CEH contract report
to the
Department of the Environment, Food and Rural Affairs

CS2000 Module 9
DATA INTEGRATION FOR LOCALISED RESULTS AND SUPPORT FOR INDICATORS OF COUNTRYSIDE CHARACTER AND QUALITY
Final Report (A5) – May 2004

R.A. Hill & G.M. Smith

Centre for Ecology and Hydrology
Monks Wood
Abbots Ripton
Huntingdon PE28 2LS

CONTRACT
No. CR0263

EXECUTIVE SUMMARY

This is one of the two Final Report of Countryside Survey Module 9 'Data Integration for Localised Results and Support for Indicators of Countryside Character and Quality'. Countryside Survey is a programme of regular environmetal assessments based on field survey and satellite mapping. Module 9 aimed to derive data for and in support of the production of indicators. Module 9A specifically aimed to determine how data from Countryside Survey 2000 (CS2000) field survey (FS) and Land Cover Map 2000 (LCM2000) could be integrated to produce consistent and robust estimates of stock and change at different scales.

LCM2000 and FS each have their strengths and weaknesses in generating land cover statistics. An inter-calibration of LCM2000 and FS has been performed, which has determined how these two datasets can be integrated to produce consistent and robust estimates of land cover at different scales, exploiting the strengths and minimising the weaknesses of these two mapping approaches.

The integration procedure involved three stages: correspondence analysis, stratification and inter-calibration.

Correspondence analysis:

- The 569 FS 1 km squares and their equivalent LCM2000 sections were inter-compared to get a broad picture of correspondence between the two datasets and to allow the generation of Broad Habitat (BH) cover statistics at the national level, equivalent to those of the FS. When differences occurred, the source or explanation was not always obvious without further analysis and interpretation.
- The correspondence analysis determined calibration matrices between FS and LCM2000. The results were examined in detail and at a range of spatial scales and for different spatial zonations (e.g. the National Land Classes (NLC), the mosaic of satellite images that made up LCM2000). This analysis highlighted an inconsistency in correspondence between LCM2000 and FS stock estimates for different satellite image pairs, and also between areas of LCM2000 that were mapped using single or two-date image classification and between target and non-target date images.

Stratification:

- The chosen spatial framework for data integration was the NLC, with summary land cover data generated for each 1 km National Grid cell in England.
- A spatial stratification based on the NLCs offered the optimal use of the spatial distribution of FS squares, as NLCs provided the stratification for the systematic random location of FS squares. Also, an NLC stratification would replace the artificial image boundaries present in LCM2000 (which were based on satellite orbiting parameters) with boundaries based on environmental variables.

Calibration:

- Calibration was carried out using the matrices from the stratified correspondence analysis. The approach used the calibration matrices to scale the BH stock statistics of LCM2000 for a particular region to resemble those that would be derived by a comprehensive field survey. The emphasis was on incorporating the strengths of FS and LCM2000 to minimise the weaknesses of these two datasets.

- The relationship between FS and LCM2000 in an NLC was represented by a calibration matrix, which was the average correspondence matrix, derived from the set of correspondence matrices for each of the FS squares within the NLC. This calibration matrix was then used to convert the land cover proportions from LCM2000 for each 1 km square to those that would be expected from the FS.
- The calibration procedure had to be more than the simple application of correspondence matrices. The results from the correspondence analysis identified areas where additional information was required to correct weaknesses and guided the formulation of a set of knowledge-based corrections. Thus, coastal, urban, elevation, soil type and woodland masks were applied to control the spatial application of the inter-calibration method as appropriate.

Ancillary data were identified as a possible means of correcting deficiencies in the calibrated 1 km data set and validating the final results. A range of data sets were obtained and extensive comparisons undertaken with FS, LCM2000 and the calibrated 1 km data set. These ancillary data sets have generated their own problems and issues with decisions necessary on quality, definition and coverage along with their suitability for a role within Module 9A.

For the datasets considered useful for additional knowledge-based correction within the calibration process, suitable rule bases were developed and tested. This resulted in a total of seven iterations of the calibrated 1 km product to a point where it is in it's final form with respect to this project. The bootstrapping procedure has been implemented to provide uncertainty information at the 1 km square level.

A full validation has been undertaken using a number of the ancillary datasets for statistical comparison, by plotting BH totals per NLC as derived from FS, LCM2000 and calibration, and by visual inspection of the 1 km calibrated maps in CIS. The calibrated product provides better stock estimates, and maps of their distribution, compared with the extrapolated FS dataset in all cases. However, there are three examples of where the calibrated product is worse that the original LCM2000 1 km summary product: standing open water & canals, littoral sediment and built up & gardens. Standing open water & canals loses its geographical distribution as linear features; littoral sediment is reduced in area by the restricted off-shore limits of the NLCs; and built up & gardens is over-estimated due to the unique situation for this land cover type in the FS.

In general, the calibration procedure often results in a very low (i.e. 1-2%) cover of BHs where they should not be expected to occur; visually this causes a general background level of a particular BH across the landscape (constrained by the spatial boundaries of NLCs); and occasionally removes features that are anomalous within an NLC or poorly sampled by the FS squares within that NLC.

To derive estimates of uncertainty for the integrated BH stock estimates per 1 km square, a range of different calibration matrices were generated for each NLC. The individual calibration matrices were generated by randomly sampling the correspondence matrices within the NLC with replacement. A thousand different calibration matrices were generated for each NLC. These were then used to produce a thousand estimates of the stock of each BH within each 1 km square. The thousand estimates were then ranked and the 50th and 950th estimates extracted to give the upper and lower limits of uncertainty.

A strategy for the measurement of landscape pattern has been developed and these values have been derived for England at a 1 km level. Metrics such as the Simpson's Index of Diversity, the Evenness Index, and the Bray-Curtis Similarity Index have been computed using the proportional composition of Broad Habitats per 1 km square. These pattern metrics are not affected by issues of how landscape patches or mosaics are defined, how well landscape features are identified as individual objects by satellite-based mapping, or the extent to which landscape features are truncated by field-based sampling using 1 km squares.

A brief investigation into the derivation of more detailed pattern metrics for each FS square has considered the various issues concerning object and boundary representation and the calculation of metrics in FS and LCM2000. At a general level it has been shown that LCM2000 identifies landscape structure, but the boundaries themselves are not necessarily meaningful or comprehensive in the way that FS would achieve.

Through the process of meetings, presentations and email distribution the procedures developed and results produced by Module 9A have been communicated to a wide range of user groups. These procedures have in general been accepted and deficiencies identified for further considerations. The data sets have been supplied to the Module 9A Technical Advisory Group, the Countryside Quality Counts (CQC) team and offered to the Countryside Information System (CIS) user community as GB mode, 1 km summary datasets for CIS version 7.

Module 9A has resulted in a clearer understanding of the issues associated with LCM2000, FS, their integration and the resulting calibrated 1 km data sets. In many ways the project has posed as many questions as it has provided answers. It has identified the strengths of each approach to landscape survey and attempted to maximise these. It has identified the weaknesses and devised mitigation strategies that compensate or minimise them.

From this analysis recommendation for future surveys can be extracted based on the specifications which are developed for their production. From this analysis recommendation for future surveys can be extracted associated with:

1. Timing of surveys.
2. The nature of different minimum mappable units (MMU).
3. Differences in class definitions or in mapping protocol.
4. The rarity and typical patch size / shape.
5. Appropriateness of thematic classification.
6. Survey information from the coastal zone.
7. Representativeness of the sampling strategy.
8. Normalisation of satellite images.
9. Use of ancillary data.

Module 9A results should be used in association with the CS2000 Module 17 - Finding Out Causes and Understanding Significance (FOCUS) results with particular relevance for the planning of future Countryside Surveys.

TABLE OF CONTENTS

INTRODUCTION

This is the final report of Countryside Survey Module 9 'Data Integration for Localised Results and Support for Indicators of Countryside Character and Quality'. Module 9 aimed to address Defra's requirement for information on the status and changing character of the countryside at a local level, including an evaluation of countryside quality through the use of indicators. In this later respect Module 9 was to interface with the Countryside Quality Counts (CQC) project which aimed to develop national indicators of change in countryside character and countryside quality for the English countryside

Module 9 had two sub-sections. The remit of Module 9A was 'To determine how data from Countryside Survey 2000 (CS2000) Field Survey (FS) and Land Cover Map 2000 (LCM2000) could be integrated to produce consistent and robust estimates of stock and change at different scales.' This part of the project was designed as follows:

Feasibility phase

1. Develop and test integration methods,
2. Demonstrate methodology for England,
3. Produce provisional estimates for English Regions, Wales & Scotland,
4. Produce prototype calibrated 1 km data set for England,

Operational phase

5. Refine methodology based on user feedback,
6. Quantify and explain sources of error,
7. Develop and evaluate measure of landscape pattern.

Modules 9B was to demonstrate how CS2000 data could be used alongside other information on countryside features to support the development of national indicators of change in countryside character and countryside quality, paying particular attention to issues of ecological character and condition.

This project has now been completed successfully. The principal outputs from Module 9A have been four new England only 1 km summary data files for the Countryside Information System (CIS). These consist of: Broad Habitat stock estimates resulting from the process of inter-calibrating CS2000 FS and LCM2000 statistics; measures of uncertainty for the Broad Habitat stock estimates derived from the calibration procedure (lower and upper confidence limits); and metrics of landscape pattern derived by comparing the land cover types and coverage of neighbouring 1 km squares. These CIS data layers provide information on Broad Habitat stocks for England and the English Regions. However, the production of a calibrated 1 km data set and regional estimates for Scotland and Wales was not undertaken as these countries did not join the CS2000 Module 9 group.

BACKGROUND

Countryside Survey 2000

The main component of CS2000 was a FS module that recorded a stratified sample of 569 1 km squares. The strata were the 40 National Land Classes (NLC); an environmental regionalisation based on physical geographical variables. The FS recorded areal features (e.g. fields), linear features (e.g. hedges) and point features (e.g. ponds) in great thematic and spatial detail, using 1:10 000 Ordnance Survey (OS) maps as the base; associated species (mostly plants) were also recorded. The main characteristics ('primary codes') denoted the type of feature (e.g. a wheat field, a hawthorn hedge, an individual tree); secondary codes recorded qualifying information (e.g. about species and cover, feature-size and management). It was possible to combine primary and secondary codes in an almost infinite variety of ways, to record some of the true complexity of the countryside. The information was necessarily simplified to generate the basic 'widespread Broad Habitat' (BH) classification of CS2000. BHs were based on selected combinations of primary and secondary codes, using objective rules. Further subdivisions were possible: for example, the CS1990 'baseline classes' give an objectively based, consistent, tried and tested classification.

The LCM2000 was also a module within the CS2000 project. The LCM2000 was based on the analysis of satellite image data with a spatial resolution of 25 m and provided a comprehensive map of widespread BHs. LCM2000 used image segmentation to identify relatively uniform areas within the images that were essentially distinct land parcels (e.g. fields, water bodies, urban areas and mosaics of semi-natural vegetation). The LCM2000 land parcels, or segments, were held in a vector format similar to the FS data. The segments were classified using the spectral character of the image data (i.e. reflectance, often from two different seasons). Enhancements were provided by knowledge-based corrections driven by ancillary data (e.g. elevation, soil sensitivity). LCM2000 used a hierarchical classification scheme consisting of 16 target classes, which were further subdivided to make 24 subclasses, with these in turn subdivided to give up to 72 class-variants. Most BHs were themselves target classes, though some were defined at the subclass level.

Policy driver

The Rural White Paper published in November 2000 confirmed the Government's intention to develop an indicator of countryside quality to provide a measure of progress toward sustainable development. For such an indicator information was needed on the status and changing character of the countryside at a local level, including an evaluation of countryside quality. No single data source was available which could provide this information and it was therefore deemed to be necessary to combine data from several different sources.

Initial results of CS2000, also published in November 2000, included an update to the Government's Quality of Life Counts indicator on landscape features. However, the results published still did not adequately assess the more localised changes in landscape and habitats usually associated with countryside quality. CS2000 Module 9 was therefore devised to provide Defra and other sponsors improved local estimates of the stock and change in broad habitat types and landscape features for a range of policy applications. Module 9 also aimed to provide indicator development driven by the improved localised estimates and additional ancillary data. The results for regions and other administrative areas were to be published and data were to be made available for use in Defra's Countryside Information System (CIS). The results of Module 9 were also designed to contribute to CQC project which ran in parallel.

Calibration of FS and LCM2000

The 569 FS 1 km squares and the equivalent LCM2000 sections were inter-compared as part of the LCM2000 production programme to get a broad picture of LCM2000 map accuracy and to allow the generation of BH cover statistics at the national level, equivalent to those of the FS. The FS data were not 'ground truth': a quality assurance sample-survey recorded 88% agreement for re-survey of the original primary codes. In the absence of 'ground truth', the process of inter-comparison was one of 'calibration' where the FS and LCM2000 were quantitatively related. When differences occurred, the source or explanation was not always obvious without further analysis and interpretation.

These comparisons between FS and LCM2000 were raster-based at 2.5 m spatial resolution. The first step assessed the need for a shift in x- and / or y-directions of the LCM2000 data relative to the FS data. Further analysis then used, where appropriate, the shifted data set. The comparisons generated correspondence matrices, one for each FS 1 km square. Correspondences were recorded per-pixel (direct) and per-segment (LCM2000 segment labelled with FS data) and per-parcel (FS parcel labelled with LCM2000 data). To provide confidence limits for the measures of correspondence, a 'bootstrapping' procedure was adopted.

LCM2000 segments, compared with FS parcels, showed a basic correspondence of 63.4% in per-parcel comparisons at BH level (allowing for the FS generalisation of *Built up areas* and the LCM2000 omission of *Boundary and linear features* and *Rivers and streams*). As correspondence cannot realistically exceed the 88% repeatability of the FS, LCM2000 seems to be scoring at least 72% of its maximum potential. About 5% of the mis-match is explained by the 25 m grid underlying the image parcels, compared with the continuously variable structure of the FS. (If the field data are resampled onto the 25 m grid, the results show 95% correspondence with the original input data). The 0.5 ha Minimum Mappable Unit (MMU) of LCM2000 contrasts with the 0.04 ha MMU of the FS and explains many of the differences, especially for BHs which occur in less extensive stands (more than 4% of the area recorded by FS comprised parcels, not linear features, which were below the LCM2000 MMU). Time-differences explain other mis-matches: the FS was predominantly undertaken in 1998; LCM2000 used images mainly from 1998-2001. Evidently up to 15% of differences can be explained by the underlying structure of LCM2000 and, additionally, by its coarser MMU, and by date-differences. This suggest that LCM2000 may record Target classes with 87% success; to quote a figure of *c*. 85% accuracy at Target class level seems realistic.

For a given NLC a single calibration matrix was produced by averaging the correspondence matrices for the FS squares that fall within it. A calibration could then made between the FS and LCM2000 by passing the LCM2000 BH proportions for the NLC through the calibration matrix. This process was repeated for each NLC and the results combined to given calibrated regional estimates of land cover. Such estimates were produced for Great Britain, England, Scotland, Wales and the six Environmental Zones. These results can be found in the Final Report of the LCM2000 project (Fuller et al., 2002; http://www.cs2000.org.uk/ Final_reports/ M07_final_report.htm).

CALIBRATION FOR SMALLER REGIONS

The calibration procedures developed during the LCM2000 production could be applied at a range of spatial scales / resolutions, although in this instance they were tailored to provide results at a 1 km^2 spatial resolution.

This work was not simply a repeat of the LCM2000 production calibration as many issues that were masked at the national and Environmental Zone level became significant when working at a 1 km spatial resolution. The developments for Module 9A therefore had three main components that considered i) the correspondence between the FS and LCM2000 results, ii) the stratification used to structure the calibration procedure spatially and iii) the calibration procedure itself. The components were dealt with in parallel at the beginning of the work, but were later merged as the calibration was refined.

The main components can be described thus:

i. ***Correspondence analysis*** was used during LCM2000 production to determine the calibration matrices between LCM2000 and FS. The same approach was used during this project, but the results were examined in more detail and at a range of scales.
ii. ***Stratification*** using the NLC framework was assessed in the light of the fact that the calibration was designed to be between FS data and data derived from satellite images. The FS data were collected in support of the NLCs, while the satellite image data collections were controlled by the satellites' orbital parameters.
iii. ***Calibration*** was carried out using the matrices from stratified correspondence analysis. The approach used the calibration matrices to scale the Broad Habitat stock statistics of LCM2000 for a particular region to resemble those that would be derived by a comprehensive field survey. The emphasis was on incorporating the strengths of FS and LCM2000 to minimise the weaknesses of these two datasets. The simple approach was enhanced in a number of ways to accommodate problems identified during LCM2000 production and from i. and ii. above.

Correspondence between LCM2000 and FS

Correspondence analysis was used to understand the similarities and differences between full resolution LCM2000 and FS data for each of the 569 FS squares. This study used correspondence matrices generated by per-pixel comparisons; a direct overlay, with no regard for the structure of either dataset. FS parcels and LCM2000 segments were sampled onto a grid with a 2.5 m cell-size. To accommodate residual errors in the geo-registration of satellite images, the LCM2000 data were shifted to improve alignment. The correspondence analysis operated using shifted extracts (where appropriate) from LCM2000 Release 1. The overall mean shift distance was 53 m, with 48 % of squares shifted one pixel (25 m) or less in x- and or y- directions and 62 % shifted two pixels or less. Per-pixel scores of correspondence between the two data sets (160 000 samples at 2.5 m) were tabulated for each 1 km square.

Table 1 shows an example of a simplified correspondence matrix. Values on the diagonal represent samples where the LCM2000 and FS agree, while those off the diagonal show confusion. The correspondence matrices from all FS squares can be analysed individually or averaged across the Environment Zones, Regions, or even to give one set of values for GB.

LCM2000 \ CS2000	Broadleaved, mixed and yew woodland	Coniferous woodland	Arable and horticulture	Improved grassland	Neutral & calcareous grassland	Acid grassland	Bracken	Dwarf shrub heath	Fen, marsh and swamp	Bog	Standing open water and canals	Montane habitats	Inland rock	Built up areas and gardens	Total	errors of comission
Broadleaved, mixed, yew woodland	10715	1319	1927	5501	504	44	20	0	173	0	86	0	0	1506	21795	50.84
Coniferous woodland	1066	2470	290	147	9	7	2	0	0	0	0	0	0	103	4095	39.69
Arable and horticulture	1364	325	23508	13233	350	8	0	0	158	0	63	0	1	2499	41509	43.38
Improved grassland	3548	250	5209	48906	2202	43	13	0	1037	0	33	0	2	4145	65389	25.24
Neutral & calcareous grassland	622	182	1772	7643	395	68	16	0	12	0	16	0	0	1193		
Acid grassland	279	1	274	1487	36	0	0	5	0	0	1	0	0	881	2964	100.00
Bracken	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
Dwarf shrub heath	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
Fen, marsh and swamp	65	0	0	3	0	0	0	0	0	0	0	0	0	14	82	100.00
Bog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
Standing open water	187	59	70	77	0	0	0	0	0	0	0	0	10	0	404	100.00
Montane habitats	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
Inland rock	64	0	354	319	26	0	0	0	0	0	0	0	0	91	852	100.00
Built up areas and gardens	435	56	1189	1308	95	10	0	0	14	0	5	0	0	7770	10881	28.60
Total	18346	4662	34593	78624	3617	180	52	5	1394	0	205	0	13	18203	160000	
errors of omission	41.59	47.02	32.05	37.80	89.07	100.00	100.00	100.00	100.00	0.00	100.00	0.00	100.00	57.31	OA	58.50 Overall Correspondence

Table 1. An example correspondence matrix for a single FS square, using aggregated classes for clarity on the page.

The overall correspondence statistics for the 569 FS squares are summarised in Figure 1. For mapping the BHs, the range of correspondence between FS and LCM2000 for individual squares was 0% to 98%, with a mean of 53%. The modal percentile range was 70-80% correspondence. The CS2000 squares with lowest correspondence were frequently found in upland areas, where the ancillary data used in the knowledge-based corrections were insufficient to distinguish accurately between Dwarf shrub heath and Bog.

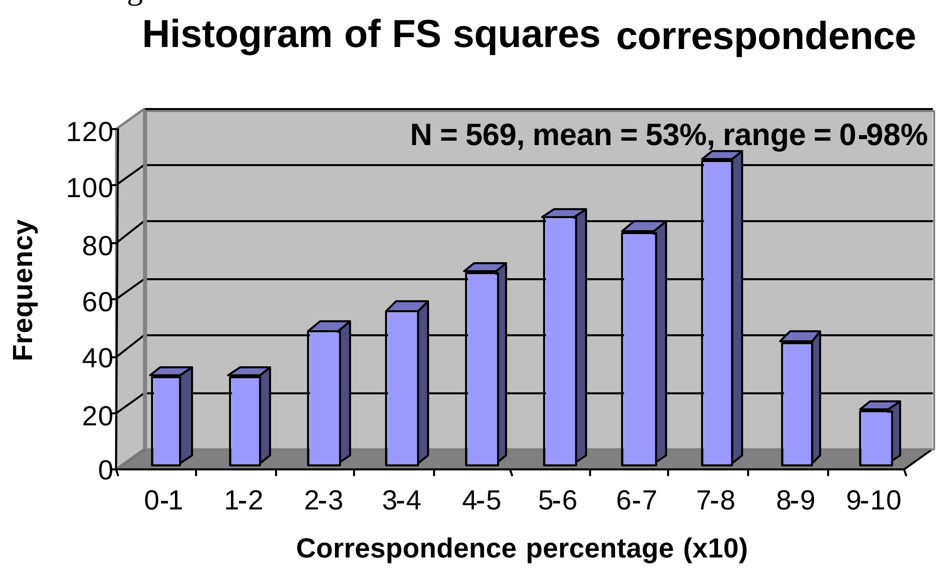


Figure 1. Correspondence between FS and LCM for mapping Broad Habitats

A surface representing the correspondence between LCM2000 and FS across GB is shown in Figure 2. Each 1 km cell has an interpolated correspondence value based on the actual correspondence between FS and LCM2000 for surrounding FS squares. This map was derived from an Inverse Distance Weighted spatial interpolation of the overall correspondence value for all BHs for each FS square. The value in each interpolated cell will be determined by the correspondence of and distance to the nearest FS squares. The location of some of the FS squares can easily be seen in Figure 2 where they have an anomalously low (pale pink) or high (dark red) correspondence compared with surrounding values. This interpolated surface demonstrates a general gradient, with higher correspondence in the managed arable and pastoral landscapes of the south east and lower correspondence in the uplands of the north and west. Areas of low correspondence may reflect a high level of inter-class confusion between two or three particular BHs or a general high level of BH inter-class confusion.

Table 2 shows the relative differences in GB-level BH stock estimates between FS and LCM2000. In summary, the following discrepancies were identified by correspondence analysis:

- LCM2000 apparently over estimates Calcareous and Neutral grassland, at the expense of Improved grassland due to the problems inherent in mapping a continuum of grassland intensity and the use of a soil sensitivity mask with deficient class boundaries;
- LCM2000 apparently over estimates Dwarf shrub heath, at the expense of Bog due to problems with the peat mask used in knowledge-based correction;
- Some land-cover classes (e.g. Bracken, Fen, marsh & swamp, Supra-littoral and Littoral classes) are too rare or of too limited an extent to be recorded consistently in LCM2000 due to limitations of the training data and the MMU of 0.5 ha;
- Montane habitats were identified in LCM2000 using a decision rule (elevation > 600m) that was too generalised and based on coarse spatial resolution vegetation records;
- Inland rock was over estimated in lowland Britain due to spectral similarity with un-vegetated arable fields, and this could not be corrected by knowledge-based procedures as both inland rock (e.g. quarries) and un-vegetated fields can occur in a lowland context;
- FS does not sample within core urban areas and so extrapolated national statistics based on FS will inherently under estimate the spatial coverage of this land class. Within squares where FS does map urban, it may be over estimated as no distinctions are made for urban green space etc.

