Comparing statistical and semantic approaches for identifying change from land cover datasets

Alexis Comber^{1*}, Peter Fisher¹, Richard Wadsworth²

¹ Department of Geography,

University of Leicester,

Leicester,

LE1 7RH, UK

Tel: +44 (0)116 252 3859,

Fax: +44 (0)116 252 3854,

email: ajc36@le.ac.uk

² Centre for Ecology and Hydrology,

Monks Wood, Abbots Ripton,

Huntingdon,

Cambridgeshire,

PE28 2LS, UK

* Corresponding author

Abstract

In this paper we examine methods for integrating spatial data which apparently should be comparable because they are of the same data type or theme, but which are incompatible or discordant because the classes of that theme are different. For a variety of reasons including changes in methods, in understanding of the resource, and in policy initiatives in the commissioning of the survey, this problem is widespread in the results of natural resources surveys. We present two generic methods: one method is grounded in a statistical approach using discriminant analysis, and the other exploits the knowledge of experts. We use the context of land cover mapping of Great Britain to explore these approaches for integrating discordant data. We demonstrate that the expert-based approach gives very good levels of identification of locations with incompatible classifications at the different times, and gives a much better rate of recognition of change. Some conclusions are made about the need to expand current metadata and data quality reporting to include descriptions of:

- data conceptualisations, semantic and ontologies;
- who decided and defined what the features of interest in a dataset are, and why;

If the benefits of spatial data initiatives such as GRID, E-science and INSPIRE are to be fully realised then some method needs to be found to communicate that information most effectively to the potential user of the data.

Key words: land cover, expert, integration, metadata

1. Introduction

The problem of data dissimilarity (discord, ontological inconsistency, variation, etc) has long been recognised both in the GIS and environmental disciplines. This is seen in the different classifications and land cover maps of CORINE (European Environment Agency, 2001), the USGS Land Use and Land Cover classification (Anderson *et al.*, 1976) and land cover classifications used in the different land cover maps of Great Britain (Fuller *et al.*, 2003; Comber *et al.*, 2003a). Spatial data integration is a necessary step for environmental model development, predicting the rates and locales of change and for initiatives that aim to respond to environmental change. Typically analysis is conducted by intersecting data to generate measures of class-to-class correspondence that are then used as probabilities in Bayesian analyses or membership functions in fuzzy approaches (e.g. Tappeiner *et al.*, 1998; Clerici *et al.*, 2002; Woodcock and Gopal, 2000). These allow multiple realisations to be generated of the phenomenon of interest by Monte Carlo simulation or Markov chains. However, correspondences take no account of the different conceptualisations embodied in different datasets.

Implicit in modelling activities is an understanding of the endemic issue of data discord: different datasets that purport to report the same environmental features may do so in different ways. Geographical information (GI) is an *interpretation* of reality and discord (or ontological variation) among datasets occurs because the same phenomenon can be recorded using different conceptualisations that may include using various objects, labels, and definitions. To date the majority of repeat natural resource inventories are different because of different methodologies, technological developments, changes in policy (as documented by Comber *et al.*, 2002; 2003a), or advances in scientific understanding. The result is inconsistency or discord between datasets recording similar environmental processes. This is illustrated in this paper using the UK land cover mappings of 1990 and 2000 as an example.

The current importance of the issue of data discordance is highlighted by the changing context within which spatial data integration activities are taking place. Current computing infrastructure initiatives such as spatial data infrastructures (SDIs) aim to increase the availability of third party data to distributed users in order to expand data access for both social and scientific objectives. For example the EU INSPIRE project seeks to make available “*relevant, harmonised and quality geographic information to support formulation monitoring*

and evaluation of Community Policies” (<http://inspire.jrc.it/home.html>). The development of the computing Grid is providing “*pervasive, dependable, consistent and inexpensive access to advanced computational capabilities, databases, sensors and people*” (<http://www.escience-grid.org.uk>). Laudable as the E-science initiatives are, their origins are in the disciplines of computing science and database management, and can be caricatured as considering only the database object. Differences in the way that similar processes may be conceptualised by different groups are not addressed by the stated objectives of SDIs nor by current meta-data or data quality reporting paradigms. This issue is given increased weight by the large numbers of GI users, brought into this arena by the increasing ease of digital data transfer, SDIs and an acknowledgement of the spatial component of much data. However, as documented by Fisher (2003), many users of GI are interested only in acquiring the (digital) map and not understanding the wider *meaning* of the data.

