  
 LLU: unused;  
 Validation: 311, 311;  
 CLC2000: 311.  
 Photos show trees, Image2000 shows forest. LUCAS LC codes are questionable.





European Environment Agency

**The thematic accuracy of Corine land cover 2000  
Assessment using LUCAS (land use/cover area  
frame statistical survey)**

2006 — 85 pp. — 21 x 29.7 cm

ISBN 92-9167-844-9

European Environment Agency  
Kongens Nytorv 6  
1050 Copenhagen K  
Denmark

Tel.: +45 33 36 71 00  
Fax: +45 33 36 71 99

Web: [www.eea.europa.eu](http://www.eea.europa.eu)  
Enquiries: [www.eea.europa.eu/enquiries](http://www.eea.europa.eu/enquiries)



Publications Office