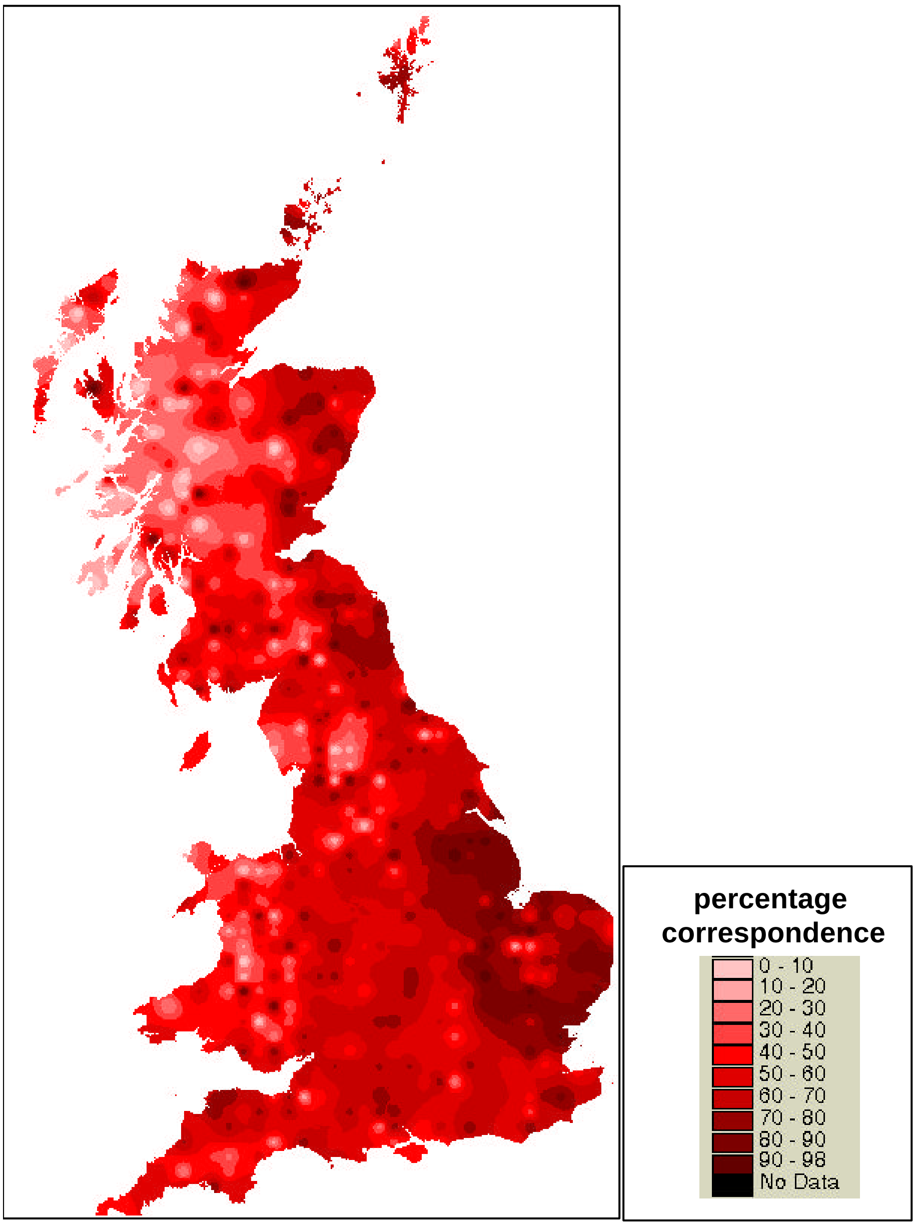


Figure 2. Interpolated correspondence between FS and LCM for Broad Habitat mapping.

BHs where LCM2000 estimates are lower than the 95% confidence limits of FS	BHs where LCM2000 estimates are higher than the 95% confidence limits of FS
Improved grassland	Neutral grassland
Bracken	Calcareous grassland
Fen, marsh & swamp	Dwarf shrub heath
Bog	Montane
Supralittoral rock	Inland rock
Supralittoral sediment	Built up & gardens*
Littoral rock	

* Note that the FS does not sample Built up & gardens within core urban areas

Table 2. The differences in land-cover statistics for BHs across GB as sampled by the FS and census of LCM2000.

The BH inter-class confusion was more complex than the above GB areal estimates of land-cover would suggest. Across GB it was possible to find examples of inter-class confusion between virtually all BHs. However, the most frequently occurring examples of BH inter-class confusion were: between Improved grassland and Arable & horticulture, Neutral grassland, or Calcareous grassland; between Neutral grassland and Calcareous grasslands; and between Bog, Acid grassland and Dwarf shrub heath.

The BH inter-class confusion can be summarised as resulting from distinctions between FS and LCM2000 and misclassification in LCM2000. The former reflect the inherent differences between the two surveying approaches, whilst the latter reflect the difficulties in mapping detailed thematic classes using satellite data. The distinctions between FS and LCM2000 include:

- Different surveying dates: only half of the LCM2000 image pairs were from the FS 'target period' and therefore land-use rotation between crops and ley grass created apparent non-correspondence;
- Different boundary positions and MMUs resulted in non-correspondence and apparent patch effects when FS and LCM2000 were compared at a 2.5 m pixel scale;
- The FS and LCM2000 had different approaches to mapping within urban and woodland boundaries;
- Varying state of tides between FS and the time of satellite image acquisition resulted in different extents of coastal BHs being mapped;
- When comparing FS and LCM2000 there is an issue of the representativeness of 569 FS squares, especially if sub-divided into a spatial stratification such as the NLCs.

The misclassification in LCM2000 occurred due to:

- Compromised image dates; early summer or late winter imagery reduced spectral distinctions between certain land cover types that are strongest in mid-summer and mid-winter imagery;
- Spectral similarity occurred between land cover types (e.g. bare and un-vegetated land, different grassland types);
- Differing illumination levels due to aspect caused increased spectral ranges of land cover types, increasing the chances of spectral confusion between land cover types;
- Varying detail and quality of ancillary data used in knowledge-based corrections resulted in some localised misclassification;
- The difficulty of identifying the rarer land cover types meant these were often under-represented in LCM2000.

Stratification

Local and regional scale patterns occur in the BH inter-class confusion described above resulting from the various boundaries in the combined data sets used to create LCM2000. These include:

- boundaries between satellite image pairs;
- boundaries within satellite image pairs resulting from summer or winter only data or local in-filling of cloud holes with LCMGB 1990 data;
- boundaries of ancillary data masks, such as soil sensitivity, peat depth, coasts;
- boundaries in the application of knowledge-based correction rules, e.g. thresholds in elevation, slope, vegetation indices, etc.

The results of correspondence analysis have been examined in detail at a range of spatial scales and for different spatial zonations. There are a number of stratifications that could be identified as the spatial framework for the calibration procedure. For instance, average calibration matrices might be better derived per-satellite image; or according to regions of LCM2000 derived with one-date or two-date images; and / or whether the data were from the target dates or based on compromised definitions of ‘summer’ and ‘winter’.

The nature of BH inter-class confusion and of apparent BH over or under estimation varies with different stratifications. As the stratification becomes increasingly multi-zonal, so the number of FS squares per zone declines, rendering the calibration matrices less representative statistically. For example, 36 pairs of satellite images were used to cover GB in LCM2000. Within these 36 image pairs, the number of FS squares varies between a minimum of 1 and a maximum of 42. However, this stratification could be subdivided according to whether classification was based on summer-winter composite, summer-only or winter-only data, or whether the image pairs were within or outside the ‘target period’.

Figure 3 shows some examples of the spatial variation in correspondence between FS and LCM2000 for sample BHs when stratified by image pair. For each BH shown, the correspondence between FS and LCM2000 has been calculated using all of the confusion matrices from FS squares located within the boundaries of each satellite image pair. For each BH shown, the correspondence is displayed per image pair and thus many of the image boundaries can easily be seen. In calculating and displaying the correspondence results in this way, the spatial variation between image pairs in the strength and direction of non-correspondence for individual BHs becomes apparent. The dark colours represent areas where LCM2000 has a higher proportion of a particular BH within an image pair compared with FS, whilst pastel colours represent areas where LCM2000 has a lower proportion of a BH. For the sample BHs shown in Figure 3, note that there is no consistent trend at the GB level in whether LCM2000 apparently over or under estimates percentage coverage compared with FS, or by what extent.

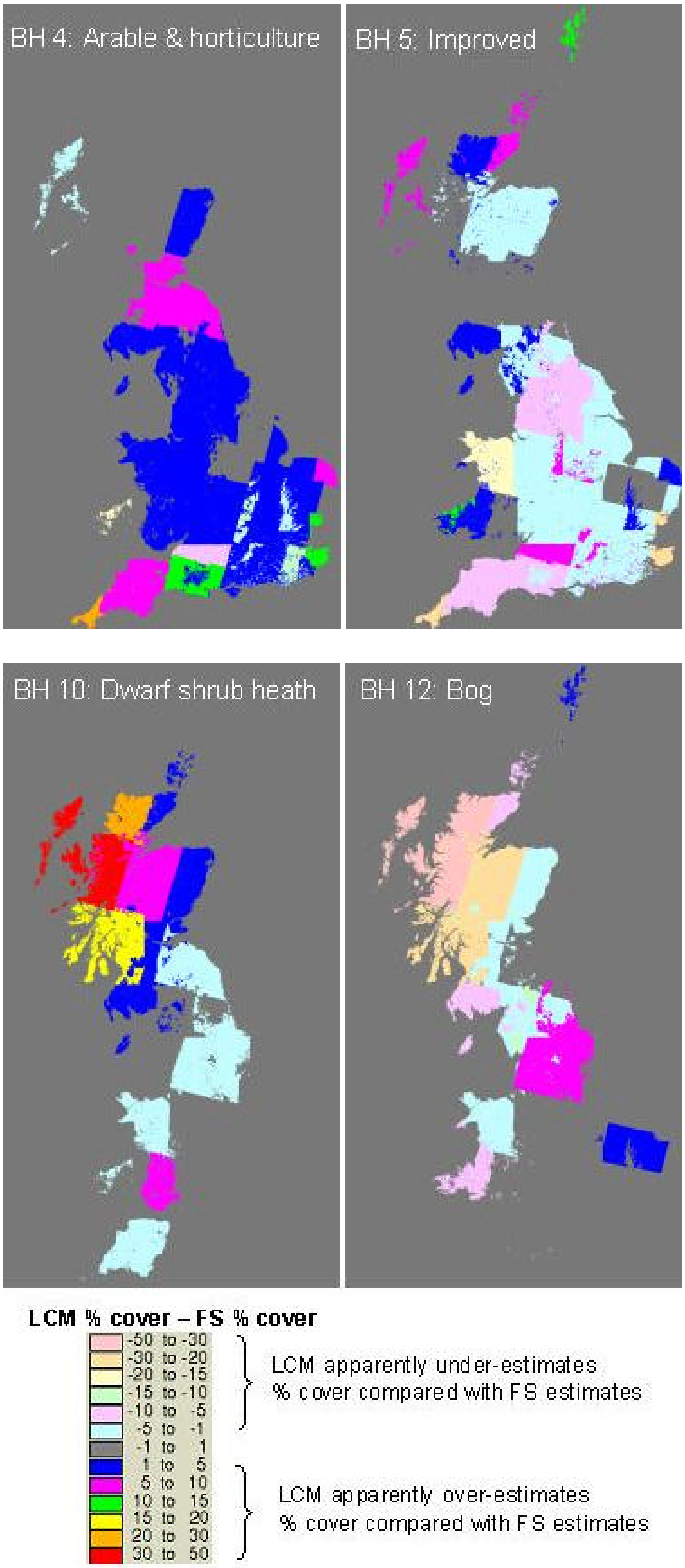


Figure 3. Correspondence between FS and LCM2000 per image pair.

The spatial patterning of correspondence between FS and LCM2000 reported above is an artificial one resulting from the distribution of images dictated by the satellite orbital parameters. Stratification using the NLC was assessed as a method of off-setting the imprints in LCM2000 of the satellite sampling framework as it is more relevant to the biogeographical pattern of GB. The number of field survey squares per NLC varies between 6 and 30. The results of correspondence analysis based on the NLC stratification show a different spatial pattern and different values for apparent over or under estimation per zone compared with the satellite scene stratification. Nonetheless, the overall nature of BH inter-class confusion remains largely consistent between the satellite scene and NLC based stratifications. Thus, for both stratifications, compared with FS estimates, LCM2000 frequently under estimates the percent coverage of Fen, marsh & swamp, Bog and Built up & gardens, and over estimates the percent coverage of Arable & horticulture, Neutral, Calcareous and Acid grasslands, and Dwarf shrub heath. However, in the satellite scene stratification LCM2000 also frequently under-estimates the percent coverage of Broadleaf, mixed & yew woodland, Improved grassland and Bracken compared with FS.

The chosen spatial framework for data integration was the NLCs, with summary land cover data generated for each 1 km National Grid cell in England from LCM2000. A spatial stratification based on the NLCs offered the optimal use of the spatial distribution of FS squares, as the original NLC provided the stratification for the systematic random location of FS squares. Also, an NLC stratification would replace the artificial image boundaries present in LCM2000 (which were based on satellite orbiting parameters) with boundaries that are based on environmental variables (such as topography and geology).

Calibration

The calibration method used the correspondence matrices for a given NLC to scale the results of LCM2000 for each 1 km square within that NLC to resemble the results that would be derived by a compromise between the comprehensive LCM2000 and sampled based FS. The emphasis in the development of the calibration method was on incorporating the strengths of FS and LCM2000 and minimising their weaknesses.

The calibration was more than the simple application of correspondence matrices (Figure 1). The results from the correspondence analysis identified areas where additional information was required to correct weaknesses and guide the formulation of a set of knowledge-based corrections. For instance, coastal, urban, elevation, soil type and woodland masks have been applied to control the spatial application of the inter-calibration method as appropriate.

The initial step of the method was the production of a set of calibration matrices, one for each NLC. Each of the NLC calibration matrices was the average correspondence matrix, derived from the set of correspondence matrices for each of the FS squares within the NLC. Thus,

$$A_k = \frac{1}{S}\sum_{l=S}^{1} M_l \qquad \text{Equation 1}$$

where A_k is the calibration matrix for NLC k, M are the individual correspondence matrices for the FS squares within NLC k and S is the number of FS squares within NLC k. Each element of the calibration matrix, A_{ij}, denotes the value for row i column j of the calibration matrix, i.e. the proportion of LCM2000 type i classified as FS type j.

The above process differs slightly when applying bootstrapping to derive correspondence, but that will be described later.

The application of the calibration matrices to the original 1 km LCM2000 BH data set used the same basic procedure as that used for the regional estimates within the LCM2000 production. For a given 1 km square the LCM2000 data were calibrated to FS equivalent values by multiplying the LCM2000 BH proportions for that 1 km square by the calibration matrix for the NLC within which the 1 km square lies, i.e.

$$FS_m = A_k \times LCM_m \qquad \text{Equation 2}$$

where FS_m and LCM_m are vectors of the proportions of each BH within the m^{th} 1 km square. LCM_m is the vector from the original LCM2000 data and FS_m is the resulting vector which forms the new calibrated data set.

As an example, consider this hypothetical calibration matrix:

LCM2000 BH	Field survey BH Broadleaved	conifer	urban
broadleaved	0.75	0.10	0.15
conifer	0.10	0.85	0.05
urban	0.05	0.05	0.90

Results have been normalised so values sum to '1' across the rows. The above matrix shows the proportions of each of the land cover types that were mapped as the same or different land cover types in the other survey. For instance, of the area mapped by LCM2000 as broadleaved, 75 % was mapped as broadleaved by the FS, but 10 % was mapped by the FS as conifer and 15 % as urban. The calibration matrix allows the LCM2000 proportions to be altered by performing the following analysis:

LCM2000 BH	Values	Field survey BH broadleaved	conifer	urban
broadleaved	1000	750	100	150
conifer	500	50	425	25
urban	200	10	10	180
Calibrated output values		**810**	**535**	**355**

The LCM2000 proportions of each BH (Values) are multiplied by the fractional amounts in the calibration matrix to give output proportions which, if summed (bold), show how the same 1 km square might have been recorded by a comprehensive FS.

The above procedure is valid only where the calibration matrix is fully representative of the landscape within the 1 km square which is being calibrated. The NLCs range in extent from just over 800 km^2 to in excess of 15000 km^2. With a maximum of 30 FS squares per NLC, not all of the spatial variability in the landscape within an NLC will be present in it's calibration matrix. Also, some of the NLCs straddle a number of landscape types which can not be mixed in a calibration exercise such as this. For instance, NLC 8 is defined as 'Coastal, often estuarine, mainly pastures, otherwise built-up' and is found along the coast of The Wash and along the river courses that cross The Fens. Therefore a simple application of the calibration matrix for NLC 8 would produce coastal habitats along the river valleys of the Nene, Ouse and Welland.

To control the operation of the calibration matrix a number of knowledge-based corrections were developed (Figure 4). These can be divided into three groups based on their method of operation and the impact on the results.

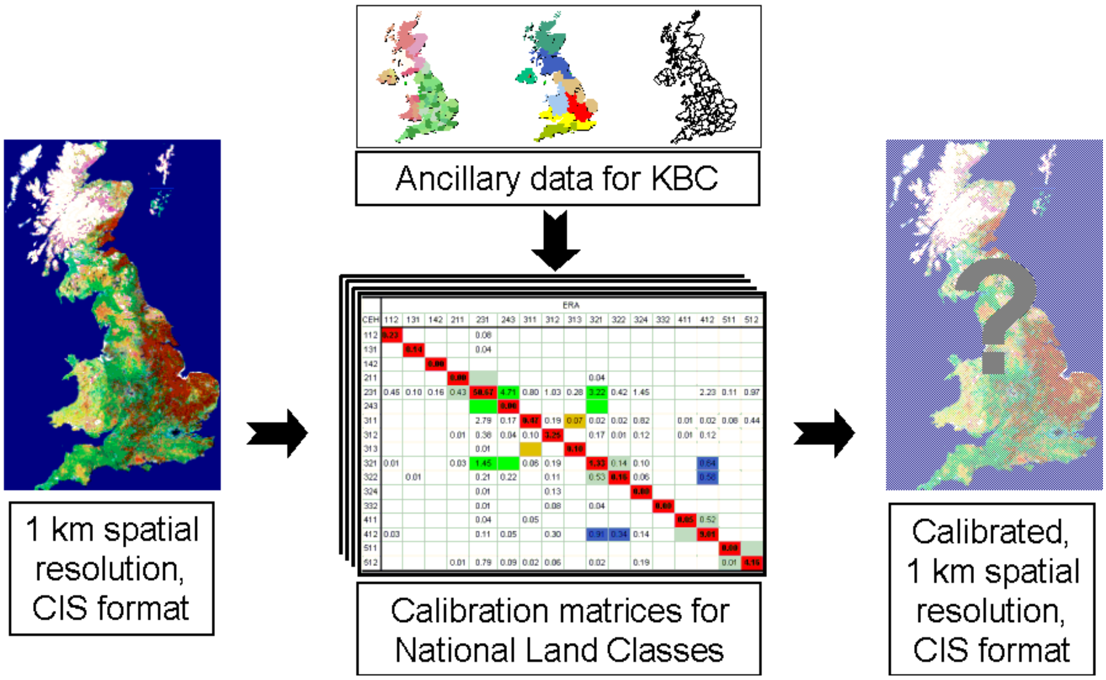


Figure 4. Calibration method to produce the calibrated 1 km data set from calibration matrices and ancillary data.

The simplest knowledge-based correction worked by disabling the calibration process completely where it worsened the results. For instance, the FS was not designed to map dense urban areas and the selection of FS squares specifically avoided areas with greater than 25 % urban. In this case, dense urban areas were therefore mapped optimally by the LCM2000 data without calibration.

The second type of knowledge-based correction related to 1 km squares where a particular class was known not to be present via some additional contextual information. In this case the output column of the calibration matrix for the class that was not present was set to zero to prevent any of it being produced by the calibration. As the input rows of the calibration matrix no longer sum to 1 it was necessary to re-normalise the calibration matrix to produce the correct total of the output proportions. This type of correction was used with the example of NLC 8 described above to prevent coastal habitats appearing in river valleys far from the sea.

The third type of knowledge-based correction related to a known mis-classification within a particular spatial context which the calibration matrix at the NLC level could not fully correct. In this case the column of values for the mis-classified class were combined with the column of values for the correct class. This correction was applied to grassland types where ancillary data, such as soils information, could identify which of the grassland types was correct.

To derive estimates of uncertainty for the calibrated BH stock estimates per 1 km square a bootstrapping approach was implemented. Rather than generate a single calibration matrix for each NLC, a range of different 'bootstrapped' calibration matrices were generated. The 'bootstrapped' calibration matrices were generated by randomly sampling the correspondence matrices within the NLC with replacement so as to give a constant number of samples per NLC. A thousand different calibration matrices were generated for each NLC in this way. These calibration matrices were then used to produce a thousand estimates of the stock of each BH within each 1 km square. The thousand estimates were then ranked and the 50^{th} and 950^{th} estimates extracted to give the upper and lower limits of uncertainty for each BH. To ease spatial visualisation of the results within this report, a single measure of uncertainty was then derived by halving the difference between the upper and lower limits.

Calibration development

The process for developing the calibration method was one of iteration (Figure **5**). The 1 km summary data set from LCM2000 Release 1 was the starting point. The calibration method was applied, the calibrated results produced and these results were validated. The results of the validation were then used to refine the calibration method and the process was repeated. Note that after the fourth iteration, the calibrated data set was sent out for peer review by members of the Module 9 Technical Advisory Group and other interested parties. Feedback from this process was incorporated into the final set of iterations.

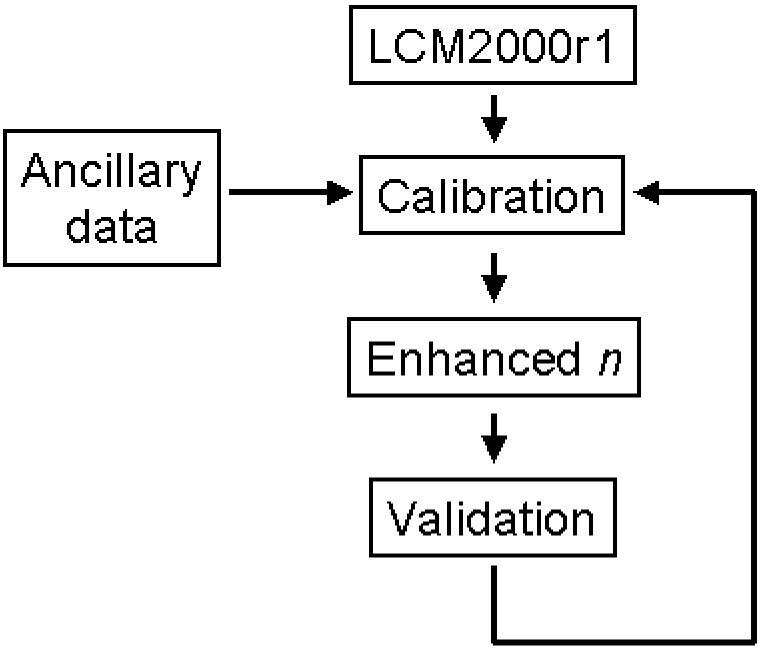


Figure 5. The iterative process for developing the calibration method.

In total there were seven iterations to produce the final version of the calibrated data set (see below for explanations of the ancillary data used):

1. Apply the calibration matrices without knowledge-based corrections,
2. As above, but with the calibration disabled in dense urban areas,
3. As above, but coastal habitats excluded outside the coastal zone,
4. As above, but montane habitats excluded outside of a montane mask and urban excluded within it,
5. As above, but with soil type data used in the calibration procedure to alter the calibration matrix for each NLC if one of the three soil types (acid, neutral or calcareous) was found to be dominant,
6. As above, but with ancillary woodland data used to prevent the creation of large amounts of woodland by the calibration procedure which altered the balance of broadleaf, mixed & yew and coniferous woodland,
7. As above, but with a calcareous grassland mask from the priority habitats added to complement the soils data.

ANCILLARY DATASETS FOR CALIBRATION & VALIDATION

The LCM2000 and the FS each have their strengths and weaknesses in generating land cover statistics, resulting from their methods of production. The key element to achieve a 'best fit' integration of LCM2000 and FS for the supply of land cover statistics was their comparison with consistent nationwide datasets of known quality. There are few such relevant datasets that are readily available for England with a 1 km or finer spatial resolution:

- National Soil Resources Institute (NSRI) National Soil Map for England and Wales (NATMAP1000),
- Ordnance Survey (OS) 1 km Geographic Reference Data for GB,
- Forestry Commission (FC) Digital Woodland Map for England,
- English Nature (EN) Grassland Inventory of England,
- EN Biodiversity Action Plan (BAP) Priority Habitat inventories,
- The June Agricultural Census.

In addition to these 1 km spatial resolution datasets on land cover (or in the case of NATMAP1000, a key environmental variable that can directly influence land cover) CEH also has a 10 km spatial resolution GB dataset on indicator species composition.