In this paper we introduce the UK land cover mappings of 1990 and 2000 (~2) and describe statistical and expert methods for data integration (~3). The results of applying these approaches to the land cover maps are presented (~4) before discussing the issues highlighted by this work (~5) and drawing some conclusions (~6). It is noteworthy that whilst the differences in the UK land cover maps will be of interest to a predominantly UK audience, the issues being addressed are relevant to a wider context of increased ease of access to digital spatial data and the objective of data integration, interoperability and standardisation.

2. Background: UK land cover mapping

The UK Land Cover Map 2000 (LCM2000) was described as an ‘upgrade’ to its predecessor the 1990 Land Cover Map of Great Britain (LCMGB) by Fuller *et al.*, (2002). There are considerable structural and thematic differences between the datasets and a full specification is given by Fuller *et al.* (2003). In brief, LCMGB is a raster dataset derived from a by-pixel supervised classification of composite (i.e. combined) winter and summer satellite imagery captured around 1990. It identified 25 ‘Target’ classes based on their spectral properties (Fuller *et al.*, 1994). LCM2000 is a vector dataset and records parcels of land cover identified by applying a segmentation algorithm to composite satellite imagery from around 1998. The central core of each parcel was then classified into one of 26 policy-related ‘Broad Habitat’ classes based on the average (mean) reflectance of the parcel constituent pixels. LCM2000 includes extensive object-level (i.e. parcel-level) meta-data (Smith and Fuller, 2002) describing parcel processing history and spectral heterogeneity.

One of the ways to understand the embedded differences between environmental data as determined by what is measured (the epistemology) and how different concepts are specified (the ontology) and recorded is to consider the context within which data are commissioned. Comber *et al.* (2003a) described the different commissioning contexts of LCMGB and LCM2000 using networks of actors. A network is described in terms of nodes or actors and their linkages. Thus an actor is any entity that interacts with other actors or serves as an intermediary between actors. These may be humans or objects since all interactions between humans can be seen as being mediated through objects of one type or another (Law, 1992). Using this analysis Comber *et al.* (2003a) showed the relative complexity of the network involved in the creation of LCM2000 compared to LCMGB:

- LCMGB had a small number of important actors contributing different types of influence as input to the production team: money came from the British National Space Centre and the Department of the Environment, information came from satellite imagery and associated training data and skills came from the remote sensing community. The output was validated by comparison with the field component of the Countryside Survey and the output (the land cover map) was then passed to the user community.
- LCM2000 had a much wider set of important actors influencing the production of the survey as input. These included most of those for LCMGB with the addition of policy and its influence articulated through the user community (environmental agencies). The tensions between the policy objectives and what was scientifically feasible were played out through a steering group with some control over the methodology, another new actor.

LCMGB was a demonstration of the utility of satellite imagery in environmental monitoring and LCM2000 was a survey that was to fulfil national and international policy obligations.

These issues are further exemplified in the classification schemes of the two surveys. In LCMGB the target classes were selected by scientists on the basis of what they knew to be of interest and what they could reliably separate spectrally. In LCM2000 the broad habitat classification of LCM2000 had to record features of interest (priority habitats) that were relevant to the users charged with the responsibility for environmental management. Unfortunately not all of these were spectrally distinct and so LCM2000 used ancillary data to identify specific habitats: a soil acidity mask for 'Acid Grassland', 'Neutral Grassland' and 'Calcareous

Grassland’, and a peat map to identify areas of ‘Bog’. Thus the two surveys embody very different conceptualisations of the landscape, not only in terms of data structures (raster vs. parcel), but also in terms of the classification and the different objectives of each survey (policy vs. science). Because of these issues (although some of them were not explicitly acknowledged), LCM2000 was released with a warning against comparing directly it with LCMGB, yet for many users these data are the only national land cover datasets available.