These datasets can potentially contribute to three important elements of the integration process: (i) providing an independent comparison for LCM2000 and FS estimates of stock, thereby identifying the strengths and weaknesses of each survey; (ii) providing an input to the calibration procedure where the calibration matrices per NLC are not sufficient to generate an accurate output; (iii) providing a means of validating the integrated 1 km summary dataset, if not used in the calibration procedure.

Soils data

The NSRI NATMAP1000 is a 1 km raster data set containing information on the dominant soil group per 1 km National Grid cell of England and Wales. The soil classification scheme is hierarchical, with 10 major soil groups which expand into 34 different soil groups (in total describing the composition and distribution of 300 soil associations). The current version of NATMAP1000 was launched in 2001 and more information is available from www.silsoe.cranfield.ac.uk/nsri.

The 34 different soil groups identified in NATMAP1000 were classified into calcareous, neutral, acid and peat soil types. This resulted in a map of the dominant soil type per National Grid 1 km cell, against which FS and LCM2000 estimates of BHs strongly influenced by soil base levels could be assessed. This included the semi-natural grasslands, dwarf shrub heath and bog, which can be distinguished much more readily by field surveyors than by the spectral reflectance characteristics in satellite imagery. For calcareous, neutral and acid grassland, each was plotted as a percentage of the total semi-natural grassland per NLC, against the percent cover of the appropriate soil type per NLC for both FS and LCM2000. Dwarf shrub heath was plotted as the percentage of the total of Dwarf Shrub Heath and Bog against the percent cover of acid soils per NLC for FS and LCM2000. Lastly, the percentage cover of bog was plotted against the percent cover of peat soils per NLC for FS and LCM2000 (see Appendix I).

OS 1km summary data

OS 1 km summary Geographic Reference Data for GB is available through the CIS and lists the percent cover for each 1 km National Grid cell of a range of features. Of relevance to the BH reporting framework of CS2000 are the classes: built up towns, built up villages, canals, inland water, rivers, foreshore, sand, and woodland. The OS Geographic Reference Data is dated as 1998.

The OS Geographic Reference Data provided a 1 km summary comparison for: Broadleaf, mixed and yew and coniferous woodland combined into a generic woodland class; Standing open water and canals by combining canals and inland water, Biult-up and gardens by combining built up villages and towns, and a generic 'coastal' habitat (BHs 18-21) by combining foreshore and sand.

In contrast to the comparisons made using the soils data which compared proportional coverage per NLC, with the OS dataset it was possible to make direct comparisons of land cover estimates. Of course, OS and LCM2000 data on land cover were both a complete census, whereas the FS land cover estimates were based on the proportional composition of samples within an NLC multiplied by the area of each NLC. However, this does not supply land cover statistics for the same spatial coverage as LCM2000 and OS since:

- the NLCs do not extent as far off shore as the LCM2000 and OS98 datasets,
- the FS statistics are not intended to be projected into National Grid cells with greater than 75% urban coverage.

Thus, the FS statistics were not extrapolated to cover as large a spatial area as LCM2000 and OS. This had obvious implications for the FS stock estimates for coastal and built up land covers. For all four land cover types examined, both the LCM2000 and FS estimate were plotted against the OS total per NLC (see Appendix II).

FC Digital Woodland Map for England

The FC Digital Woodland Map for England is a vector dataset based on interpretation of 1:25 000 aerial photography (flown in 1991-2000) and plotted against OS 1:25 000 mapping. Woodland parcels consist of areas of tree cover with a crown density of at least 20 %, with a minimum width of 50 m and a minimum size of 2 ha. Woodland classes are: coniferous, broadleaved, mixed, shrub, coppice and young trees. In addition, parcels also identify ground prepared for planting, felled woodland, and young trees. Data were updated by Woodland Surveys for the National Inventory of Woodland and Trees to include FC new planting and New Woodland Grant Schemes, as at 31^{st} March 2000.

The FC dataset is a complete survey of all woodland parcels > 2 ha size and all 136 286 polygons have a land cover label. This supplied overall woodland land cover statistics for England, which could be compared with LCM2000 and FS estimates. In addition, the vector data was compressed into proportional composition of the different 'woodland' classes and 'other' non-woodland cover per 1 km National Grid cell and combined as appropriate to match woodland widespread BHs identified in LCM2000 and FS.

In the FC dataset, the classes broadleaved, mixed, coppice, and shrub were combined into a broadleaved & mixed woodland BH, whilst the classes conifer and young trees were combined into a coniferous woodland BH. Scatter plots were produced of the LCM2000 and FS estimate of broadleaved & mixed woodland and coniferous woodland against the FC total for the 21 NLCs of England (see Appendix III).

Grassland Inventory

The EN Grassland Inventory of England is a vector dataset of semi-natural lowland grass communities (i.e. sites of enclosed grassland occurring at or below 300 m above sea level). The inventory is based on a range of sources, but the Phase 2 level surveys form the bulk of the data used, with sites mapped at the 1:50 000 level. Other data sources include surveys carried out by organisations such as Wildlife Trusts and Local Authorities. Both statutorily designated sites (e.g. National Nature Reserves and Sites of Special Scientific Interest) and undesignated sites are included within the inventory. Grassland sites were included according to the following criteria: high botanical diversity; post-1980 survey data; minimum size of 0.5 ha (100 m length for linear features); and information which is easily located and held as a readily accessible record. Only semi-natural communities were included in the inventory: neutral, calcareous and acid grassland, Calaminarian grassland (metallophyte vegetation), fen meadows and rush pastures, and selected swamp mire and mire communities. Maritime cliff grassland, salt marsh and sand dune grasslands were mostly excluded from the definition, as were improved or semi-improved grassland communities. These data were collected from sources spanning a 16 year timescale and consequently some sites in the inventory will have been lost to agricultural improvement or development.

The Grassland Inventory is not a complete measure of the extent and distribution of lowland grassland communities in England since the coverage of the Phase 2 grassland surveys were not comprehensive and some grassland types (e.g. calcareous grassland) have better coverage than others (e.g. acid grassland). Additionally, of the 8109 parcels making up the dataset, 481 have no information on grassland types present (i.e. are ‘empty polygons’ in which grassland type was not designated). This includes substantial areas of grassland at Salisbury Plain, Teesdale and the Brecklands. Furthermore, 25 % of the parcels list more than one grassland type, but the proportional composition is not stated. The EN Grassland Inventory could thus provide a dataset for comparing or validating the land cover composition of sample areas from LCM2000 or FS for which overlapping data exist. However, since the EN Grassland Inventory is not a complete census, has unlabelled parcels and the sampling strategy was not systematic, then this dataset was not particularly useful for comparing the land cover statistics per NLC or as input to the calibration procedure.

BAP Priority Habitat inventories

The Grassland Inventory will soon be superseded by Biodiversity Action Plan (BAP) Priority Habitat inventories. In total, 23 BAP Priority Habitat inventories will be available for England, which includes seven grassland habitats. These data are currently being quality assessed by EN before release in April 2004 via the website http://natureonthemap.org.uk. The dataset will still contain no information as to the location or proportional coverage of each Priority Habitat within the boundaries of designated areas. Nonetheless, compared with the Grassland Inventory, the dataset will be more of an exhaustive survey and will contain habitat labels for all parcels.

This dataset was not fully available from EN prior to public release in April as a QA procedure was still being undertaken. However, the inventories for lowland and upland calcareous grassland were acquired and these were simplified to recorded presence or absence per 1 km National Grid square and used as a spatial mask in the calibration procedure for semi-natural grasslands.

June Agricultural Census

The June Agricultural Census is an annual census of agricultural activity conducted by a postal questionnaire which collects information from farm holdings including land use, crops, livestock, and horticulture. In 9 years out of every 10, the Census is in fact conducted as a sample survey (1998 was a sample year). A stratified random sampling approach was adopted in which holdings are divided into groups (strata) on the basis of their economic size, with higher sampling rates being used in the larger strata. National and local figures were then estimated by Defra based on the data received.

The June Agricultural Census data for 1998 were used to provide a statistical comparison of arable & horticulture and improved grasslands, as recorded in FS and LCM2000. The Census data from 1998 have been made available from the CQC project and have been processed into Countryside Character Areas. This dataset thus offers the ability to examine the validity of the calibrated product for the two most prevalent English BHs and for smaller geographic units than to which the calibration process was applied.

Indicator Species

The Biological Records Centre (BRC) at CEH Monks Wood has a GB dataset recording species composition in a 10 km x 10 km grid. Preston et al (2003) identified all species associated with BAP BHs and identified the percentage of acid, calcareous and neutral species of the GB total in each grid square. Therefore, the score of each grid cell could be up to 300 %. This dataset has been used in the Critical Loads project (see http://critloads.ceh.ac.uk) to refine the LCM2000 data on semi-natural grassland identification. They used a cut-off of 50% to determine a 10 km square of calcareous grassland (i.e. a grid cell where at least 50 % of the calcareous species pool for GB is present). They used a cut-off of 40 % for acid and neutral grasslands. The nature of the dataset and these rules means that grid cells can be deemed to have more than one type of semi-natural grassland (in some cases all three). No information is provided in the dataset as to where within a 10 x 10 km grid cell the areas dominated by acid, neutral or calcareous indicator species occur, or indeed of how much of a 10 x 10 km cell could be deemed as being acid, neutral or calcareous grassland. As a result, these data were not used for calibration or validation purposes.

Results of comparisons of FS and LCM2000 stock estimates with ancillary data

Comparisons of land cover statistics for the generic 'coastal' and 'built up' land cover types derived from FS and LCM2000 with OS estimates clearly demonstrate that LCM2000 offers by far the better set of statistics. This occurred for several reasons. Firstly, LCM2000 can map further off-shore than FS and maps the percentage cover of all National Grid 1 km cells, regardless of the percentage urban content. This obviously has the greatest impact on the coastal and built up land cover types respectively. Secondly, even for the areas of overlap in both FS and LCM2000, LCM2000 provides better estimates of total coverage for coastal and built up land cover types. FS apparently over-estimates the cover of both coastal and built up land classes. In the case of coastal land classes, it is perhaps not surprising that the FS statistics over-estimate the total land cover, when the estimates are based on the area of each NLC rather than the length of coastline. There is not likely to be a relationship between the length of coastline and the area of hinterland for each NLC. The over-estimation of urban total land coverage in FS probably relates to the way in which urban areas are mapped by the FS. The whole area within an urban boundary is classed as built up, irrespective of whether there are open spaces, woods, lakes etc. These features are all identified in LCM2000 if above the MMU of 0.5 ha. Thus, it was concluded that it would be detrimental to the quality of the LCM2000 statistics to use FS data to calibrate the **total coverage** per NLC of 'built up' or 'coastal' land classes. However, using the calibration procedure to re-distribute the 'coastal' land class between littoral and supra-littoral BHs is worth attempting, since these are not well distinguished in LCM2000.

The comparison of land cover statistics for inland water derived from FS and LCM2000 with OS estimates also demonstrated that LCM2000 offers by far the better set of statistics. This almost certainly relates to the fact that water bodies are amongst the most readily mapped 'land' cover types as water has very identifiable spectral reflectance characteristics compared with vegetation, bare soil or artificial surfaces. The poorer statistics for FS in this particular case reflects the distinction between extrapolated sample data and well identified census data. Thus, once more it was concluded that it would be detrimental to the quality of the LCM2000 statistics to use FS data to calibrate inland water.

The difficulties in mapping deciduous woodland in single date or two-date imagery acquired early or late in the growing seasons (as occurred for considerable areas of LCM2000 production) resulted in poor land cover statistics for this land cover class compared with FS statistics. Calibration of LCM2000 broadleaved & mixed woodland and coniferous woodland using FS statistics was identified as beneficial. This process was improved by using the FC Digital Woodland Map for England to provide a spatial mask to maintain the within-NLC spatial distributions.

The inadequacies of the soil sensitivity and drift maps used in the post-classification knowledge-based correction procedures in LCM2000 production are highly apparent in the comparisons of FS and LCM2000 cover statistics with soil type. FS estimates per NLC are much closer to what would be predicted based on soil type, especially in the case of neutral and acid grasslands. Thus, calibration of the LCM2000 estimates for the semi-natural grasslands, dwarf shrub heath and bog made use of FS statistics and the soil type data as a mask to maintain spatial distribution patterns of these BHs within NLCs. Acid grassland and dwarf shrub heath both occur on acid soils and the comparisons with FS data showed that mis-classification occurs between these two BHs in LCM2000. Whilst the soil type data were used to correct for mis-classification in LCM2000 between these two BHs and calcareous or neutral grassland and bog, they could not be used to correct mis-classification between acid grassland and dwarf shrub heath.

For the more rare BHs of bracken, fen marsh & swamp, inland rock and montane, no consistent nationwide datasets are currently available with which to compare the land cover estimates of FS and LCM2000. In the case of these relatively rare, often isolated or fragmented, and, in the case of bracken, temporally variable BHs, both sample-based field survey and satellite-based land cover mapping struggle to generate meaningful nationwide statistics. Under these circumstances it must be asked whether either survey technique should be expected to provide robust land cover statistics for these BHs at a range of spatial scales, and whether the calibration of these land cover statistics by integration is a worthwhile aim. Finally, the two most prevalent land cover types in England, arable & horticulture and improved grassland are known to show mis-classification in LCM2000 when compared with FS. This results from rotation farming which is picked up when LCM2000 satellite images were not acquired in the target period coinciding with the timing of FS. Because these two BHs are so abundant in all NLCs in England, the stock estimates of FS were considered statistically representative to be used for calibrating the LCM2000 estimates.

THE CALIBRATED 1 KM SUMMARY DATASET

The calibrated product

The calibrated product has been released as a GB mode, 1 km census file for the CIS (version 7). For each of 20 BHs, there is an individual data layer showing the total coverage in hectares per km^2 (Examples can be seen in Figure 6 or the data is available through the CIS web site).

An obvious distinction between LCM2000 (Release 1) and the calibrated product (Iteration 7) for the terrestrial BHs is a shift away from areas with 0 % coverage to a smoothing effect whereby a background level proportional cover is visible across NLCs. As a result, the boundaries of NLCs are often highly apparent in the calibrated dataset, replacing any satellite image boundaries that were present in LCM2000.

The calibration procedure not only has the effect of 'smoothing' the spatial distribution of BHs across the NLC zones, but also of removing features that are anomalous within a zone or poorly sampled by the FS squares within that zone. This reflects the fact that the FS was designed to be representative for generating land cover statistics at the national or regional level, rather than for precise spatial mapping of land cover within NLC zones. The most notable example of this is Salisbury Plain, which is not sampled by a FS square and calcareous grasslands in this NLC are under-represented.

Uncertainty of calibrated stock estimates

The uncertainty estimates have been released as two GB mode, 1 km census files for the CIS (version 7) to complement the calibrated stock estimates. For each of 20 BHs, there is an individual data layer showing the ± coverage in hectares per km^2 that represents the 95 % confidence limits.

On first examination the uncertainty information appears very similar to stock information and contains many of the same complex spatial structures. This is not surprising as it was derived from multiple realisations of the calibration procedure to estimate stock. To interpret the uncertainty information effectively it must be considered in the context of the calibrated stock estimates.

By way of illustration the stock and a single measure (as described earlier) of uncertainty for three BHs will be compared to identify the main features of which a user should be aware (Figure 6). For the acid grassland BH the stock and uncertainty appear almost identical in relative proportions and pattern. The black areas represent very small uncertainties in areas where there is little if any of the BH present. This demonstrated the fact that for most BHs there is a positive relationship between stock and uncertainty. Closer examination does show some interesting features, such as the relatively high uncertainty associated with this class for the moorlands of the southwest. The broadleaf, mixed and yew woodland BH shows a very similar relationship between stock and uncertainty to the acid grassland BH, but the relatively uniform distribution of this BH across England has allowed the effects of the NLC to become apparent. The sampling of the BHs by the FS 1 km squares within each NLC will impact on the uncertainty estimate. The stock and uncertainty estimates for the arable & horticulture BH show the case where in some NLCs they are inversely proportional. When considering the Midlands and East Anglia, the uncertainty is smaller where the stock is larger. This demonstrates the success of the FS, LCM2000 and calibration procedure at mapping this BH in these areas. There are a number of interesting anomalies, for instance the East Riding of Yorkshire and Thetford Forest which have small stock values but large uncertainties. In the Thetford case the sampling of the FS in this region may not be representative of the wider landscape.

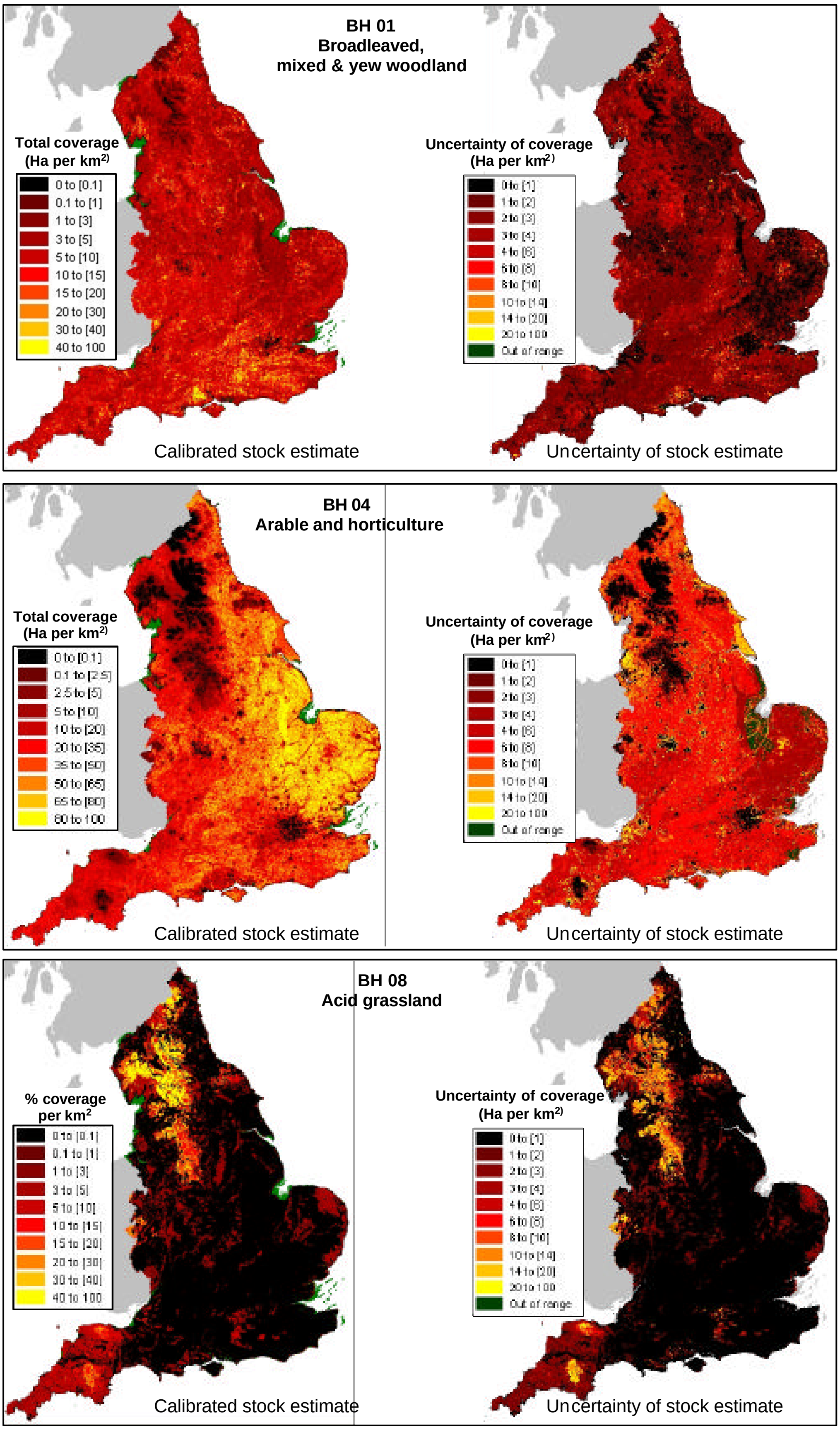


Figure 6. Stock and uncertainty estimates from the calibration procedure for three sample Broad Habitats.

Estimates of land cover stock for England & the English Regions

The total stocks of the 19 BHs in England, as estimated from extrapolated FS statistics, census LCM2000 data and the calibrated 1 km summary product, are shown in Table 3. Similar tables breaking the England statistics down into The English Regions are given in Appendix IV. The statistics were extracted from the three different datasets, using the country and region boundaries provided in the CIS.

The lower and upper confidence intervals for the FS statistics were derived using two standard errors of the mean to estimate the 95-percentile range. For the calibrated product, the upper and lower limits of uncertainty represent the 50th and 950th ranked estimates of stock from one thousand different bootstrapped calibration matrices.

	Field Survey			LCM2000	Calibrated 1 km data		
		Bias corrected confidence interval				Bias corrected confidence interval	
Broad Habitats	**Mean**	**Lower**	**Upper**	**Mean**	**Mean**	**Lower**	**Upper**
Broadleaved, mixed and yew woodland	10121.2	8611.2	11631.2	10929.4	9533.8	6350	12720
Coniferous woodland	2985.3	1807.3	4163.3	2979.8	2888.0	1410	4560
Arable and horticulture	46359.0	42447.0	50271.0	48393.6	46823.4	38870	55460
Improved grassland	36864.9	33548.9	40180.9	32017.0	39479.9	31420	47670
Neutral grassland	4033.8	3115.8	4951.8	5003.3	3472.3	1710	5480
Calcareous grassland	353.8	23.8	683.8	7849.1	1120.5	550	1630
Acid grassland	3819.4	2793.4	4845.4	2784.6	4048.5	2280	6090
Bracken	1658.7	1014.7	2302.7	701.0	1398.2	590	2630
Dwarf shrub heath	3621.8	2305.8	4937.8	2650.3	2888.6	1190	4780
Fen, marsh and swamp	1492.4	926.4	2058.4	179.6	1237.2	600	2330
Bog	983.4	429.4	1537.4	1054.8	1061.2	320	2280
Standing open water and canals	895.9	19.9	1771.9	588.6	534.3	250	880
Montane habitats	9.7	-6.3	25.7	0.0	0.4	0	0
Inland rock	120.3	46.3	194.3	1101.5	169.1	40	290
Built up areas and gardens	10707.9	8721.9	12693.9	13807.5	15665.9	7100	16730
Supra littoral rock	140.2	62.3	218.3	0.6	158.7	80	260
Supra littoral sediment	233.2	13.2	453.2	105.0	379.7	200	510
Littoral rock	0.0	0.0	0.0	16.6	0.0	0	0
Littoral sediment	1181.7	345.7	2017.7	1117.4	891.3	650	1060

Table 3. The stocks of Broad Habitats for England, as recorded by FS, LCM2000 and the calibrated product.

The mean stock estimate of the calibrated product is within the lower and upper confidence limits of the FS estimates for all BHs, except calcareous grassland and built up & gardens. By contrast, the LCM2000 mean stock estimate was outside the 95 % confidence limits of the FS estimates for 13 of the 19 BHs (see Table 2). It is not surprising that the calibrated estimate of Built up & gardens is significantly greater than the FS estimate, since the FS value does not include any 1 km grid cell that is > 75 % urban. In the case of calcareous grassland, the fact that the calibrated stock estimate is not within the 95 % confidence limits of the FS mean, highlights the issue that FS under-represents calcareous grassland. The mean stock estimate for each BH, as derived from FS, LCM2000 and the calibrated product is shown in Figure 7.

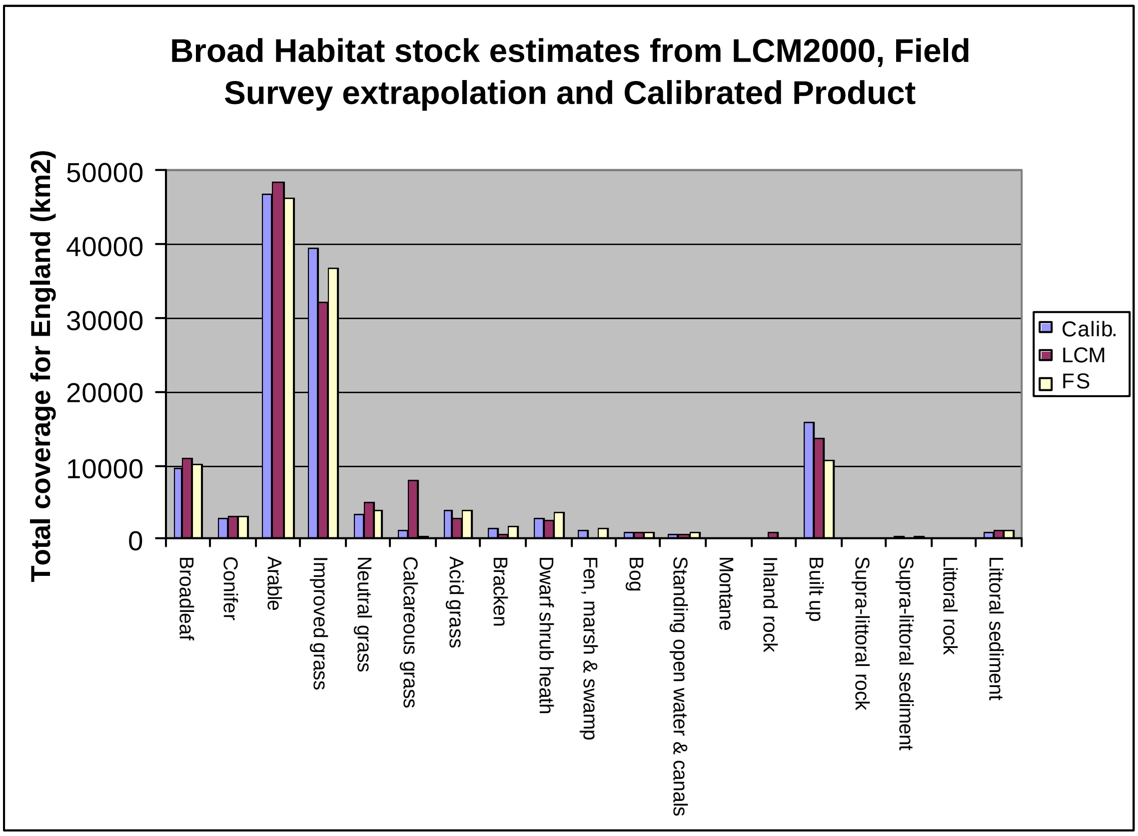


Figure 7. The mean stock estimate for each BH, as derived from FS, LCM2000 and the calibrated product.

Validation of the calibrated product

A visual QA and discussions with field ecologists at CEH Monks Wood has led to the general conclusion that the calibrated product is in almost all cases better than the original 1 km summary LCM2000 dataset. The two obvious cases of where the calibrated product is worse that the original LCM2000 1 km summary product is for standing open water & canals/rivers & streams and littoral sediment; the former loses its geographical distribution as linear features, the later is reduced in area by the restricted off-shore limits of the NLCs. It has to be accepted that the calibration procedure:

(i) often results in a very low (i.e. 1-2%) cover of Broad Habitats where they should not be expected to occur;

(ii) causes a general redistribution of land cover across the landscape (constrained by the spatial boundaries of NLCs);

(iii) occasionally removes features that are anomalous within an NLC or poorly sampled by the FS squares within that NLC.

In terms of statistical analyses, the first task was to compare the LCM2000, FS and calibrated product land cover estimates per NLC. The plots in Appendix V show LCM2000 against FS, LCM2000 against calibrated, and FS against calibrated stock estimates per NLC for the 19 BHs. In all cases the fit of either LCM2000 or FS against the calibrated stock estimates is better than fit of LCM2000 against FS. This is particularly so for the relationship of FS stock estimates against the calibrated stock estimates.

Broadleaf, mixed & yew and coniferous woodland validation of the calibrated stock statistics was achieved by comparison with the FC Digital Woodland Map of England. The total stock estimates for England, as derived from FS, LCM2000, the calibrated product and FC are shown in Table 4. All three CS2000 datasets apparently under-estimate the total coverage of coniferous woodland in England (by 613 - 708 km^2), but over-estimate the total coverage of broadleaved & mixed woodland (by 3 024 - 4 420 km^2).

	FS*	**LCM2000**	**Calibrated**	**FC**
Broadleaf & mixed woodland	9 970	10 930	9 534	6 510
Coniferous woodland	2 980	2 980	2 888	3 593

* Excludes National Grid cells with > 75% urban coverage

Table 4. Total coverage of woodland (in km^2) in England as reported in CIS by FS, LCM2000, and the calibrated product compared with FC data.

Appendix VI plots the LCM2000, FS and calibrated estimate of broadleaved & mixed woodland and coniferous woodland against the FC total for the 21 NLCs that cover England. For coniferous woodland the estimates of LCM2000 and the calibrated product lie much closer along the $y = x$ line than for the FS estimates. For broadleaved & mixed woodland, all three datasets show a tendency to over-estimate the total coverage in almost every NLC; although the scatter is least for the calibrated estimates. This apparent over-estimation of broadleaved & mixed woodland may relate to the MMU of the FC data (2 ha) compared with LCM2000 (0.5 ha) and FS (0.04 ha). By default the calibrated product also has a MMU of 0.5 ha.

For arable & horticulture and improved grassland validation of the calibrated stock statistics was achieved by comparison with the June Agricultural Census data for 1998. The plots in Appendix VII show the LCM2000, FS and calibrated estimate of arable & horticulture and improved grassland against the June Census total for 158 of the Countryside Character Areas (CCAs) that cover England. For both BHs, the LCM2000 estimates have the least scatter about the $y = x$ line, whilst the FS estimates have the greatest scatter. This reflects the inappropriate spatial scale of estimating land cover from the FS for areas as small as the CCAs. By contrast, the relative lack of scatter about the $y = x$ line in the calibrated dataset demonstrates the robustness of the calibration procedure for deriving land cover estimates at a smaller spatial scale than the NLCs.

For standing open water & canals, built up & gardens and a generic coastal class (composed of BHs 18-21) the validation of the calibrated stock statistics was achieved by comparison with OS 1 km Geographic Reference data. Appendix VIII plots the LCM2000, FS and calibrated estimate of these three land cover classes against the OS total for the 21 NLCs that cover England. A similar pattern emerges as with the previous example, that the LCM2000 estimates have the least scatter about the $y = x$ line, the calibrated product is similar, whilst FS estimates have the greatest scatter. The total stock estimates for England, as derived from FS, LCM2000, the calibrated product and OS are shown in Table 5.

	*FS	LCM2000	Calibrated	OS
Inland water	890	590	534	500
Coastal	1 560	1 900	1 430	2 000
Built up	10 420	13 800	15 666	13 700

* FS has a restricted spatial coverage compared with LCM2000, calibrated product and OS

Table 5. Total coverage of land cover types (in km^2) in England as reported by FS and LCM2000, compared with OS98 data.

For inland water, the calibrated product offers the most similar estimate for the England total to the OS data. The FS estimate is too high by 390 km^2 compared with the OS data, LCM 2000 estimate by 90 km^2, and calibrated estimate by 34 km^2. This reflects the fact that the sampling density of FS squares in England was not sufficient to represent adequately the spatial coverage of inland water bodies. Nevertheless, the process of calibrating stock estimates based on LCM2000 and FS land cover statistics has clearly been a beneficial process.

By contrast to inland water, the calibrated product offers the least similar estimate to the OS data for the England total of coastal and built up land cover types. LCM2000, FS and the calibrated product all apparently under-estimate the total England coverage of coastal land cover compared with OS; by 100 km^2, 440 km^2 and 570 km^2 respectively. This reflects the restricted off-shore area for which these coastal classes are calculated using the NLCs. For built up it is noticeable that whilst the total England estimate for LCM2000 is within 100 km^2 of the OS total; FS apparently under-estimates the total by 3 280 km^2, whilst the calibrated product apparently over-estimates the total by 1 966 km^2. This in fact relates to the restricted urban mapping that the FS supplies. However, we can calculate from the urban mask used in the CIS that 5 038 1 km National Grid cells which have an urban coverage of greater than 75 % are absent from the FS estimate of built up. As these 1 km cells must have an urban composition of between 75 % and 100 %, we can calculate the total urban estimate of England for FS to be in the range of 14 200 km^2 to 15 458 km^2. This is an over-estimate compared with the OS data of 500 km^2 to 1 758 km^2. This results from the FS method of placing a boundary around any built area and assigning a land cover of 'built up' irrespective of the actual variations in land cover within that built up boundary. Because the calibration procedure retained LCM2000 statistics for the areas under the FS urban mask, but used FS statistics which over-estimate urban coverage for the rest of England, this explains the apparent over-estimation of the built up class in the calibrated product.

PATTERN

Detailed pattern analysis

Landscape structure refers to the individual features or building blocks which make up a landscape, whilst landscape pattern refers to the spatial configuration of those building blocks. LCM2000 data provides a measure of landscape spatial structure identified in land parcels which were based on spatial boundaries in spectral reflectance characteristics of the Earth's surface recorded in satellite imagery. As a result, the landscape boundaries identified in LCM2000 often do not match those identified in the field. Statistical measures of landscape structure and pattern derived from LCM2000 and FS will thus be different, as the two datasets identify different building blocks. However, at a more aggregate level (i.e. combining neighbouring parcels of the same land cover type) LCM2000 and FS can show a more similar structure and pattern in landscape patch dynamics. In other words, the overall picture may be similar, but the breakdown of patch dynamics into individual parcels may be very different. Therefore, measures of landscape structure and pattern derived from LCM2000 and FS will vary depending on whether the focus is on the parcel or patch level. Which of these takes the focus depends largely on whether the interest is on land cover areas or the boundaries of landscape features.

In LCM2000, a single large field containing a crop with a variable growth pattern may be subdivided, whilst a landscape composed of several small fields of the same land cover type and condition may be amalgamated as a single feature. Examples of both of these can be seen in Figure 8, a FS square in Wiltshire. In enclosed landscapes, the parcel boundaries of LCM2000 are generally (but not exclusively) less reliable than the FS boundaries. In Figure 8, the FS identifies 86 parcels (either complete or in part) and a total of 15.7 km of boundaries. Of these boundaries, only 8.7 km were labelled by field surveyors as representing a real object on the ground (such as a hedge or wall). It is clear by comparison with the high resolution remotely-sensed image in Figure 8 that much of the linework separating land cover parcels in this FS square is over-segmenting the landscape. LCM2000 identifies 59 parcels (again, either complete or in part) within this 1 km square, and there are a total of 18.3 km of boundaries. The satellite mapping procedure did not attempt to assign a land cover label to parcel boundaries, as landscape features such as hedgerows and walls were below the MMU.

In the example (Figure 8), the FS and LCM2000 thus show different landscape boundaries. The two mapping approaches identify different shape and size landscape parcels, and boundaries of different lengths and locations. Clearly there is some mis-classification in LCM2000 between arable, inland bare and built up areas (resulting from early summer imagery in which much of the arable land was bare soil and thus difficult to distinguish from other non-vegetated surfaces). Also, the effects of field rotation are apparent, with different fields showing improved grassland and arable & horticulture in FS and LCM2000. Nevertheless, in this example FS square, FS and LCM2000 identify the same basic landscape structure. Appendix IX shows six sample FS squares from an enclosed landscape (two examples each from arable, pastural and marginal landscapes) mapped by FS and in LCM2000. These reinforce the above conclusions that at a general level LCM2000 identifies landscape structure, but the boundaries themselves are not necessarily meaningful or comprehensive in the way that FS would achieve.

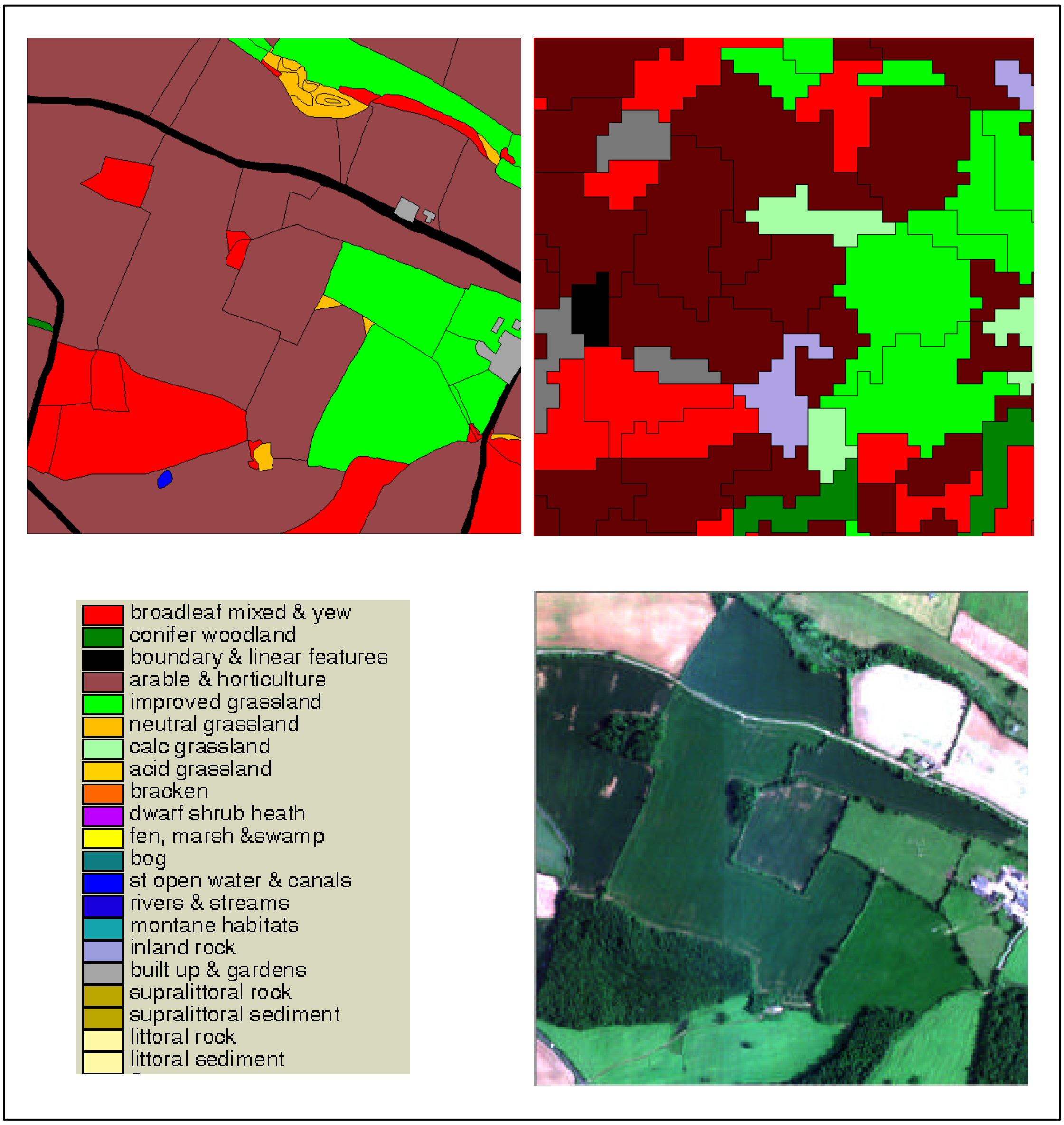


Figure 8. An example FS square (in Wiltshire) showing how this area appears from an aircraft and how it was mapped by FS (top left) and LCM2000 (top right).

The FS contains information on boundary type and length, all of which is in relation to land cover type and area. The calibration procedure has produced an integrated product on land cover type and area per km^2 based on FS and LCM2000 statistics per NLC. It should therefore in theory be possible to achieve a similar product based on boundary length from FS and LCM2000 within a stratification such as the NLCs. This could potentially be performed in relation to the land cover statistics of the calibrated stock estimates, so that boundary length calibration was carried out appropriately to context. An issue that would have to be considered here is that LCM2000 data offer complete spatial coverage, whilst the FS squares are 1 x 1 km extracts. Thus, it is possible to extract LCM2000 data selecting all parcels that intersect an area (giving a set of parcels that extend beyond the 1 km survey square boundaries). The FS data however are trimmed down to a 1 x 1 km area only. As a result many of the parcels in the FS data are truncated at the FS square boundary, thereby placing artificial boundaries on the landscape. In a recent study carried out at CEH Monks Wood

(Swetnam, unpublished data) it was found that for 23 FS squares spanning the English and Welsh NLC, on average only 40 % of parcels (range 15-73 %) were entirely contained within the 1 km square.

In unenclosed (semi-natural) environments, it could be argued that neither FS nor LCM2000 provide boundaries that are meaningful or reliable. This is shown in Figure 9, for two FS squares in an upland environment. Here, boundaries are highly subjective as land cover types tend to form mosaics or ecotones.

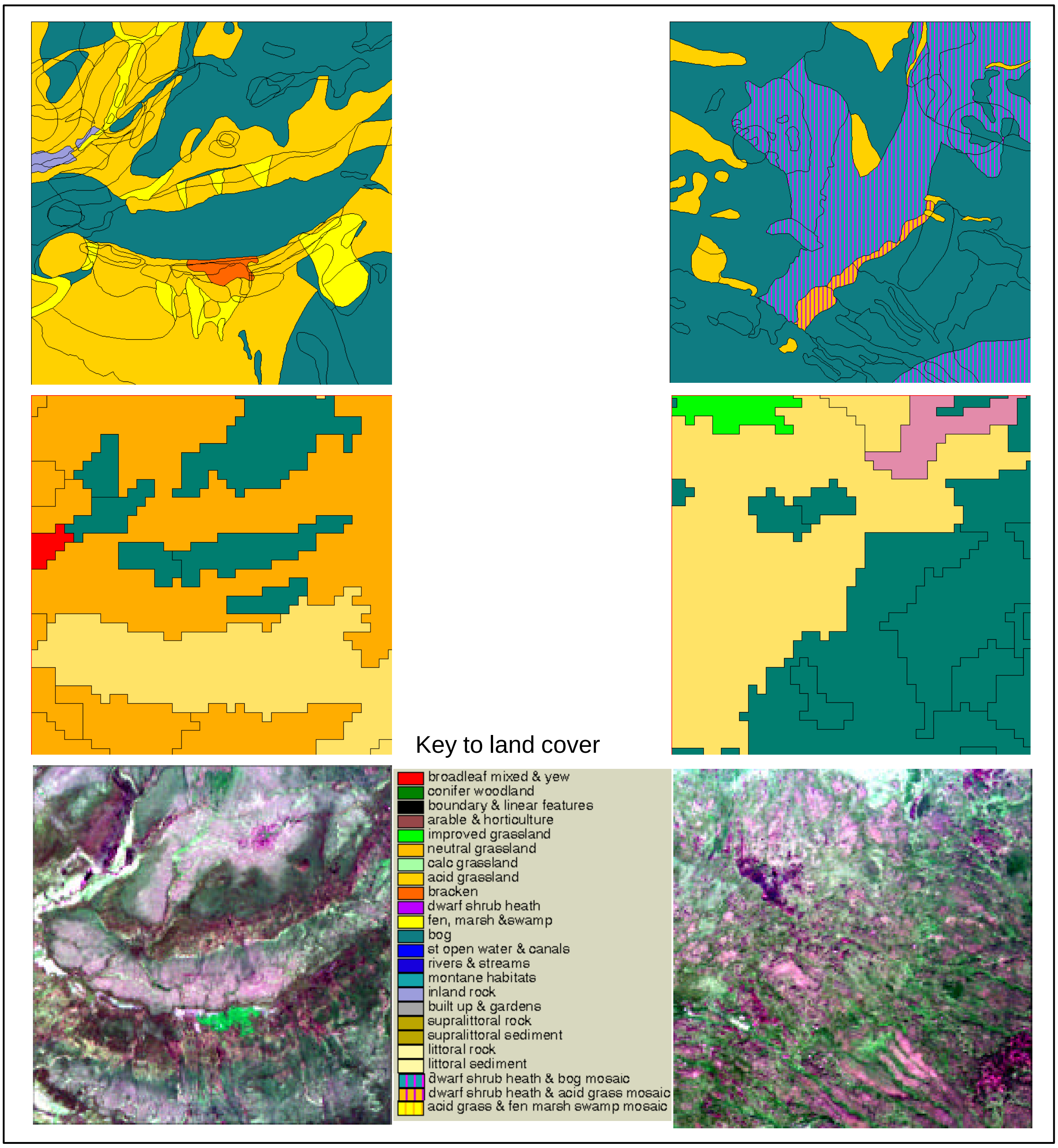


Figure 9. Two sample of upland FS squares showing how this area was mapped by FS (top) and LCM2000 (middle) and how it appears from an aircraft.

From the discussion above it becomes apparent that the calculation and comparison of detailed pattern metrics from FS and LCM2000 are fraught with difficulty. Area metrics would be influence by the truncation of parcels at the edges of FS squares, the different MMUs of the FS and LCM2000 and the pixelated nature of the LCM2000.

Line features in the FS are accurate mappings of real surface features whereas the lines in LCM2000 are locations where the spectral contrast of the image data was sufficient to halt the growth of an image segment. These factor together with the pixelation of LCM2000 make line metrics unreliable. Shape metrics will be influenced by the differing MMUs, the FS mapping of linear features narrower than the pixel size of the original LCM2000 imagery and the pixelation of LCM2000 parcel boundaries.

1 km pattern metrics for England

Landscape structure and pattern are important attributes that may describe countryside character. Therefore it has been important to derive meaningful landscape pattern metrics at a 1 km scale for all of England that have been supplied to the CQC project. Metrics such as the Simpson's Index of Diversity, the Evenness Index, and the Bray-Curtis Similarity Index have been computed for England using the proportional composition of BHs per 1 km square. These pattern metrics are not affected by issues of how landscape patches or mosaics are defined, how well landscape features are identified as individual objects by satellite-based mapping, or the extent to which landscape features are truncated by field-based sampling using 1km squares. The pattern data have been released as a GB mode, 1 km census file for the CIS (version 7). There are four data layers, showing diversity, evenness, similarity and the number of BHs per km^2 (see Appendix X). Because the pattern metrics were calculated using the calibrated product to supply information on the number and proportional composition of BHs per 1 km square, the influence that the NLC spatial stratification had on the calibrated stock estimates is readily apparent in the pattern data.

The metrics were calculated as follows:

Simpson's Index of Diversity

$D = 1 / S\ p_i^2$

D = Simpson's Index, p_i = % cover of land cover types.

Simpson's Diversity Index (D) has a range of 1 to *lc* (where *lc* is the number of land cover types present in a 1 km^2). This Index gives more weight to common land cover types. *D = lc* when all land cover types are equally represented.

Evenness

Evenness = D / lc

D = Simpson's Index of Diversity, *lc* = number of land cover types

Land Cover Evenness has a range of 1/*lc* to 1. Evenness = 1 when all land cover types are equally represented

Bray-Curtis Similarity Index

$I_{BC} = 1 - S|x_i - y_i| / S (x_i + y_i)$

$\boldsymbol{I_{BC}}$ = Bray-Curtis measure of similarity, $\boldsymbol{x_i}$ = % cover of a land cover type in one square, $\boldsymbol{y_i}$ = % cover of the same land cover type in an adjoining square.

Bray-Curtis Similarity Index (I_{BC}) has a range of 0 to 1 (0 = no land cover types shared, 1 = all land cover types shared). The similarity score for each 1 km grid cell was calculated as the average similarity score for the surrounding eight squares. (Note that England was treated as a 'data island' and so the similarity scores along the Scottish & Welsh borders were not based on information of land cover across the border).

CONCLUSIONS

Through the process of meetings, presentations and email distributions the procedures developed and results produced by Module 9A have been communicated to a wide range of user groups. These procedures have in general been accepted and deficiencies identified for further considerations.

Ancillary data has been identified as a possible means of correcting deficiencies in the calibrated 1 km data set and validating the final results. A range of data sets have been obtained and extensive comparisons undertaken with FS, LCM2000 and the calibrated 1 km data set. These ancillary data sets have generated their own problems and issues with decisions necessary on quality, definition and coverage along with their suitability for a role within Module 9A.

Data sets useful for additional knowledge-based correction within the calibration process have been identified and suitable rule bases developed and tested. This has resulted in a total of seven iterations of the calibrated 1 km product to a point where it is in it's final form with respect to this project. The bootstrapping procedure has been implemented to provide uncertainty information at the 1 km square level.

A full validation has been undertaken using a number of the ancillary datasets for statistical comparison, by plotting BH totals per NLC as derived from FS, LCM2000 and calibration, and by visual inspection of the 1 km calibrated maps in CIS. The calibrated product provides better stock estimates, and maps of their distribution, compared with the extrapolated FS dataset in all cases. However, there are three examples of where the calibrated product is worse that the original LCM2000 1 km summary product: standing open water & canals, littoral sediment and built up & gardens. Standing open water & canals loses its geographical distribution as linear features; littoral sediment is reduced in area by the restricted off-shore limits of the NLCs; and built up & gardens is over-estimated due to the unique situation for this land cover type in the FS.