The impact of the differences in conceptualisation is illustrated by the class of ‘Bog’. Table 1 contains the different definitions and quantities of bog from LCMGB and LCM2000 in a 100km x 100km area in the UK (Ordnance Survey square SK). It is evident that they are conceived in very different ways: LCMGB by standing water and LCM2000 by peat depth. An uninformed user comparing the two different datasets might conclude that in this area there has been considerable increase in bog over the 10 years, despite the inertia of bog plant communities. So there is a situation where the datasets are nominally very similar (both map land cover and refer to ‘Bog’), but they are difficult to compare or to integrate because they embody different views of the landscape. The extent of this issue cannot be overstated in natural resource inventory: the same problems exist for most land cover classes (for example, see Lund (2004) for the different ways that forest is conceived around the world).

[Insert Table 1 about here]

3. Methods

The LCM2000 and LCMGB land cover data was for a 100km by 100km square corresponding to Ordnance Survey National Grid square SK in the UK. The datasets were intersected such that each LCM2000 parcel had attributes recording the number and type of LCMGB pixels contained within the parcel area. That is, the LCM2000 parcel was used as a footprint to extract the coincident 1990 pixels from LCMGB. This generated a database containing a record for each of the LCM2000 parcels. Each record had 26 attribute fields describing the 2000 broad habitat class and each of the 25 1990 target classes. The objective was to compare the performance of statistical and expert-based approaches in detecting actual changes identified during field survey. Each approach used the number and type of the 1990 pixels to determine whether the class allocated to the parcel in 2000 represented a change or not.

3.1 Discriminant analysis approach (statistical)

Discriminant analysis builds a predictive model of group membership based on the characteristics of each record (i.e. each LCM2000 parcel). It generates a discriminant function based on combinations of the predictor variables (in this work the number and type of 1990 LCMGB pixels intersected by the LCM2000 parcel) that provide the best discrimination between the groups. In this case the groups were the LCM2000 broad habitat classes and the variables were the proportions of LCMGB target classes contained within the polygons of the LCM2000 dataset. Each individual was allocated to the group to which it was closest, with a probability figure calculated from relative Mahalanobis distances, to take account of the correlations among variables (Manley, 1994). The groups were the LCM2000 broad habitat classes and the predictor variables are the 25 LCMGB target classes. The prior probabilities were set to be equal for each group (i.e. the probabilities were not weighted by group abundance) and were then weighted based on the number of cases (parcels). So the approach was to identify a predicted broad habitat class membership for each parcel based on how its particular set of intersecting LCMGB pixels matched the modelled distribution for that class. In common, with many probabilistic analyses some prior belief (in the form of prior probabilities) has to be fed into the calculation. In this case weighted and non-weighted prior probabilities were compared for completeness.

3.2 Expert approach

The expert approach has been described elsewhere (Comber *et al.*, 2003b). In brief it involves the following stages:

1. The 1990 pixels intersecting each 2000 parcel were characterised (see Figure 1):
 - a. The number and type of intersecting LCMGB target classes are determined for each parcel.
 - b. The distribution of LCMGB classes is interpreted via the description given by the expert of how the classes of LCMGB relate to each LCM2000 broad habitat. According to the expert descriptions each type of LCMGB pixel will be “unexpected”, “expected” or “uncertain” (U, E Q) relative to the broad habitat class of the LCM2000 parcel. That is, each LCM2000-LCMGB class pair is described as being “unexpected”, “expected” or “uncertain” in the LUT. A triple for each parcel is generated by summing all the LCMGB pixels (that intersect with the LCM2000 parcel) falling under these descriptions when they are run through the expert LUT.

- c. The (U, E, Q) triple values are converted to proportion of the sum of (U + E + Q). As the LCMGB pixels represent an area, this is an area normalisation.