In general, the calibration procedure often results in a very low (i.e. 1-2%) cover of BHs where they should not be expected to occur; causes a general background level of land cover across the landscape (constrained by the spatial boundaries of NLCs);

and occasionally removes features that are anomalous within an NLC or poorly sampled by the FS squares within that NLC.

A strategy for the measurement of landscape pattern has been developed and these values have been derived for England at a 1 km level. Metrics such as the Simpson's Index of Diversity, the Evenness Index, and the Bray-Curtis Similarity Index have been computed using the proportional composition of Broad Habitats per 1 km square. These pattern metrics are not affected by issues of how landscape patches or mosaics are defined, how well landscape features are identified as individual objects by satellite-based mapping, or the extent to which landscape features are truncated by field-based sampling using 1 km squares.

A brief investigation into the derivation of more detailed pattern metrics for each FS square has considered the various issues concerning object and boundary representation in FS and LCM2000. At a general level it has been shown that LCM2000 identifies landscape structure, but the boundaries themselves are not necessarily meaningful or comprehensive in the way that FS would achieve. In theory, it should be possible to achieve a calibration similar to the production of aerial estimates per 1 km square based on the detailed pattern information from FS and LCM2000 within a stratification such as the NLCs. Unfortunately, the problems of deriving comparable metrics from the FS and LCM2000 have made this impractical within the context of Module 9.

The results of Module 9A have been supplied to CQC and offered to the CIS user community as GB mode, 1 km summary datasets for CIS version 7.

Module 9A has resulted in a clearer understanding of the issues associated with LCM2000, FS, their integration and the resulting calibrated 1 km data set. In many ways the project has posed as many questions as it has provided answers. It has identified the strengths of each approach to landscape survey and attempted to maximise these. It has identified the weaknesses and devised mitigation strategies that compensate or minimise them.

From this analysis recommendation for future surveys can be extracted based on the specifications which are developed for their production:

1. Timing of surveys. Due to the dynamic nature of some of the BHs (e.g. agricultural and coastal) any time difference between the surveys should be minimised (especially within a stratum) or incorporated as a controlling factor in any correspondence analysis.
2. The nature of the two surveys results in very different MMUs. The MMUs of the surveys should be normalised (most likely to the largest) prior to any correspondence analysis to prevent features that could not exist at the coarser MMU being seen as error.
3. Differences in class definitions or in mapping protocol between the surveys (e.g. the different treatment of urban green space) should be avoided.
4. The rarity and typical patch size / shape should be included when considering the thematic class definitions (classes with limited extent compared to the largest MMU should be avoided).
5. A thematic classification should be devised which can be addressed equally by both surveys or with classes that can be separated by ancillary data.

6. Survey information from the coastal zone should be cropped to the highest tidal conditions found within the surveys.
7. The representativeness of the sampling strategy should be considered in the context of local calibration as well as national estimation (sufficient samples are required to address major within stratum features).
8. The normalisation of satellite images should be improved and the time frame of data collection restricted to make the spectral information more consistent across the survey area.
9. The use of ancillary data can not be avoided, but care should be taken to select only data sets with suitable thematic and spatial specifications, temporal similarity and appropriate uncertainty information.

Module 9A results should be used in association with the CS2000 Module 17 - Finding Out Causes and Understanding Significance (FOCUS) results with particular relevance for the planning of future Countryside Surveys.

REFERENCES

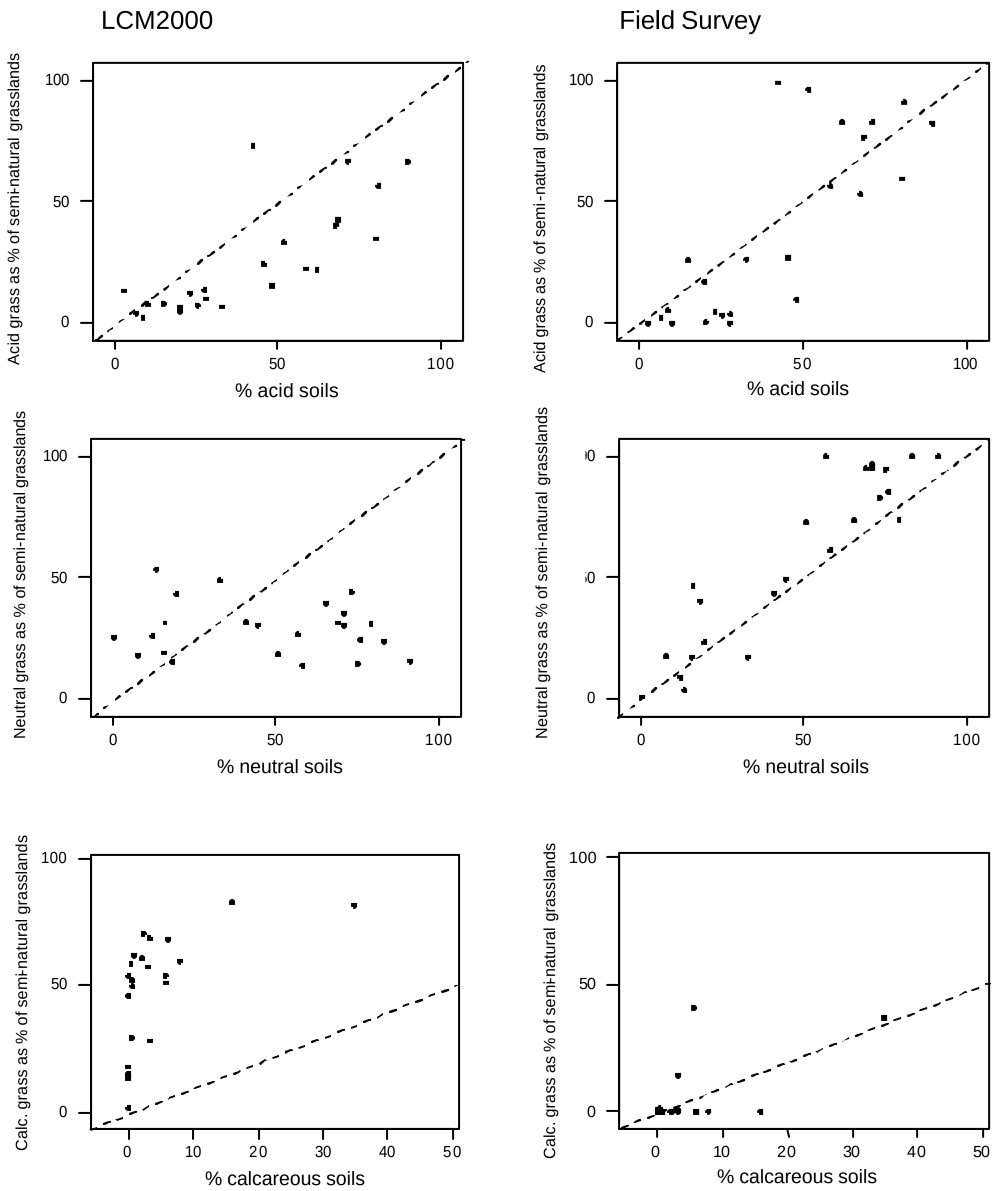
Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Cox, R., Brown, N.J., & Gerard, F.F. (2002) *Countryside Survey 2000 Module 7: Land Cover Map 2000. Final Report, CSLCM/Final.* Unpublished CEH report to DEFRA.

Preston, C.D., Telfer, M.G., Arnold, H.R., Carey, P.D., Cooper, J.M., Dines, T.D., Hill, M.O., Pearman, P.D., Roy, D.B., & Smart, S.M. (2003). The Changing Flora of the UK. (DEFRA, London).

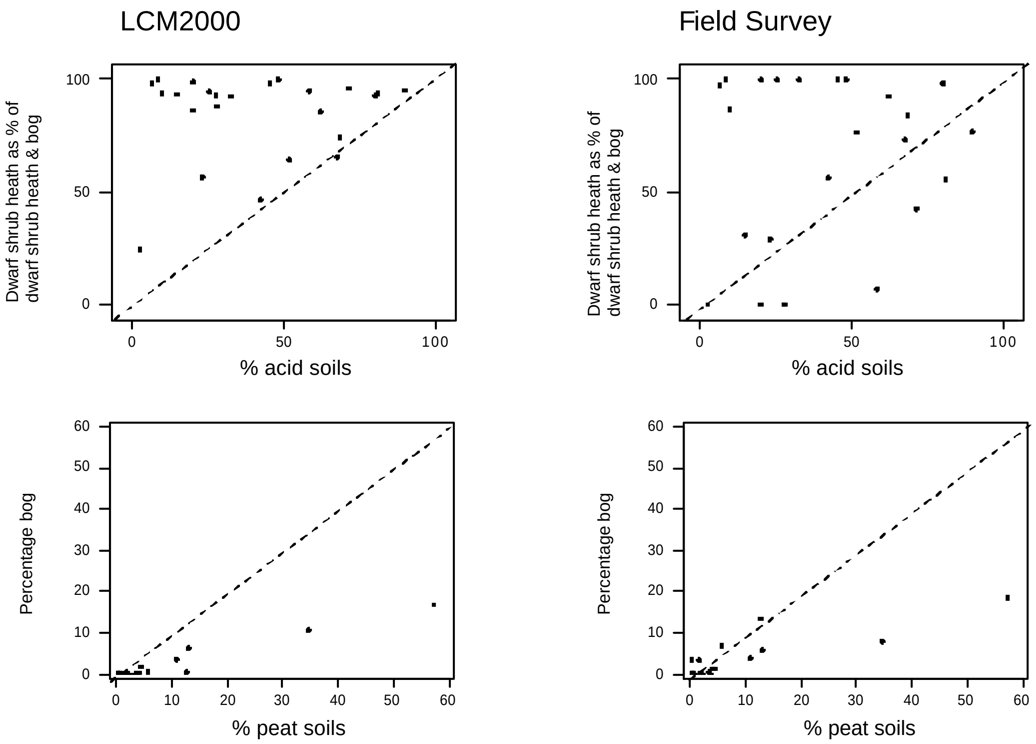
ACKNOWLEDGEMENTS

We are grateful to the National Soil Resource Institute, English Nature, Forestry Commission, Roy Haines-Young and Gemma Gregory at Nottingham University, and Dave Roy of CEH for providing the ancillary data used in Module 9A. We are also grateful to Ruth Cox, Jane Hall, Jackie Ullyett, Peter Rothery, Owen Mountford and Mark Hill of CEH Monks Wood for assistance at various stages of this project.

APPENDIX I
SCATTERPLOTS OF FS AND LCM2000 LAND COVER ESTIMATES AGAINST SOIL TYPE

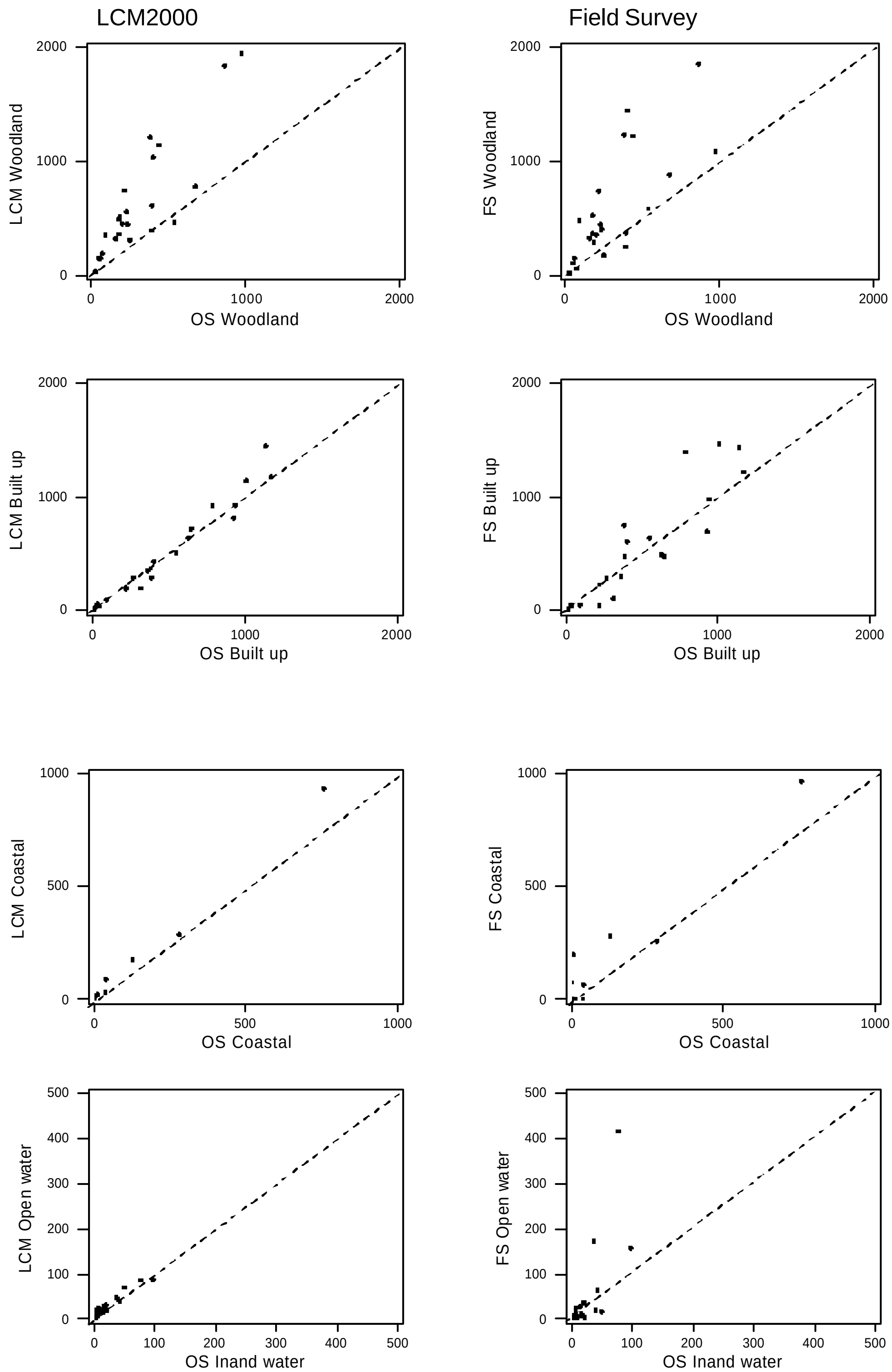


Scatter plots of calcareous, neutral and acid grassland as a percentage of the total semi-natural grassland against the percent cover of the appropriate soil type per NLC for LCM2000 and FS. (The $y = x$ line is also plotted for each graph).



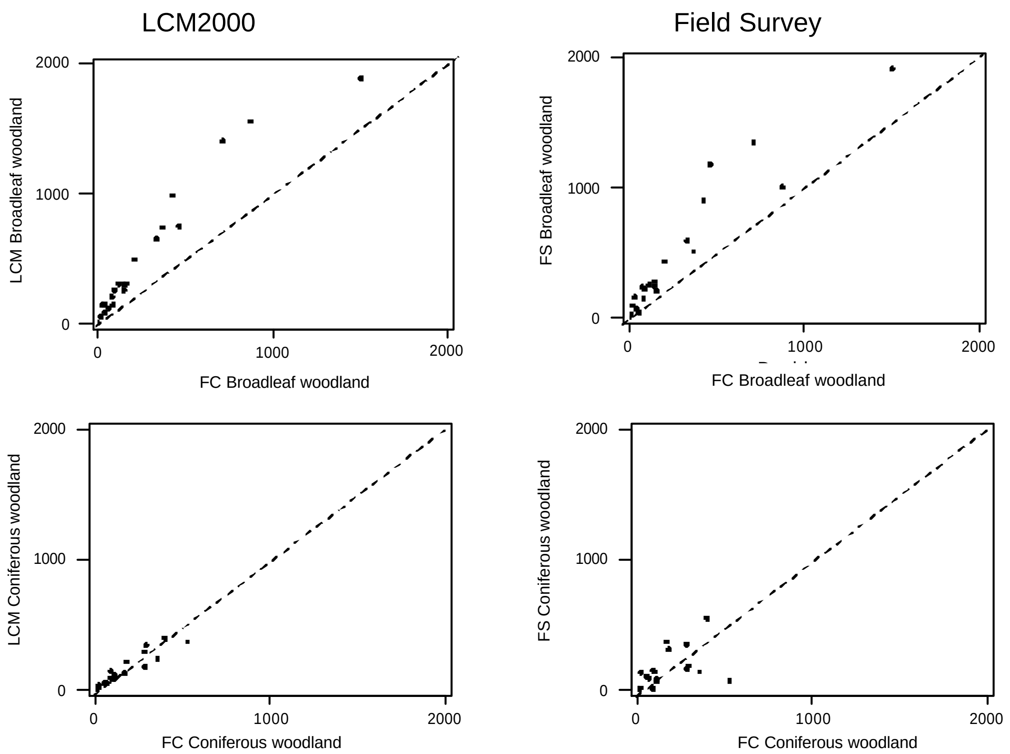
Top row: the cover of dwarf shrub heath as a percentage of BH 10 + BH 12 plotted against the percent cover of acid soils per NLC for LCM2000 and FS. Bottom row: the percentage cover of bog plotted against the percent cover of peat soils per NLC for FS and LCM2000. (The $y = x$ line is also plotted for each graph).

APPENDIX II
SCATTER PLOTS OF FS AND LCM2000 LAND COVER ESTIMATES AGAINST OS DATA



Scatter plots of the LCM2000 and FS estimate of woodland, built up, coastal and inland water land cover against the OS98 total per National Land Class. (The $y = x$ line is also plotted for each graph).

APPENDIX III
SCATTER PLOTS OF FS AND LCM2000 LAND COVER ESTIMATES AGAINST FORESTRY COMMISSION DATA



Scatter plots of the LCM2000 and FS estimate of broadleaved & mixed woodland and coniferous woodland against the Forestry Commission total for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

APPENDIX IV
STOCK ESTIMATES (KM^2) FROM FIELD SURVEY, LCM2000 AND CALIBRATION FOR THE ENGLISH REGIONS

English Regions – North East

	Field Survey			LCM2000	Calibrated LCM2000		
		Bias corrected confidence interval				Bias corrected confidence interval	
Description	Mean	Lower	Upper	Mean	Mean	Lower	Upper
Broadleaved, mixed and yew woodland	464	307	621	459.1	476.7	268.2	757.8
Coniferous woodland	488	73	903	600.8	615.4	356.1	863.7
Arable and horticulture	1748	1215	2281	2191.3	2006.3	1428.5	2536.1
Improved grassland	2492	1947	3037	1868.5	2424.9	1767.9	3117.2
Neutral grassland	235	136	334	680.0	424.9	217.7	640.4
Calcareous grassland	9	-2	20	560.0	55.3	30.2	73.2
Acid grassland	969	588	1350	404.6	676.6	403.9	954.9
Bracken	172	91	253	98.7	150.8	47.8	290.3
Dwarf shrub heath	740	287	1193	665.8	708.8	301.5	1234.7
Fen, marsh and swamp	211	136	286	0.3	194.1	86.7	368.8
Bog	217	39	395	251.7	211.6	28.1	545.9
Standing open water and canals	40	14	66	37.9	16.8	1.9	39.2
Montane habitats	5	-4	14	0.0	0.0	0.0	0.0
Inland rock	8	2	14	63.6	13.1	0.5	26.5
Built up areas and gardens	366	245	487	680.3	652.8	272.4	723.2
Supra littoral rock	21	8	34	0.0	24.2	11.9	38.2
Supra littoral sediment	12	-1	25	3.7	19.0	6.1	30.1
Littoral rock	0	0	0	1.0	0.1	0.0	0.0
Littoral sediment	67	-8	142	9.3	14.4	12.4	15.0

English Regions – North West

	Field Survey			LCM2000	Calibrated LCM2000		
		Bias corrected confidence interval				Bias corrected confidence interval	
Description	Mean	Lower	Upper	Mean	Mean	Lower	Upper
Broadleaved, mixed and yew woodland	803	580	1026	911.4	886.3	505.5	1276.6
Coniferous woodland	442	181	703	307.3	375.3	198.2	581.1
Arable and horticulture	2015	1323	2707	1743.7	1868.7	1050.8	2849.0
Improved grassland	4785	3899	5671	4765.5	5463.5	4182.5	6659.4
Neutral grassland	446	318	574	1348.9	540.8	268.5	762.1
Calcareous grassland	4	0	8	865.2	82.4	41.9	82.0
Acid grassland	1194	844	1544	917.0	1337.5	813.3	1863.2
Bracken	469	173	765	462.3	380.4	159.0	657.9
Dwarf shrub heath	1054	624	1484	448.0	700.2	299.1	1128.3
Fen, marsh and swamp	325	208	442	5.3	278.9	136.9	500.0
Bog	354	135	573	329.2	291.7	69.5	645.5
Standing open water and canals	171	6	336	106.6	100.9	42.4	144.9
Montane habitats	3	0	6	0.0	0.5	0.0	1.2
Inland rock	19	2	36	94.6	28.3	1.9	69.8
Built up areas and gardens	856	491	1221	1666.0	1703.7	640.6	1742.0
Supra littoral rock	9	3	15	0.0	12.3	6.2	20.8
Supra littoral sediment	25	-2	52	19.6	51.4	27.1	67.9
Littoral rock	0	0	0	0.0	0.0	0.0	0.0
Littoral sediment	209	50	368	144.6	294.9	214.8	344.8

English Regions – Yorkshire and The Humber

Description	Field Survey Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper	LCM2000 Mean	Calibrated LCM2000 Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper
Broadleaved, mixed and yew woodland	1073	776	1370	1032.3	936.8	611.7	1261.4
Coniferous woodland	432	161	703	308.0	335.1	171.6	522.4
Arable and horticulture	4272	3440	5104	5768.8	5311.6	4237.7	6383.0
Improved grassland	4435	3743	5127	2761.4	4150.9	3099.3	5330.5
Neutral grassland	481	283	679	1159.8	416.6	191.4	666.2
Calcareous grassland	9	-1	19	957.2	136.6	69.9	186.3
Acid grassland	813	555	1071	370.0	844.8	477.1	1244.8
Bracken	283	158	408	58.3	258.6	96.1	472.8
Dwarf shrub heath	808	454	1162	899.5	857.7	458.2	1216.9
Fen, marsh and swamp	214	138	290	4.4	186.9	93.4	317.0
Bog	171	72	270	328.2	268.4	103.5	532.1
Standing open water and canals	144	40	248	51.2	74.6	23.9	139.5
Montane habitats	2	-1	5	0.0	0.0	0.0	0.0
Inland rock	16	4	28	191.2	16.9	2.9	32.5
Built up areas and gardens	1249	822	1676	1494.1	1676.0	804.9	1812.8
Supra littoral rock	22	9	35	0.0	11.7	5.6	19.7
Supra littoral sediment	18	3	33	0.3	33.9	18.6	42.9
Littoral rock	0	0	0	1.0	0.0	0.0	0.0
Littoral sediment	86	7	165	7.4	18.9	13.3	25.4

English Regions – East Midlands

Description	Field Survey Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper	LCM2000 Mean	Calibrated LCM2000 Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper
Broadleaved, mixed and yew woodland	1114	740	1488	896.0	946.5	575.6	1375.5
Coniferous woodland	232	74	390	181.4	186.7	69.9	309.4
Arable and horticulture	7677	6693	8661	8290.9	8393.1	7329.4	9501.6
Improved grassland	3310	2631	3989	2674.9	3461.2	2603.6	4349.6
Neutral grassland	440	296	584	303.6	352.0	159.6	620.7
Calcareous grassland	0	-1	1	1191.1	94.1	38.0	144.3
Acid grassland	154	82	226	147.6	181.6	72.0	318.1
Bracken	118	57	179	30.7	105.6	48.0	236.0
Dwarf shrub heath	195	89	301	113.9	97.0	23.7	187.5
Fen, marsh and swamp	89	34	144	20.0	51.5	17.9	119.2
Bog	54	-8	116	71.9	90.3	48.5	150.4
Standing open water and canals	200	20	380	77.9	70.4	29.4	141.6
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	11	1	21	86.8	11.9	3.4	19.6
Built up areas and gardens	1269	842	1696	1487.8	1557.7	714.5	1914.0
Supra littoral rock	3	1	5	0.0	1.4	0.7	2.9
Supra littoral sediment	13	-1	27	1.7	17.3	10.2	22.3
Littoral rock				0.0	0.0	0.0	0.0
Littoral sediment	95	-5	195	44.3	70.3	45.2	87.0