[Insert Figure 1 about here]

2. The parcel was characterised based on its 2000 attribution (see Figure 2):
 - a. For each parcel the number and type of spectral heterogeneity attribute spectral subclasses are extracted. This attribute (Per pixel list) indicates the top five spectral subclasses within the segment by percentage area. These are the 1000's of primary spectral classes from which habitat classes are constructed. A full description is given in Smith and Fuller (2002).
 - b. Each spectral subclass is interpreted via knowledge of expected, unexpected or possible spectral overlap relative to the parcel broad habitat class (this knowledge comes from accompanying the LCM2000 data). These descriptions are used to generate a second “unexpected”, “expected” and “uncertain” triple based on the percentage of each spectral subclass within each LCM2000 parcel.

[Insert Figure 2 about here]

3. Changes in U, E and Q are calculated for each parcel (ΔU , ΔE and ΔQ).
4. ΔU and ΔE were normalised to a standard distribution function for each broad habitat class.
5. The normalised changes in unexpected and expected scores provided beliefs in a hypothesis of change. These were combined using a standard Dempster-Shafer approach to evidence combination.

Earlier work using just only the semantic LUT has shown that in a sample of parcels that were visited, inconsistency between LCMGB and LCM2000 was identified in 100% of the parcels, with 41% of the inconsistency being attributable to change and 59% attributable to error in one or the other dataset (Comber *et al.*, 2004a; 2004b). Note that this result was for expert opinion based on idealised semantics, and that experts were also asked about their understanding of the classes from “technical” and “change” viewpoints, describing their knowledge of common land cover class confusions and likely land cover changes.

3.3 Validation by field survey

The statistical and expert methods were evaluated using parcels that were visited in the field. In December 2002 and March 2003 some 345 parcels were visited and assessments were made of whether the land cover matched LCM2000 and whether it had changed since 1990. The field data was comprised of 151 parcels selected because they were identified as potential change areas (as described in Comber *et al.*, 2004a, 2004b,

2004c) and 194 parcels selected at random. Having a partially random dataset ensured that a large enough number of change parcels were included in the field survey. The assessment of each method was by comparing the proportions of the set of visited parcels that were identified as having changed.

4. Results

Applying both methods to the data and selecting the 345 field visited parcels resulted in dataset with the following attributes:

- LCM2000 broad habitat class;
- Predicted broad habitat class (statistical);
- Probability (statistical);
- Belief in change (expert);
- Observed broad habitat class (field validation)

A probability threshold of greater than or equal to 0.95 was used to identify “change” parcels for both approaches. In the expert method the intersecting LCMGB pixels were interpreted through the different expert relations (semantic, technical and change) and compared with the parcel spectral heterogeneity attribute. In the statistical method the distribution of LCMGB pixels for each broad habitat class served to generate a model describing each broad habitat in terms of the overall distribution of LCMGB target classes. Each parcel was in turn assessed against this model of typical broad habitats and used to populate cross-tabulations for comparison.

The results in Table 2 show the extent to which the different approaches correctly partitioned the field data.

The following proportions were calculated:

- *actual* changes identified by the method (changes identified);
- the errors of omission – how many of the changes were not indicated as having changed by the method;
- the errors of commission – how many of the ‘no changes’ were indicated as having changed;
- An overall reliability figure, given the proportion of all parcels (change or no change) that were correctly partitioned by each approach.

[Insert Table 2 about here]

Statistical

The initial analysis showed that, when using equal priors, only 27.7% of the parcels were assigned the same broad habitat class as was mapped in the LCM2000 and 46.6% when unequal priors were used. Note that these results were not subject to any probability threshold and that there is a considerable variation in the consistency between broad habitats. In the case of unequal priors, 'Neutral Grass' (6.1) and 'Open Dwarf Shrub Heath' (10.2) are particularly inconsistent (both less than 10%), while 'Continuous Urban' (17.1) and 'Water' (13.1) are the most consistent (62% and 57% respectively). The figures, however, suggest a huge rate of landscape change over the 10 years.