English Regions – West Midlands

Description	Field Survey Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper	LCM2000 Mean	Calibrated LCM2000 Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper
Broadleaved, mixed and yew woodland	1138	830	1446	1088.8	1001.5	709.7	1280.1
Coniferous woodland	299	101	497	283.8	285.2	126.9	440.2
Arable and horticulture	3800	3140	4460	4174.0	3939.1	3120.9	4852.1
Improved grassland	4207	3669	4745	4240.2	4978.5	4050.6	5833.4
Neutral grassland	345	203	487	401.1	325.1	149.0	552.1
Calcareous grassland	12	-6	30	805.6	17.1	4.7	17.1
Acid grassland	151	74	228	126.4	209.5	90.8	376.2
Bracken	166	70	262	46.9	123.3	52.8	270.2
Dwarf shrub heath	251	87	415	74.2	98.3	25.1	203.0
Fen, marsh and swamp	114	40	188	1.8	67.3	21.2	161.9
Bog	79	-20	178	1.0	34.5	2.4	101.2
Standing open water and canals	114	27	201	64.0	67.3	16.9	120.6
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	7	-1	15	124.5	21.7	1.6	11.3
Built up areas and gardens	1279	885	1673	1571.1	1845.7	875.0	1795.9
Supra littoral rock	3	0	6	0.0	0.0	0.0	0.0
Supra littoral sediment	3	-1	7	0.0	0.0	0.0	0.0
Littoral rock	0	0	0	0.0	0.0	0.0	0.0
Littoral sediment	20	-14	54	0.0	0.0	0.0	0.0

English Regions – South West

Description	Field Survey Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper	LCM2000 Mean	Calibrated LCM2000 Mean	Bias corrected confidence interval Lower	Bias corrected confidence interval Upper
Broadleaved, mixed and yew woodland	2240	1775	2705	2329.6	1957.7	1403.4	2524.3
Coniferous woodland	752	199	1305	520.1	528.7	246.0	817.7
Arable and horticulture	5574	4628	6520	7293.9	6240.3	4884.2	7783.6
Improved grassland	8768	7827	9709	9149.4	10169.2	8713.5	11634.1
Neutral grassland	739	506	972	302.4	415.3	257.8	629.3
Calcareous grassland	140	12	268	1184.4	296.6	157.1	461.5
Acid grassland	408	137	679	593.5	641.4	349.3	1051.0
Bracken	380	66	694	0.3	320.4	161.1	583.8
Dwarf shrub heath	486	128	844	289.0	361.9	62.4	714.6
Fen, marsh and swamp	309	139	479	20.6	265.9	121.0	546.9
Bog	73	-29	175	67.2	120.1	46.3	230.9
Standing open water and canals	134	42	226	57.7	53.6	31.2	97.1
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	12	4	20	277.3	15.8	1.8	34.9
Built up areas and gardens	2289	1549	3029	1669.0	2384.9	1234.0	3071.2
Supra littoral rock	58	24	92	0.0	81.3	42.1	128.3
Supra littoral sediment	78	7	149	10.1	116.6	57.7	166.3
Littoral rock	0	0	0	2.1	0.1	0.0	0.0
Littoral sediment	199	34	364	45.8	103.8	58.3	142.7

English Regions – East of England

	Field Survey			LCM2000	Calibrated LCM2000		
Description	Mean	Bias corrected confidence interval		Mean	Mean	Bias corrected confidence interval	
		Lower	Upper			Lower	Upper
Broadleaved, mixed and yew woodland	1210	891	1529	1241.3	1131.0	754.6	1495.4
Coniferous woodland	100	22	178	316.3	230.6	123.8	421.3
Arable and horticulture	10830	9520	12140	12239.8	11935.1	10951.3	13007.2
Improved grassland	3380	2442	4318	1824.5	2925.2	2268.8	3707.7
Neutral grassland	630	278	982	443.9	435.7	221.8	728.2
Calcareous grassland	10	-4	24	869.6	111.7	54.3	177.1
Acid grassland	80	-70	230	138.5	95.2	45.1	174.1
Bracken	40	-8	88	5.8	37.3	16.5	67.7
Dwarf shrub heath	30	-18	78	12.0	26.1	14.6	39.5
Fen, marsh and swamp	100	0	200	90.6	92.6	69.8	132.4
Bog	20	-10	50	7.2	28.2	13.1	45.5
Standing open water and canals	290	-159	739	69.3	78.6	56.8	108.2
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	30	-12	72	52.0	39.1	21.5	54.6
Built up areas and gardens	1390	868	1912	1696.1	1945.6	966.8	2217.9
Supra littoral rock	0	-6	6	0.0	3.1	1.5	5.6
Supra littoral sediment	30	-9	69	15.0	23.6	13.4	29.6
Littoral rock	0	0	0	0.0	0.0	0.0	0.0
Littoral sediment	260	-57	577	79.5	260.5	232.6	279.1

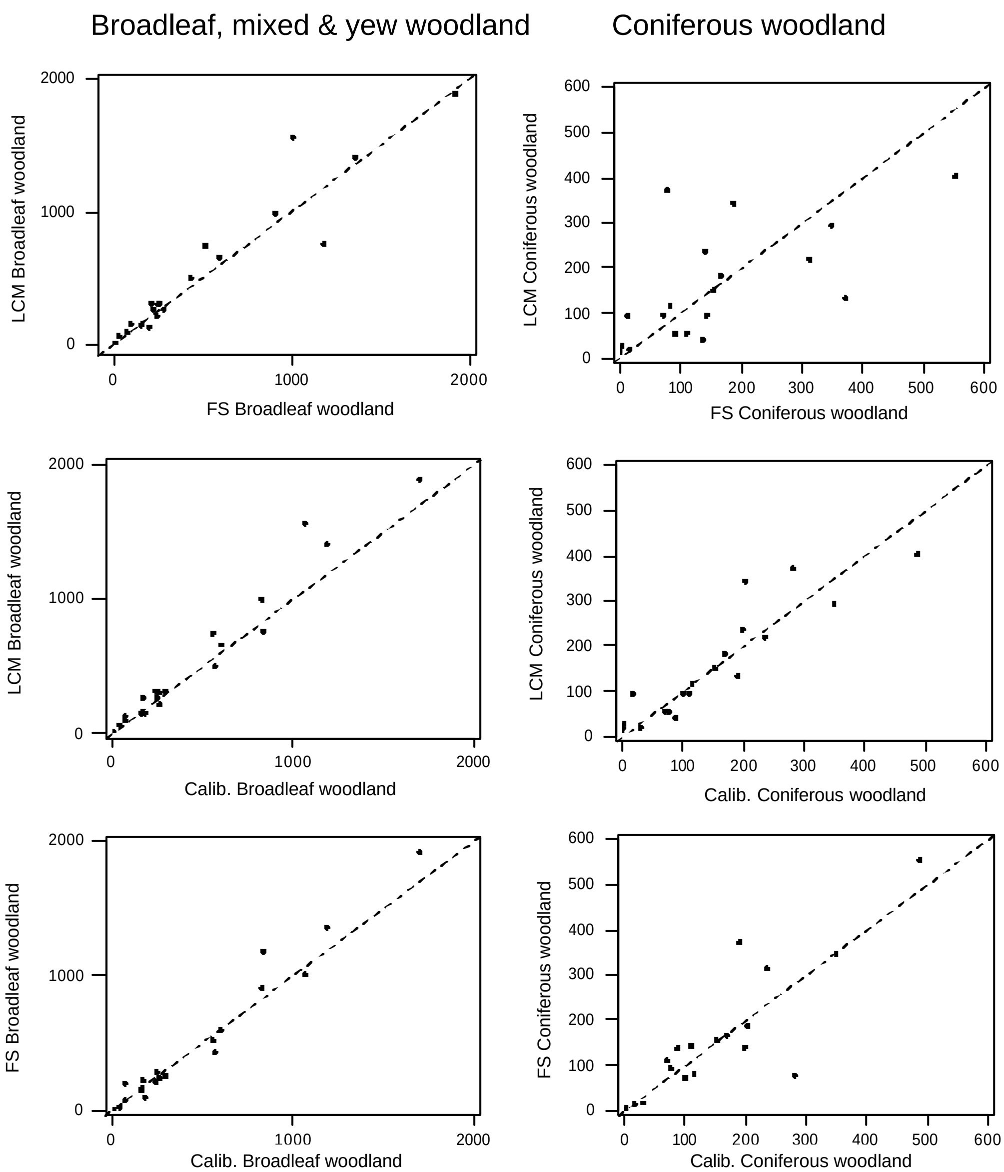
English Regions – London

	Field Survey			LCM2000	Calibrated LCM2000		
Description	Mean	Bias corrected confidence interval		Mean	Mean	Bias corrected confidence interval	
		Lower	Upper			Lower	Upper
Broadleaved, mixed and yew woodland	70	52	89	131.1	120.5	59.4	109.9
Coniferous woodland	12	2	21	5.5	8.4	2.3	15.4
Arable and horticulture	215	179	250	103.0	157.7	108.0	218.3
Improved grassland	244	210	278	138.4	239.3	162.5	239.2
Neutral grassland	21	12	30	76.4	46.0	9.6	32.9
Calcareous grassland	5	0	9	52.6	8.7	2.0	5.9
Acid grassland	3	-1	6	8.0	2.7	0.7	2.3
Bracken	1	0	1	0.0	0.7	0.1	1.4
Dwarf shrub heath	1	-1	4	2.5	1.7	0.1	0.4
Fen, marsh and swamp	9	-1	19	5.4	7.8	3.3	10.2
Bog	1	0	1	0.0	1.1	0.5	2.1
Standing open water and canals	8	-4	20	14.6	12.8	6.1	7.5
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	0	0	1	8.6	4.2	0.2	1.2
Built up areas and gardens	74	52	95	1027.4	988.9	140.3	273.0
Supra littoral rock	1	0	1	0.0	0.2	0.1	0.5
Supra littoral sediment	4	-1	8	0.0	3.6	2.1	4.2
Littoral rock	0	0	0	0.0	0.0	0.0	0.0
Littoral sediment	17	3	31	0.4	2.3	1.3	3.5

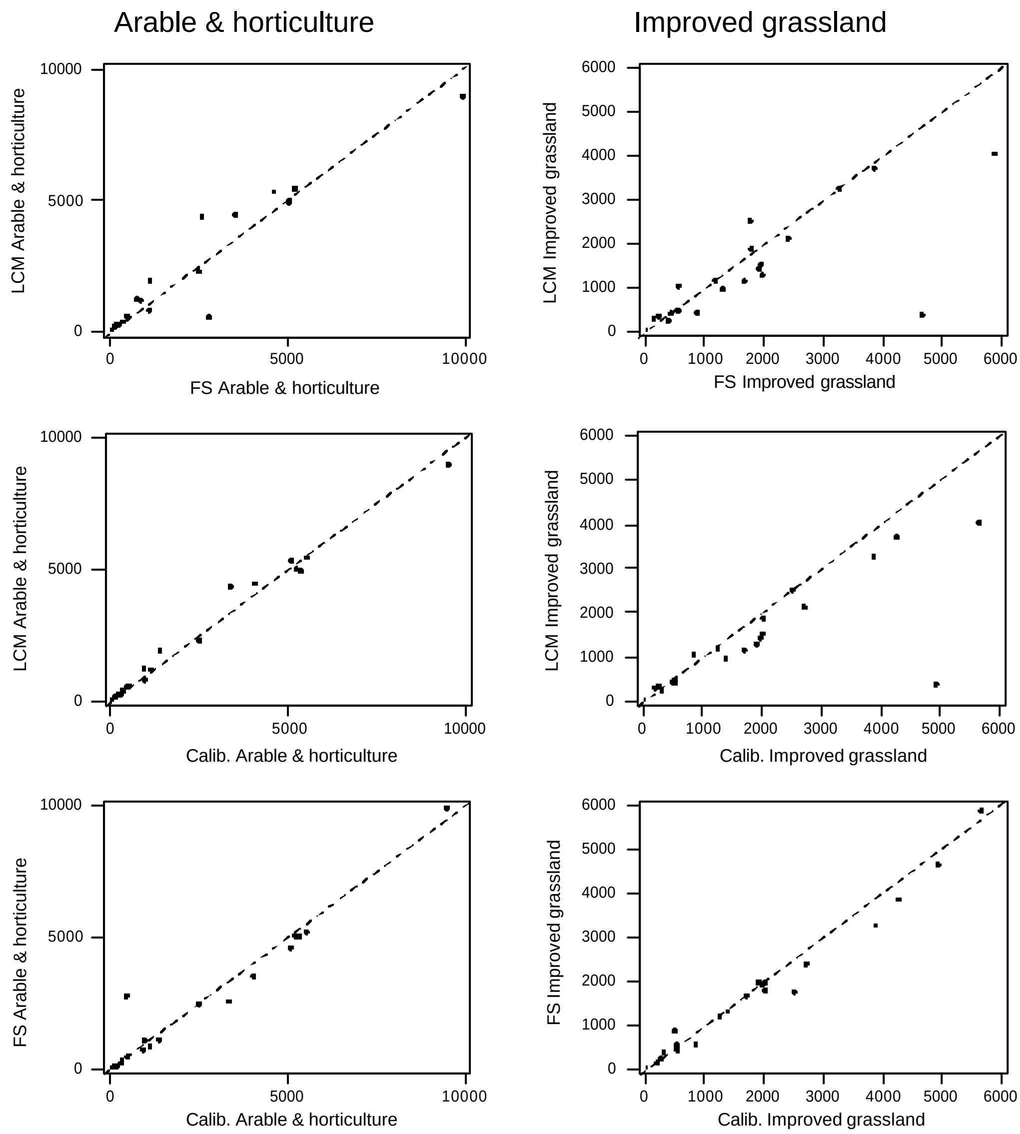
English Regions – South East

	Field Survey			LCM2000	Calibrated LCM2000		
Description	Mean	Bias corrected confidence interval		Mean	Mean	Bias corrected confidence interval	
		Lower	**Upper**			**Lower**	**Upper**
Broadleaved, mixed and yew woodland	1857	1403	2311	2865.9	2077.4	1465.0	2640.7
Coniferous woodland	229	112	346	465.6	321.5	112.3	584.2
Arable and horticulture	7756	6749	8763	6439.6	6969.1	5755.8	8329.9
Improved grassland	5015	4193	5837	4524.0	5667.0	4573.0	6799.0
Neutral grassland	573	355	791	285.4	515.5	235.2	851.3
Calcareous grassland	162	-16	340	1394.6	318.3	156.4	483.6
Acid grassland	49	-20	118	80.4	58.7	24.5	109.0
Bracken	27	7	47	2.3	23.0	6.4	48.9
Dwarf shrub heath	54	-31	139	143.4	36.5	8.5	50.3
Fen, marsh and swamp	117	-7	241	27.7	92.4	47.4	175.6
Bog	11	-3	25	0.0	14.5	3.6	26.9
Standing open water and canals	153	-57	363	88.2	59.0	43.0	86.2
Montane habitats	0	0	0	0.0	0.0	0.0	0.0
Inland rock	18	1	35	212.9	18.6	5.6	35.8
Built up areas and gardens	1648	1172	2124	2474	2909.0	1453.5	3181.9
Supra littoral rock	17	6	28	0.2	24.8	11.7	44.6
Supra littoral sediment	51	-6	108	29.9	115.2	63.2	144.1
Littoral rock				0.3	0.0	0.0	0.0
Littoral sediment	230	30	430	33.1	126.4	75.5	166.5

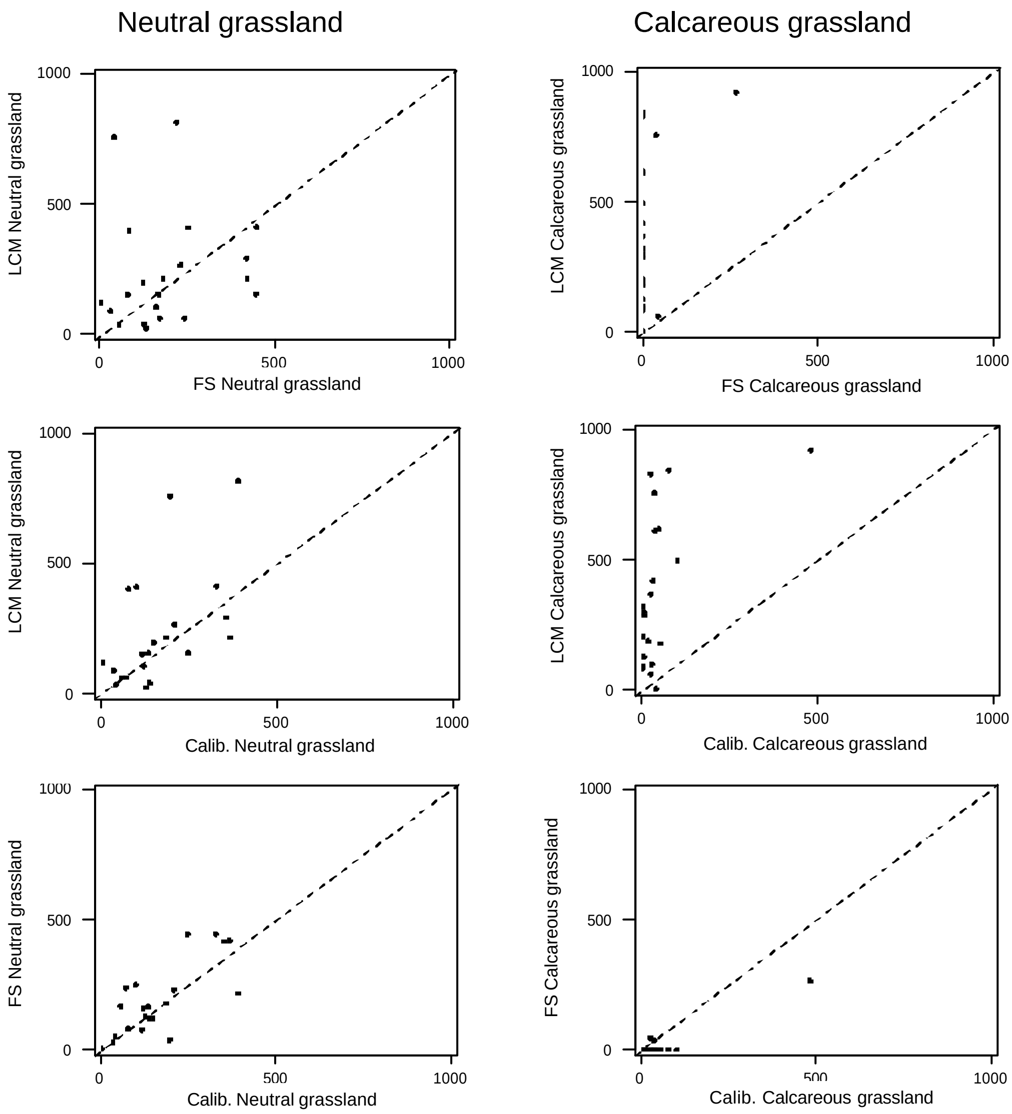
APPENDIX V
SCATTER PLOTS OF FIELD SURVEY, LCM2000 AND CALIBRATED LAND COVER ESTIMATES



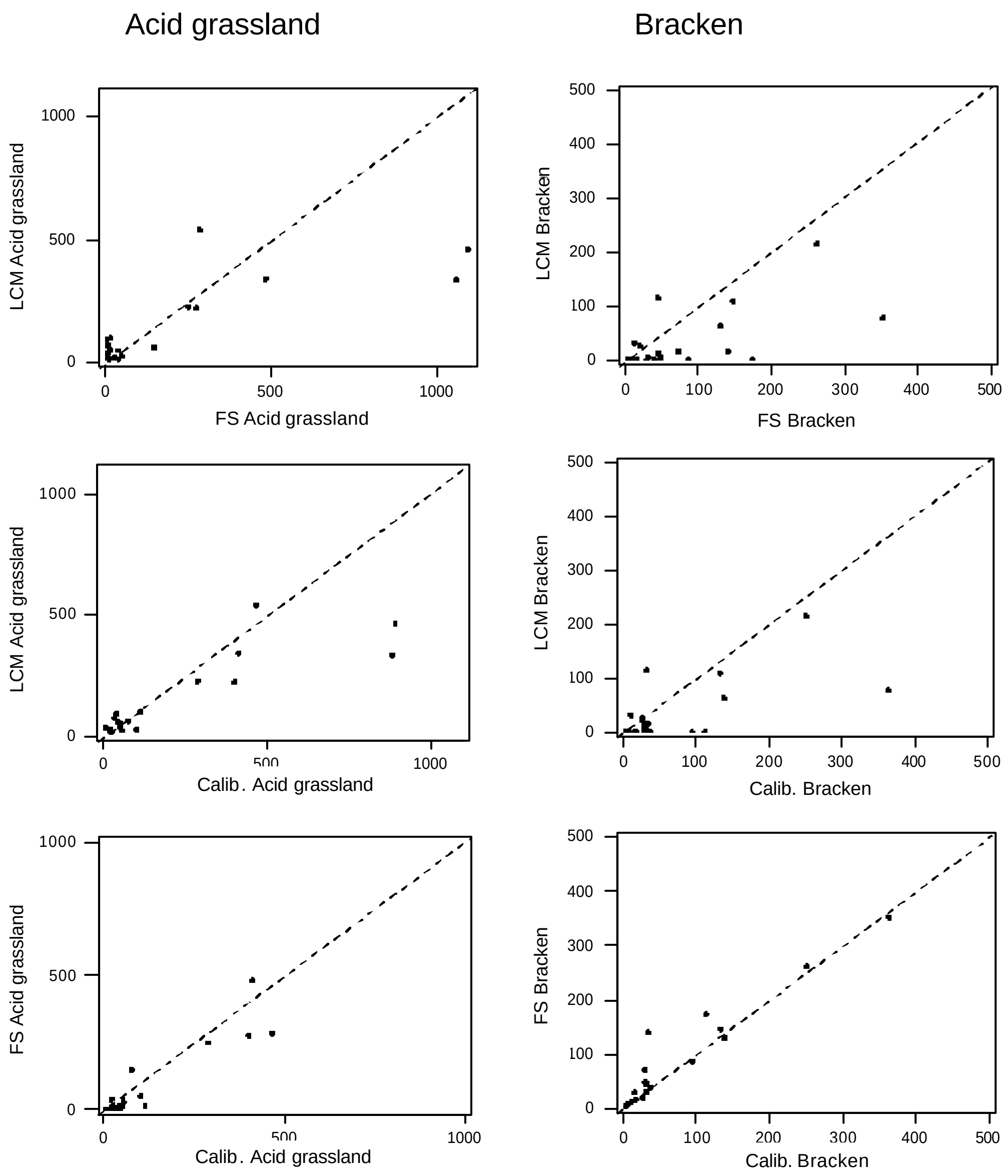
Scatter plots of FS, LCM2000 and calibrated estimates of broadleaved & mixed woodland and coniferous woodland for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



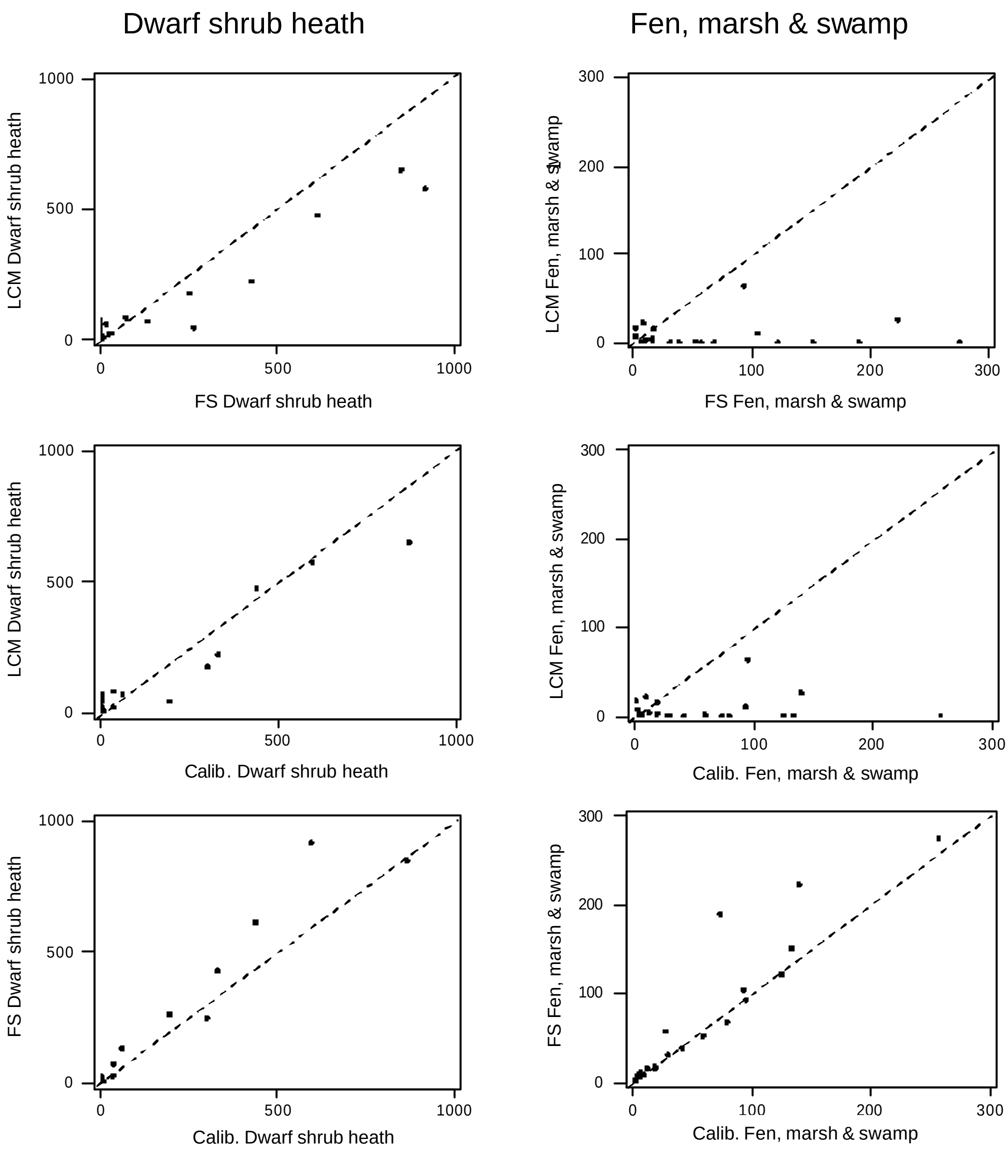
Scatter plots of FS, LCM2000 and calibrated estimates of arable & horticulture and improved grassland for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