The results show that the statistical approaches do not identify many of the changes found in the field, the discriminant analysis with equal prior probabilities performing slightly better than that with weighted ones. The errors of omission are high, showing the high number of changes not identified by this approach. The errors of commission are low due to the low number of parcels identified in total (38). These are reflected in the reliability of statements about 'change' and 'no change' and of overall error

Expert

The expert approach identifies more change parcels than the statistical approach. The proportion of change parcels identified sequentially increases as more evidence is included. Whilst the errors of omission decrease, the errors of commission increase such that when the semantic relations are combined with the technical and the change, 41% of the parcels identified as having changed are 'no change' parcels. The reliability of statements about 'no change' are high, but low for 'change' due to the high error of commission.

Comparison

The different approaches are nominally similar in terms of their overall errors if the last column in Table 2 is taken at face value. However, the statistical approaches identify far too few of the actual changes to be of analytical value. The danger with combining expert information using Dempster-Shafer is that as more information is incorporated, the number of parcels identified as potential change parcels increases. This is illustrated in Figure 3 for woodland parcels in Sherwood Forest.

[Insert Figure 3 about here]

The two approaches identify different subsets of parcels as shown in Figure 4 for the visited parcels. We see that the agreement between the expert and the statistical approach using unequal priors is 48% (i.e. $[174 + 25] / 345$) but the majority of that is composed of parcels not suspected of change by either approach.

[insert Figure 4 about here]

Figures 3 and 4 demonstrate that whilst the expert approach over-estimates change, the statistical approach misses many of the actual changes. This is reflected in the low error scores for the statistical approach: if few parcels are indicated as being candidate change areas then the error terms will be low. The inverse is true for the expert approach, with a high number of errors of commission and low errors of omission. The *conservative* behaviour of the statistical approach is confirmed by Figure 5 which shows that few of the parcels have high probabilities either among the overall set of visited parcels and among those where change was identified. If both approaches were equally good as predictors then we would expect a distribution of beliefs / probabilities which would separate the ‘change’ from the ‘no change’ parcels.

[Insert Figure 5 about here]

5. Discussion

5.1 Results

Discriminant analysis must be performed carefully because there are many choices in the way that group membership is allocated. We have illustrated the differences caused by the choice of prior probabilities. When priors are defined according to the number of parcels the classes with the greatest populations dominate and many of the less frequent classes are not present in the output. For instance, with weighted priors, no parcels were allocated to the broad habitat classes of Arable Cereal, Arable Horticulture, Improved Grasslands, Open Dwarf Shrub Heath, Neutral Grassland, Fen Marsh Swamp or Water. When equal priors were used these less frequent classes were over-estimated to the extent that spurious changes were indicated such as changes to Sea and Bog. This is a generic issue with probability-based approaches where the selection of the prior

probabilities influence profoundly the results of the analysis. This is a fundamental problem with statistical and probabilistic approaches, one that Comber *et al.* (2004d) discuss at length in relation to modelling land cover and expert knowledge. For instance discriminant analyses makes assumptions that the variables are not highly correlated with each other, that the correlation amongst predictors is constant across groups and that the values of each predictor have a normal distribution. These assumptions may be violated in thematic data, classified statistically from remotely sensed data. Some follow-up analyses have shown that performing a PCA on the predictor variables prior to analysis does not improve the discrimination.

A more profound difference can be seen in the different way that the two generic approaches address the problem of relating information from discordant databases. The expert approach is able to take account of different landscape scenarios, the knowledge that is embedded within each of the two maps and their construction, and offers a more holistic view of the data and the landscape processes that are being (differently) represented.

This work supports the work of Fuller *et al.* (2003) who warned against the direct comparison of ontologically discordant land cover datasets. Fuller *et al.* (2003) focussed on the limitations to identifying change due to the relative accuracies of the two datasets: with a 17% rate of actual change and the levels of accuracy estimated for LCMGB and LCM2000, 43% of the combined data area would record differences between the two maps and 57% would show agreement. However in this work we have adopted a different approach by using expert information to identify locales of inconsistency between the datasets, shown to be effective in other work. Whilst it is possible to partition inconsistency between two datasets into error and change, it should be remembered locales of inconsistency are *relative*: they indicate that the view of the world as represented in one dataset are inconsistent with that represented another.