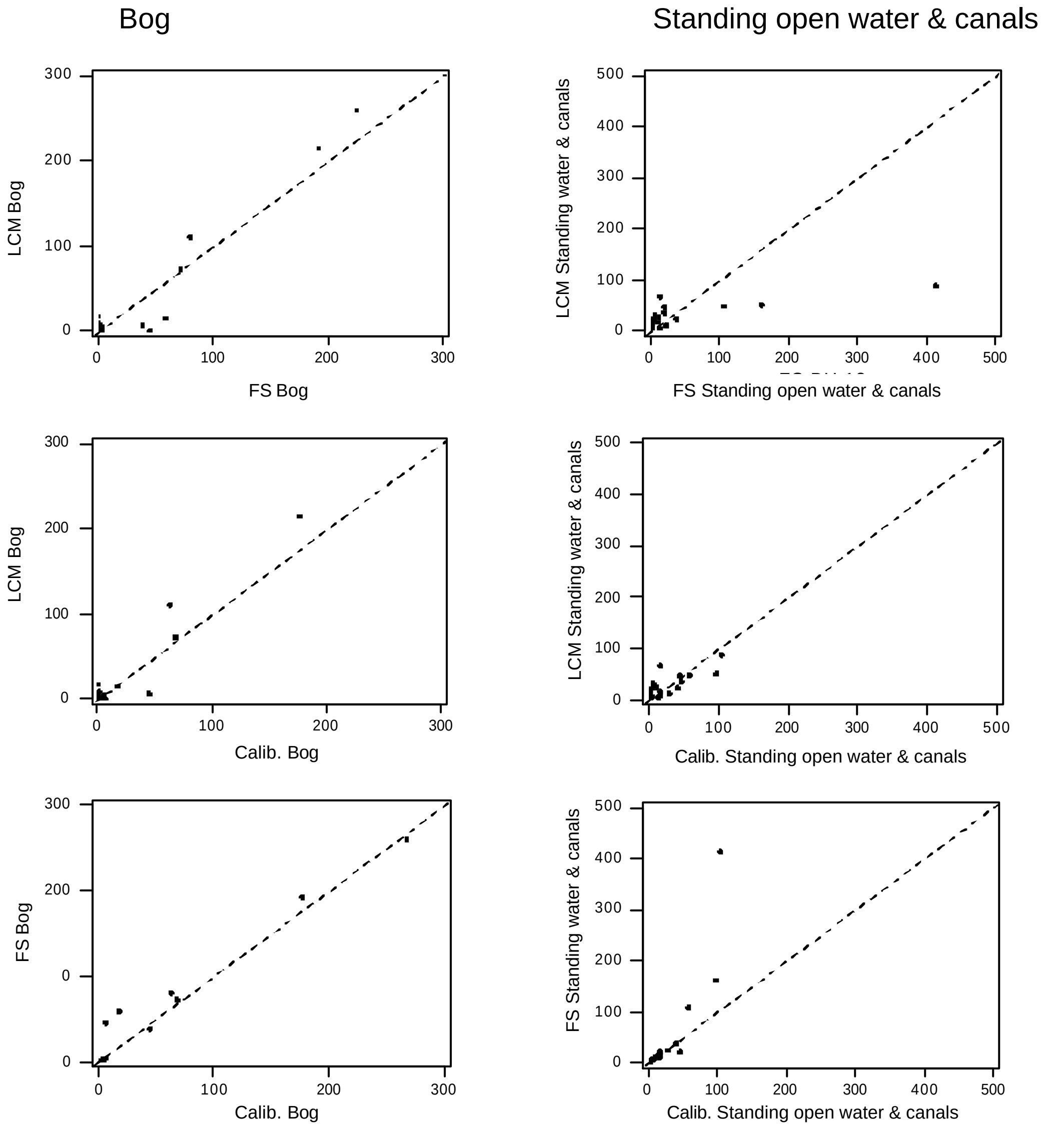
Scatter plots of FS, LCM2000 and calibrated estimates of calcareous and neutral grassland for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



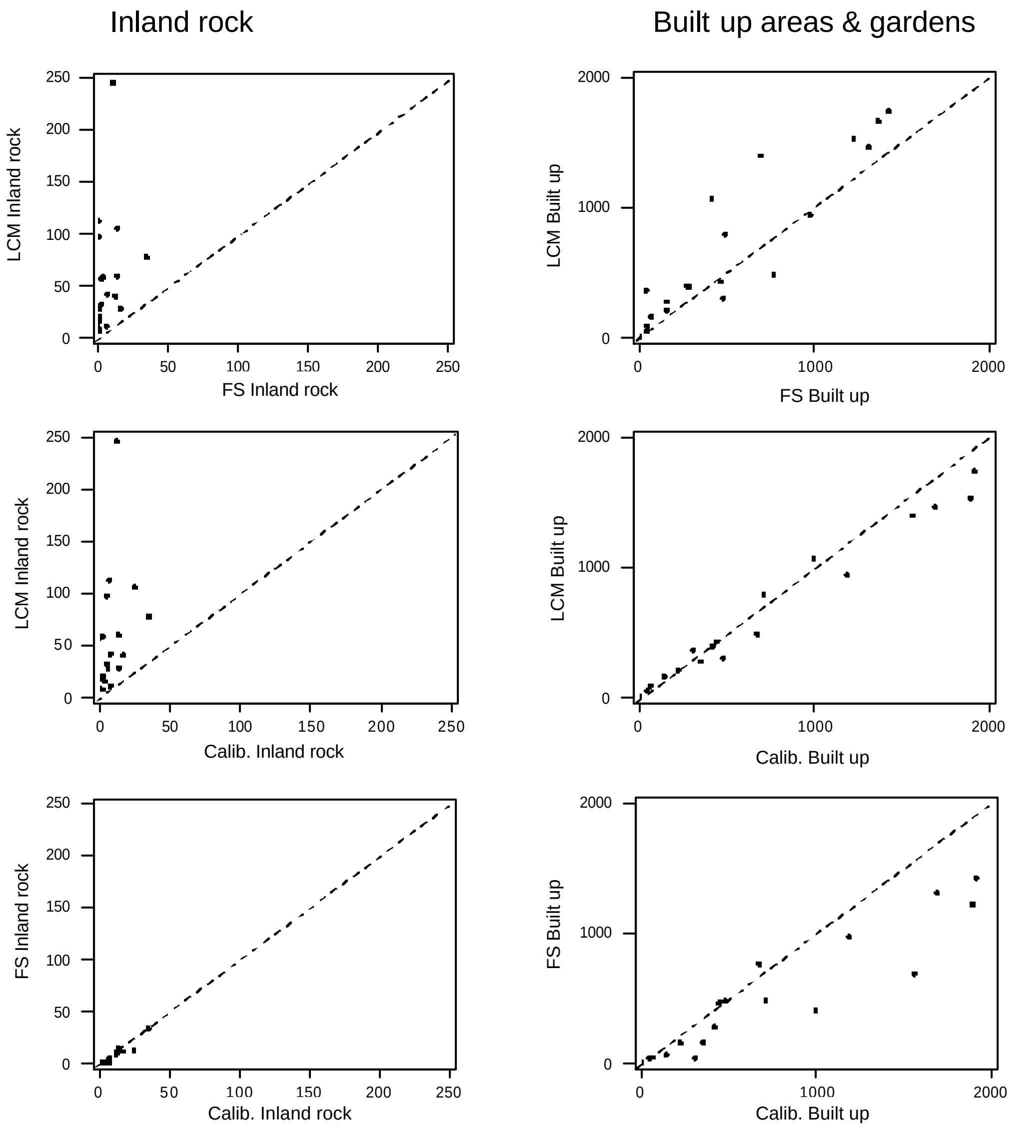
Scatter plots of FS, LCM2000 and calibrated estimates of acid grassland and bracken for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



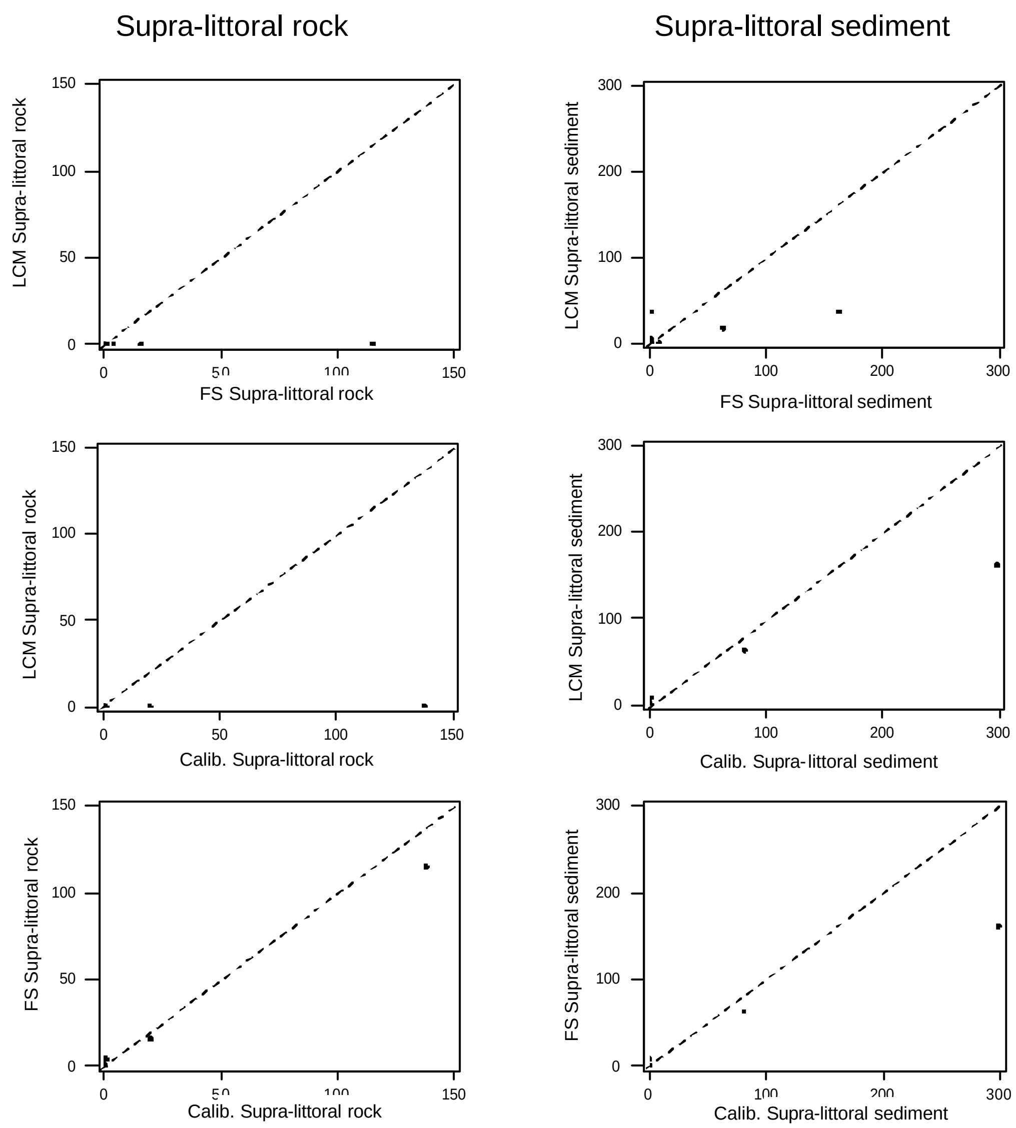
Scatter plots of FS, LCM2000 and calibrated estimates of dwarf shrub heath and fen, marsh & swamp for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



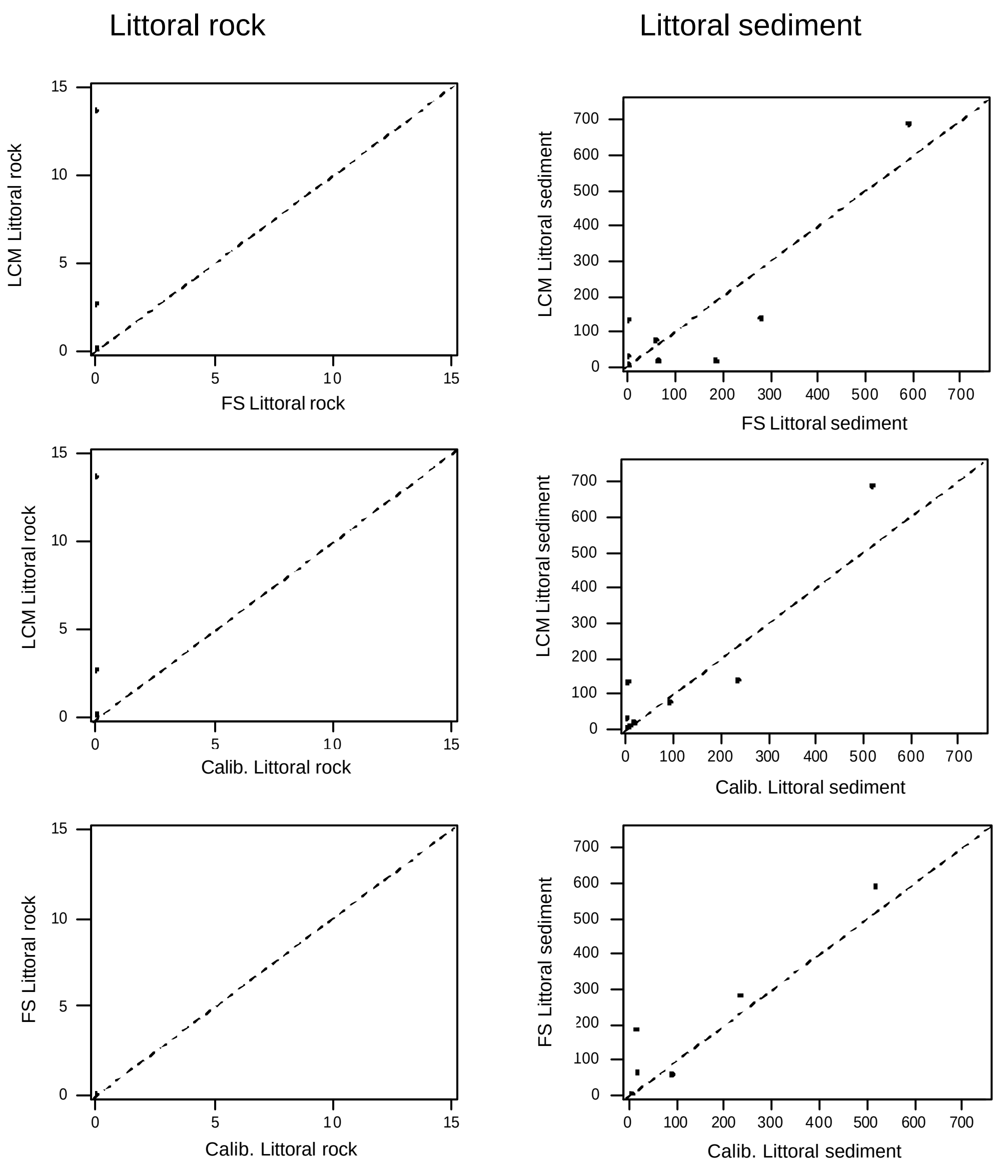
Scatter plots of FS, LCM2000 and calibrated estimates of bog and standing open water & canals for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



Scatter plots of FS, LCM2000 and calibrated estimates of inland rock and built up & gardens for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

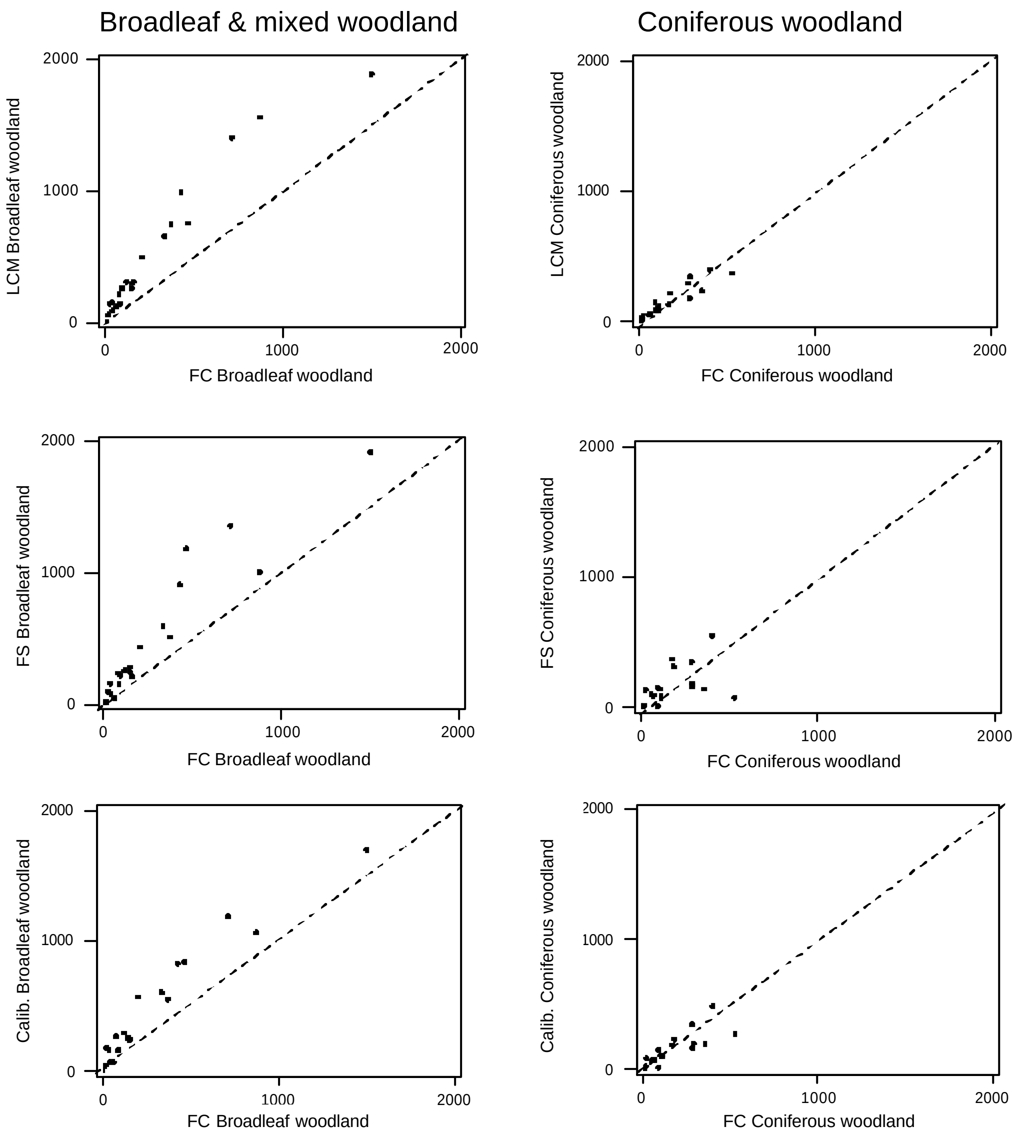


Scatter plots of FS, LCM2000 and calibrated estimates of supra-littoral rock and supra-littoral sediment for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



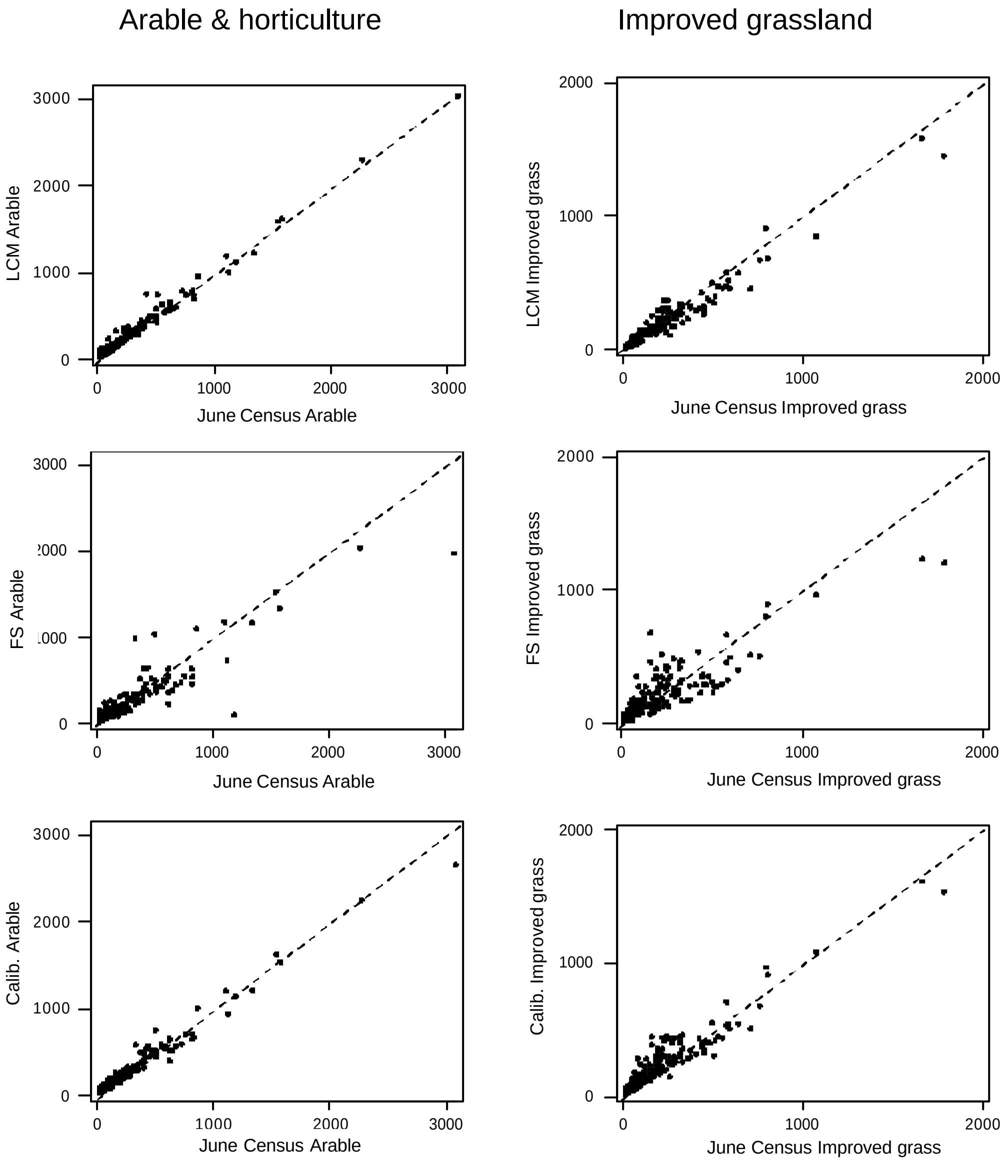
Scatter plots of FS, LCM2000 and calibrated estimates of littoral rock and littoral sediment for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