Whilst it is expected that many inconsistencies are false positives for a hypothesis of change, it is possible to provide some measure of confidence in the form of errors of omission and commission. For instance, we might expect 32% of parcels that had actually changed to be missed by the expert approach, and 41% of the parcels identified as having changed, not to have done so. In this context these results are encouraging given the stated accuracy of the data when compared to a sample field survey, and the poor results of the statistical approach.

5.2 Implications

The problem of data discordance addressed by this work has highlighted the paucity of current meta-data and data quality reporting. Historically the natural resource map was a *supporting* supplement to the description of the phenomenon being measured which was included in the report, memoir, or bulletin, *vide* Stamp (1937-1944), Coleman and Maggs (1968), Soil Monographs (e.g. Hodge *et al.*, 1984). However the role of metadata has decreased with the advent of digital representations (Fisher 2003): now the documentation *supports* the map (if it is produced at all) and the user has less understanding of the origins and meanings of the data.

These phenomena are compounded by increasing numbers of non-expert users able to access spatial data through initiatives such as E-science, GRID and INSPIRE, who may be unaware of the wider meaning of the data they incorporate into their analyses. They may be unaware that they have a different conceptual model (ontological uncertainty) than the data producer and are not helped by current metadata standards and data quality reporting paradigms that record what is easily measured, (often showing only that the data producer is able to follow a recipe) but report nothing on semantic or ontological issues.

Metadata reporting needs to change and be enhanced. It is essential that a clear textual description from the producer of the theme and the classes of that theme should be included with the map data. Where a wide-area coverage is being prepared, such as a national or international mapping, it would be very useful to the user to have a self-critical evaluation by the producer specifying their confidence in the classification overall, by class, and by political or landscape region. Where a new dataset is being prepared and there is a pre-existing mapping of the same theme (such as LCM2000 and LCMGB), a commentary on the compatibility between the two classifications should be essential, possibly as a semantic lookup table from all classes to all other classes (Comber *et al.*, 2004a). This evaluation should be included whether the description of classes in the second survey is the same or not; producers should again be self-critical in evaluating this because there may have been subtle changes in the class descriptions as applied.

6. Conclusions

The specific conclusion we are able to make from the results of this comparative analysis is that an expert approach identifies inconsistency between data sets, is able to provide a probability that the inconsistency is

due to land cover change and is more reliable at identifying change than discriminant analysis. However the approach reported is dependent on parcel metadata being available, such as is present in LCM2000, and this is rare; although adaptations for situations without metadata may be possible. Statistical approaches are implemented independent of the metadata and are unable to accommodate differences in ontologies, conceptualisations and semantics. Thus the expert offers a more holistic view of the landscape than is given by measures of statistical similarity.

The wider conclusions from this work are perhaps unanswered questions, but ones that need to be addressed if the full potential of data interoperability and third party access are to be realised:

- 1) How should metadata report on different data *conceptualisations* ?
- 2) How should metadata record who decides and defines what the features of interest in a data set are?
- 3) How is that information to be communicated to the user in an effective way?

Answering these would allow the user to make informed decisions about the suitability of particular datasets for specific analyses.

Acknowledgements

This paper describes work done within the REVIGIS project funded by the European Commission, Project Number IST-1999-14189. We wish to thank our partners in the project, especially Andrew Frank, Robert Jeansoulin, Alfred Stein, Nic Wilson, Mike Worboys and Barry Wyatt.