APPENDIX VI
SCATTER PLOTS OF LCM2000, FS AND CALIBRATED LAND COVER ESTIMATES AGAINST FORESTRY COMMISSION DATA



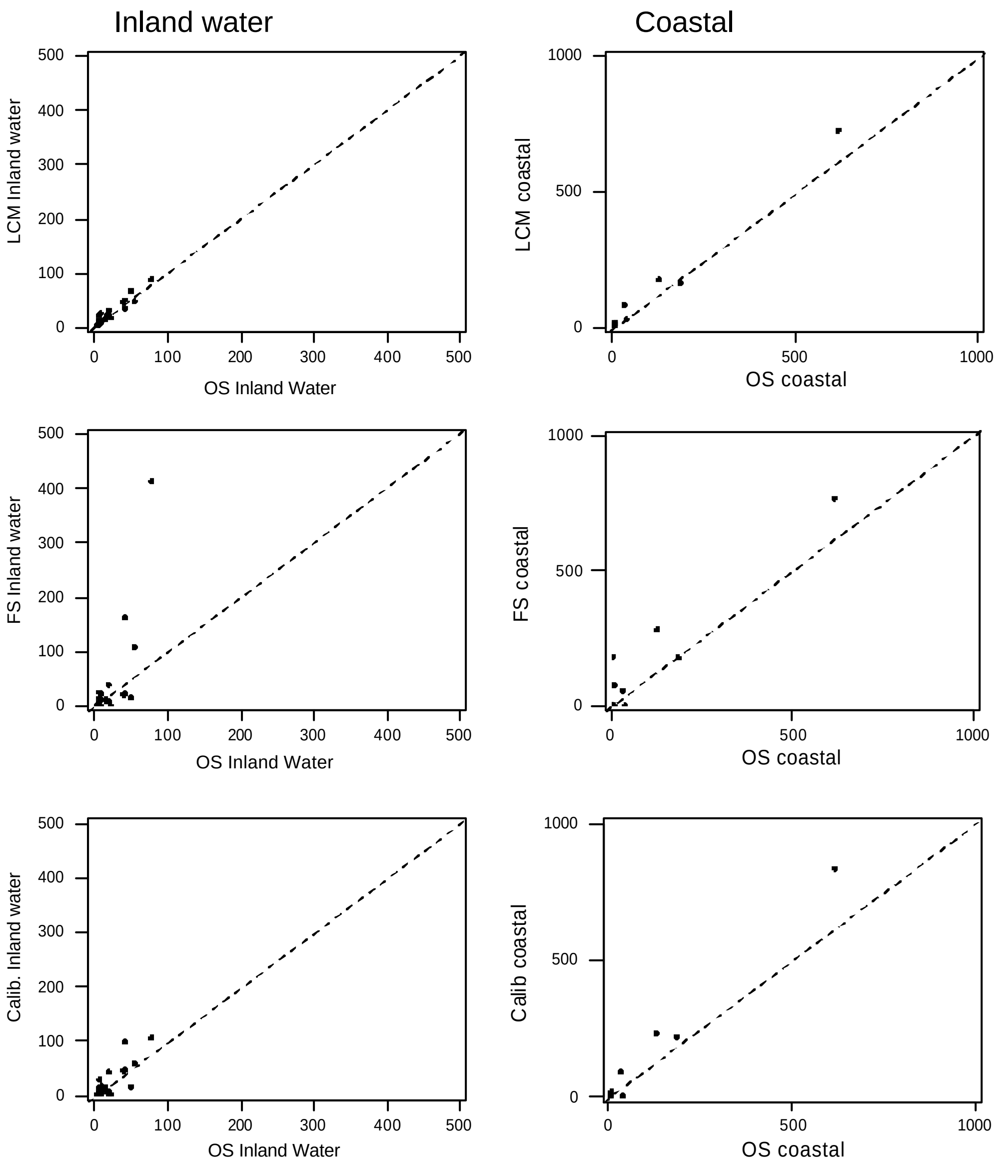
Scatter plots of LCM2000, FS and calibrated estimate of broadleaved & mixed woodland and coniferous woodland against the Forestry Commission total for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

APPENDIX VII
SCATTER PLOTS OF LCM2000, FS AND CALIBRATED LAND COVER ESTIMATES AGAINST JUNE CENSUS 1998 DATA

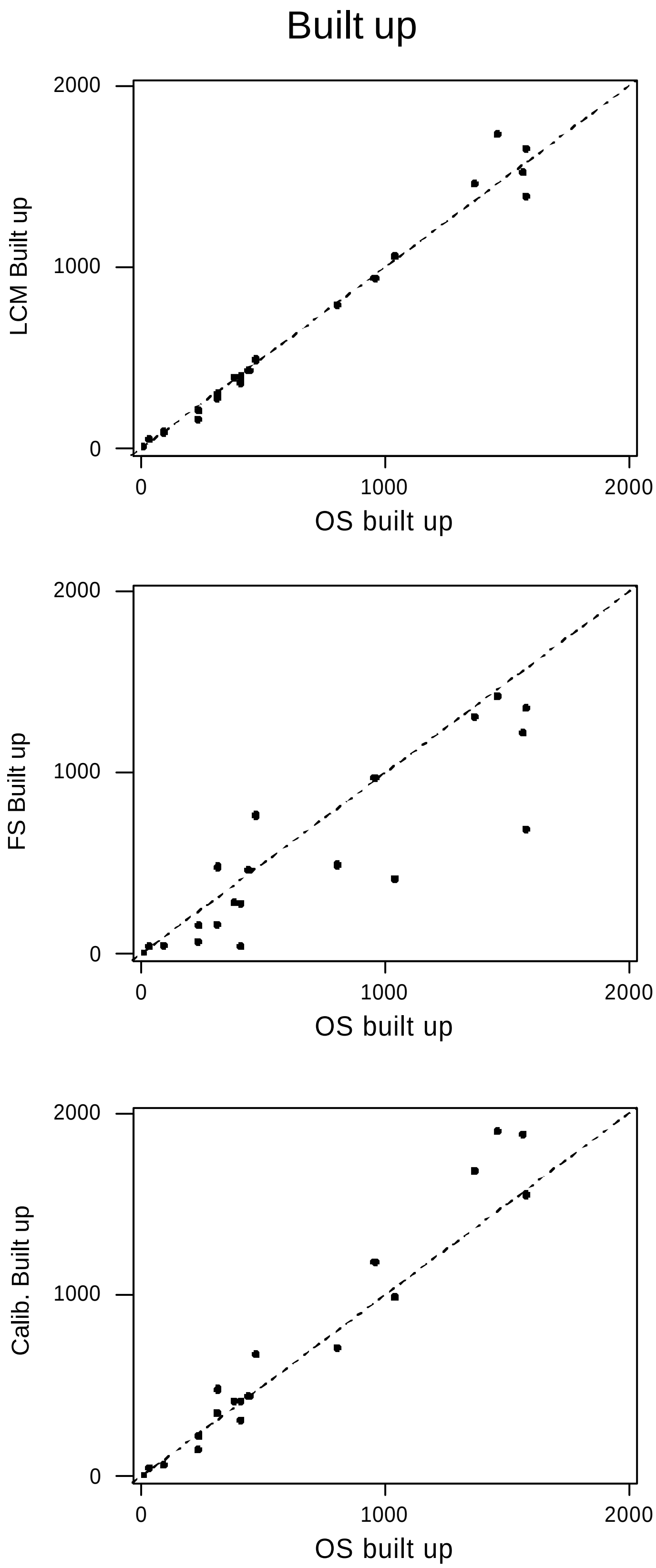


Scatter plots of LCM2000, FS and calibrated estimate of arable & horticulture and improved grassland against the June Census 1998 estimate for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

APPENDIX VIII
SCATTER PLOTS OF LCM2000, FS AND CALIBRATED LAND COVER ESTIMATES AGAINST OS DATA



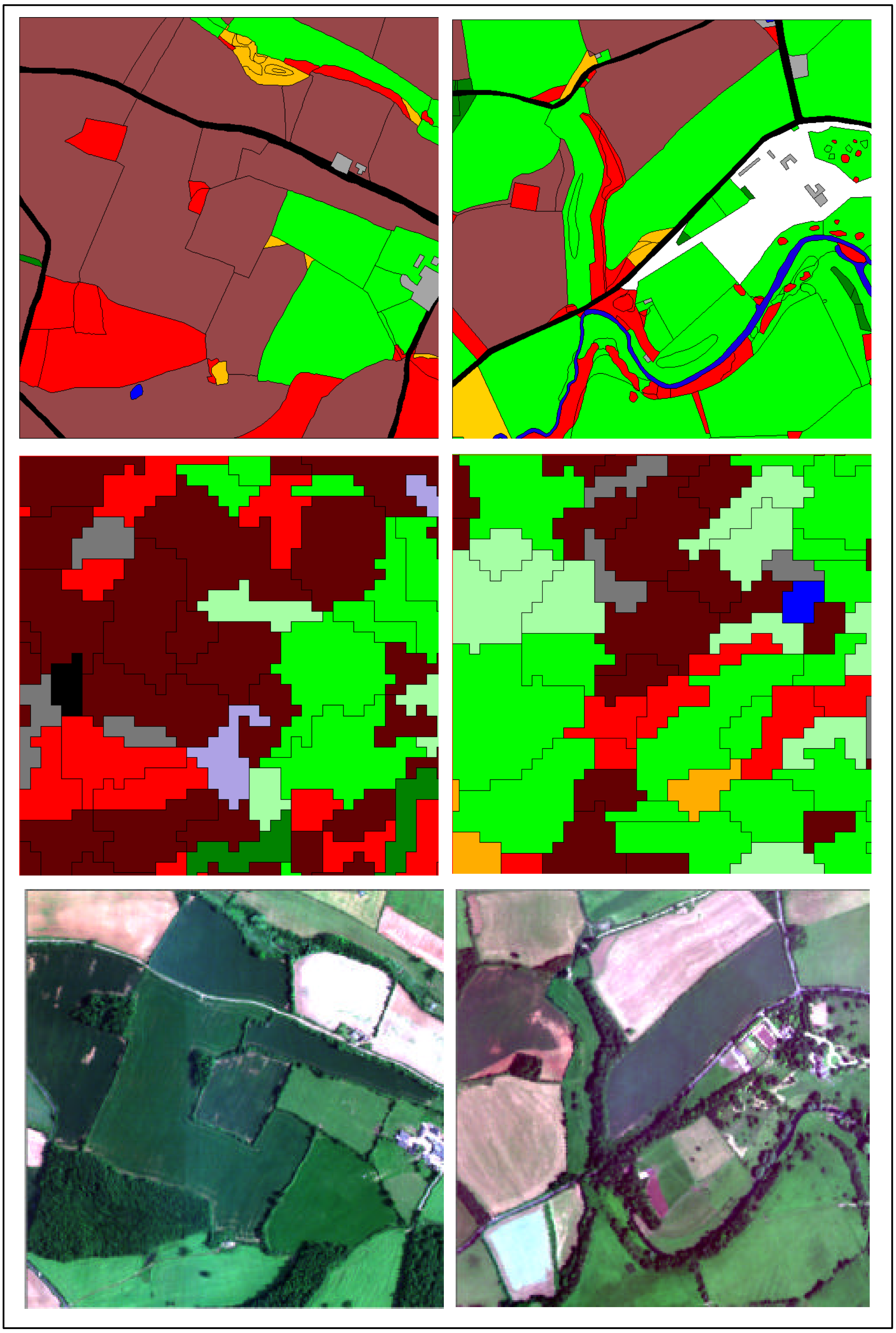
Scatter plots of LCM2000, FS and calibrated estimate of inland water and ‘coastal’ against the OS Geographic Reference data total for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).



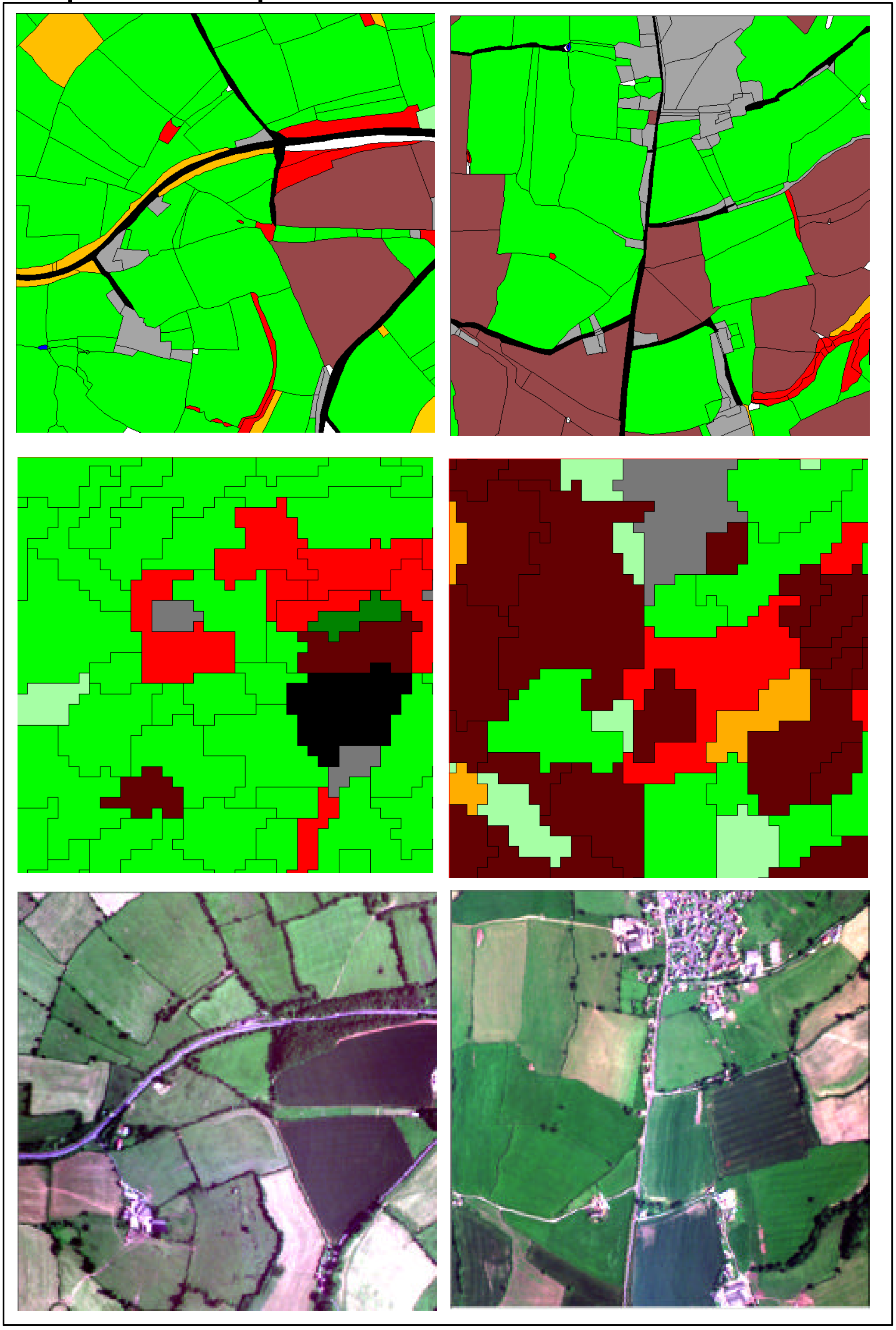
Scatter plots of LCM2000, FS and calibrated estimate of built up & gardens against the OS Geographic Reference data total for the 21 National Land Classes of England. (The $y = x$ line is also plotted for each graph).

APPENDIX IX
FS AND LCM2000 DATA FOR PAIRS OF FIELD SURVEY SQUARES

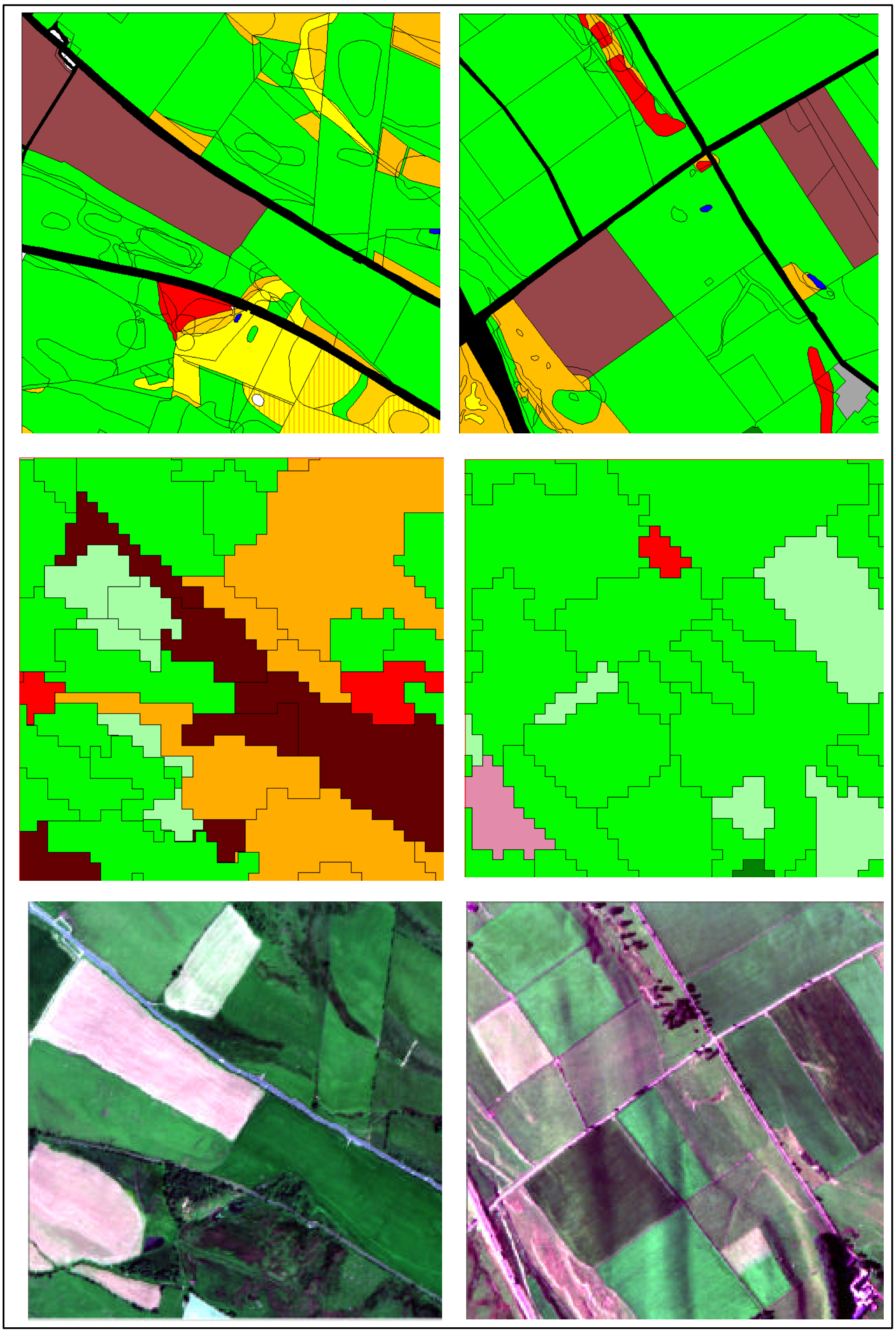
Sample Arable squares



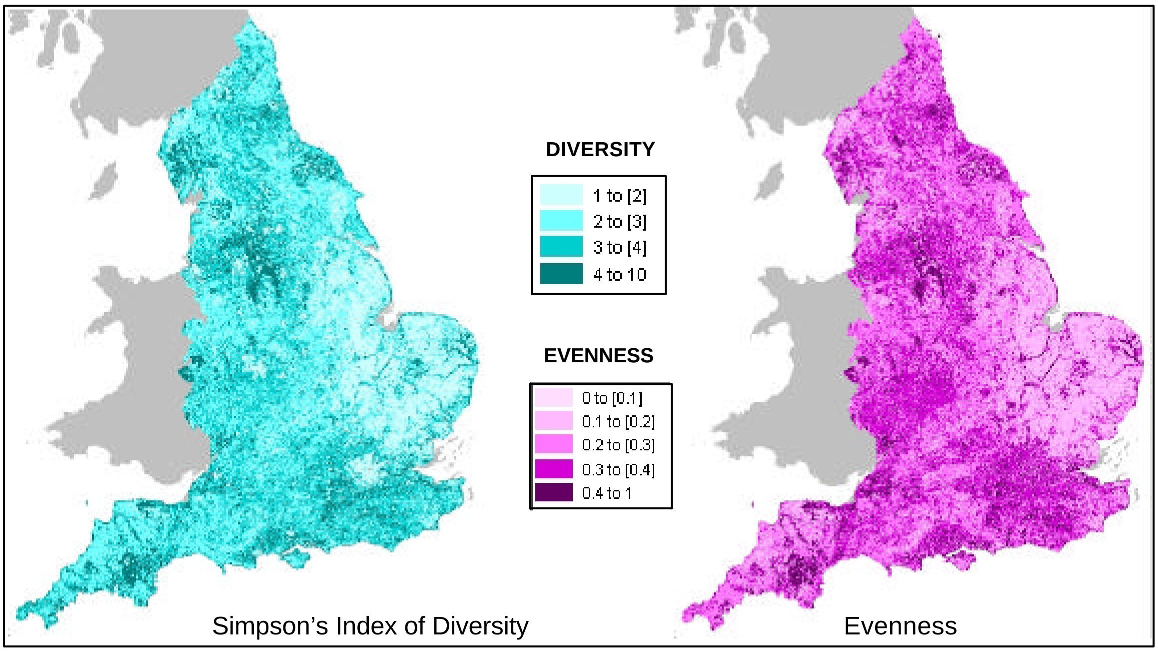
Sample Pastural squares



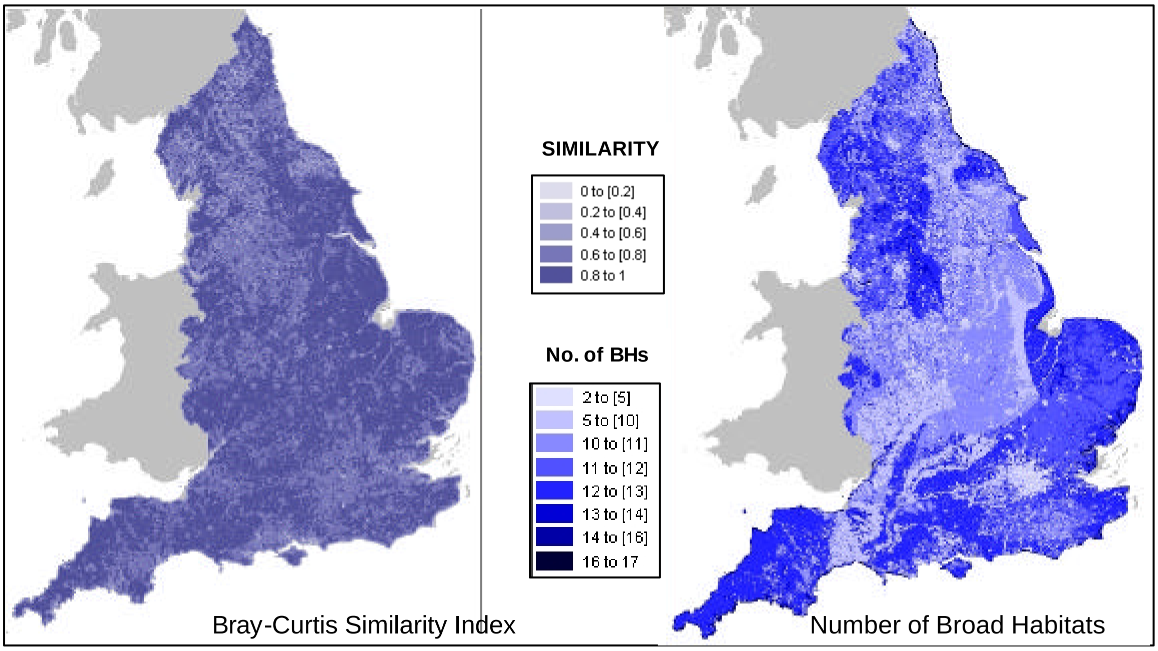
Sample Marginal squares



APPENDIX X
PATTERN METRICS PER 1 KM^2 FOR ENGLAND



Simpson's Index of Diversity
Evenness



Bray-Curtis Similarity Index
Number of Broad Habitats