References

- Anderson, J.R., Hardy, E.E., Roach, J.T. and Witmer, R.E., 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Geological Survey, Professional Paper 964, Reston, VA
- Clerici, A., Perego, S., Tellini, C. and Vescovi, P., 2002. A procedure for landslide susceptibility zonation by the conditional analysis method. *Geomorphology*, 48 (4), 349-364
- Coleman, A. and Maggs, K.R.A., 1968. Land Use Survey Handbook: An Explanation of the Second Land Use Survey of Britain on the Scale of 1: 25,000. Second Land Use Survey, London.
- Comber A.J., Fisher, P.F. and Wadsworth, R.A., 2002. Creating Spatial Information: Commissioning the UK

- Land Cover Map 2000. In Dianne Richardson and Peter van Oosterom (eds.), *Advances in Spatial Data*, Springer-Verlag, Berlin. pp. 351-362.
- Comber, A. Fisher, P., and Wadsworth, R., 2003a. Actor Network Theory: a suitable framework to understand how land cover mapping projects develop? *Land Use Policy*, 20, 299–309.
- Comber, A.J., Fisher, P.F., Wadsworth, R.A., 2003b. A Semantic Statistical Approach to Negotiating Heterogeneous Ontologies. In Shi, W., Goodchild, M., Fisher, P.(eds.), *Proceedings of the 2nd International Symposium on Spatial Data Quality*, 19th-20th March 2003, Honk Kong Polytechnic University, Hong Kong, pp. 349-358.
- Comber, A. Fisher, P. and Wadsworth, R., 2004a. Integrating land cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science*, 18(7): 691-708.
- Comber, A.J., Fisher, P.F., Wadsworth, R.A., 2004b. Assessment of a Semantic Statistical Approach to Detecting Land Cover Change Using Inconsistent Data Sets. *Photogrammetric Engineering and Remote Sensing*, 70(8): 931-938.
- Comber, A.J., Fisher, P.F., Wadsworth, R.A., 2004c. Identifying Land Cover Change Using a Semantic Statistical Approach. In Peter Atkinson, Giles Foody, Stephen Darby and Fulong Wu (eds.), CRC Press, Boca Raton, pp. 73-86.
- Comber, A.J., Law, A.N.R., Lishman, J.R., 2004d. A comparison of Bayes', Dempster-Shafter and endorsement theories for managing knowledge uncertainty in the context of land cover monitoring, *Computers, Environment and Urban Systems*, 28(4), 311-327.
- European Environment Agency, 2001. CORINE Land Cover - Technical guide, http://etc.satellus.se/the_data/Technical_Guide/ [available 27/08/01].
- Fisher, P.F., 2003. Multimedia Reporting of the Results of Natural Resource Surveys, *Transactions in GIS*, 7, 309-324.
- Fuller, R.M., Groom, G.B., and Jones, A.R., 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, 60, 553-562.
- Fuller, R.M., Smith, G.M., and Devereux, B.J., 2003. The characterisation and measurement of land cover change through remote sensing: problems in operational applications? *International Journal of Applied Earth Observation and Geoinformation*, 4, 243-253.

- Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., and Thomson, A.G., 2002. Land Cover Map 2000: construction of a parcel-based vector map from satellite images. *Cartographic Journal*, 39, 15-25.
- Hodge, C.A.H., Burton, R.G.O., Corbett, W.M., Evens, R. and Seale, R.S., 1984. Soils and their use in Eastern England. Herpenden, Soil Survey of England and Wales Bulletin No13.
- Law, J., 1992. Notes on the theory of the actor-network: ordering, strategy and heterogeneity. *Systems Practice* 5(4), 379–393.
- Lund, H.G., 2004. Definitions of Forest, Deforestation, Afforestation, and Reforestation. <http://home.comcast.net/~gyde/DEFpaper.htm>. [Available 05/05/04].
- Manley, B.F.J., 1994. Multivariate statistical methods; a primer. Chapman & Hall / CRC, Boca Raton.
- Smith, G.M. and Fuller, R.M., 2002. Land Cover Map 2000 and meta-data at the land parcel level. In Foody, G.M. and Atkinson, P.M. (eds.), *Uncertainty in Remote Sensing and GIS*, John Wiley and Sons, London, pp. 143-153.
- Stamp, D.L., 1937-1944. *The Land Utilisation Survey of Britain*, Geographical Publications, London.
- Tappeiner, U., Tasser, E. and Tappeiner, G., (1998). Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an alpine area. *Ecological Modelling*, 113 (1-3), 225-237.
- Woodcock, C.E. and Gopal, S. (2000). Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science*, 14 (2), 153-172.

List of Tables and Figures

Table 1. The different descriptions of the class of Bog in LCM2000 and LCMGB, along with the area of each class in the Ordnance Survey 100km by 100km tile SK (the significant features in the descriptions have been underlined by the authors).

Table 2. Proportions of the field data partitioned by Statistical and Expert approaches

Figure 1. Characterisation of the LCM2000 parcel by the intersecting LCMGB pixels

Figure 2. Characterised of the LCM2000 parcel by its spectral subclass attributes

Figure 3. Change parcels identified by the different approaches with some context from OS 1:50000 Raster scanned maps (© Crown Copyright Ordnance Survey. An EDINA Digimap / JISC supplied service).

Figure 4. A Venn diagram of the intersection of the visited parcels identified as potential changes by the two statistical approaches and by the combined evidence of the expert Semantic, Technical and Change scenarios.

Figure 5. Comparison of the expert beliefs and discriminant analysis probabilities. In each case the Expert are on the y-axis and the statistical probabilities on the x-axis.

Survey	Bog Definitions	Area in SK
LCMGB	<p>Bog</p> <p><u>Permanent waterlogging</u>, resulting in depositions of acidic peat mostly herbaceous communities of wetlands with <u>permanent or temporary standing water</u></p> <p>Lowland Bogs (subclass) Carry most of the species of upland bogs, but in an obviously lowland context, with <i>Myrica gale</i> and <i>Eriophorum spp.</i> being highly characteristic.</p> <p>Upland bogs (subclass) Have many of the species of grass and dwarf shrub heaths Characterised by <u>water-logging, perhaps with surface water</u>, especially in winter. species such as bog myrtle (<i>Myrica gale</i>) and cotton grass (<i>Eriophorum spp.</i>) in addition to the species of grass and dwarf shrub moorlands.</p>	12 pixels ($<1 \text{ km}^2$)
LCM2000	<p>Bogs include ericaceous, herbaceous and mossy vegetation in areas with <u>peat</u> <u>$>0.5 \text{ m deep}$</u>; ericaceous bogs are distinguished at subclass level. Inclusion of Ericaceous bogs contrasts with LCMGB 1990 where bogs were herbaceous or mossy in seasonal standing water</p>	120728 pixels (75 km^2)

Table 1. The different descriptions of the class of Bog in LCM2000 and LCMGB, along with the area of each class in the Ordnance Survey 100km by 100km tile SK (the significant features in the descriptions have been underlined by the authors).

		Changes identified	Omission	Commission	Overall error
Statistical	Discriminant Analysis (equal priors)	0.211	0.789	0.090	0.206
	Discriminant Analysis (unequal priors)	0.158	0.842	0.101	0.223
Expert	Semantic	0.614	0.386	0.288	0.304
	Semantic and Technical	0.649	0.351	0.375	0.371
	Semantic, Technical and Change	0.684	0.316	0.413	0.397

Table 2. Proportions of the field data partitioned by statistical and expert approaches

Figure 1. Characterisation of the LCM2000 parcel by the intersecting LCMGB pixels

Figure 2. Characterised of the LCM2000 parcel by its spectral subclass attributes

Figure 3. Change parcels identified by the different approaches with some context from OS 1:50000 Raster scanned maps (© Crown Copyright Ordnance Survey. An EDINA Digimap / JISC supplied service).

Figure 4. A Venn diagram of the intersection of the visited parcels identified as potential changes by the two statistical approaches and by the combined evidence of the expert Semantic, Technical and Change scenarios.

Figure 5. Comparison of the expert beliefs and discriminant analysis probabilities. In each case the Expert are on the y-axis and the statistical probabilities on the x-axis.