Quality Assurance Report: MAPPING QUALITY ASSURANCE EXERCISE

LISA NORTON, PAUL SCHOLEFIELD, LINDSAY MASKELL, SIMON SMART

CEH Lancaster

JOHN MURPHY, PENNY WILLIAMS[1], PETER SCARLETT & RALPH CLARKE[2]

Centre for Ecology and Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP
UK

[1] Pond Conservation, Oxford Brookes University, Gipsy Lane, Headington, Oxford, OX4 2PD.
[2] School of Conservation Sciences, Bournemouth University, Christchurch House, Talbot Campus, Poole, Dorset, BH12 5BB, UK

Countryside Survey 2007

Mapping Quality Assurance Exercise

Lisa Norton, Paul Scholefield, Lindsay Maskell, Simon Smart – CEH Lancaster

1. Introduction

A key output of Countryside Survey in 2007 was the provision of high-quality data on habitats and landscape features which was compatible with previous approaches. The data also needed to be in a form which enabled speedy analysis and in a database which allowed investigation of CS data alongside other national scale spatial datasets (a geodatabase). As a result, digital mapping was undertaken in the field for the first time in 2007. One of the reasons for deciding on this approach was to eliminate a major potential source of error in previous surveys. Interpretation and digitisation of overlaid paper maps post-survey by lab based digitisers, as well as being a very lengthy process, has also proved to be a complex and difficult task. The adoption of a digital approach avoids subjectivity in interpretation as well as ensuring clarity of the mapping itself and of the associated attribute data.

The use of a specially designed software package (Surveyor) in 2007 made it possible for the surveyors (following a relatively lengthy period of appropriate training) to directly map in the field and required them to provide the necessary information for each of the features present on the map. The new approach had significant and positive implications for the quality of the data coming back from the field. The Surveyor software contains mandatory fields, provides prompts for the data expected for each of the mapped feature types, avoids issues of poor handwriting /spelling and also provides 'visit status' layers, enabling surveyors to identify features not yet visited. Additionally, the data was uploaded to a 'wiki' (inter-active website) soon after completion which enabled data checking (by staff in the office) early on in the survey. The wiki was used by both surveyors and staffs involved with data receipt, survey design and analysis, to communicate any issues arising and provide guidance on approaches to take. Quality Control teams visited all survey teams in the early part of the survey to ensure that protocols were being followed appropriately.

In terms of quality assurance (QA), digital mapping ensures that an assessment of the surveyor mapped data (which is immediately the final data – without the interim step of being digitised) against that of the CS team can be carried out digitally using a range of approaches. For this report we have adopted three of these approaches including one which uses a similar technique to that used in previous quality assessments of the mapped data. In previous years the same squares were used for both plot and mapping QA. In 2007, because of the new mapping methodology used, it was decided that the CS team who had developed the methodology would be best placed to carry out the QA on that part of the data collection. As a result of this, and to avoid excess pressure on particular land-owners, the majority of the mapping QA was carried out in different squares to those used for the plot and freshwater QA exercises. Additionally, rather than just a quarter of a CS square being mapped (as for previous QA) as far as possible (where access was not refused) in 2007 whole squares were mapped in the QA exercise and mapping included area, linear and point features. This resulted in a far more in-depth QA exercise on the mapped data.

Efficacy of mapping can be tested at a range of scales and for a range of factors. At the coarsest scale frequency of different Broad Habitat codes provides a population estimate per square which indicates any overall discrepancy in the areas of mapped Broad Habitats. The accuracy of mapping at a specific point can be assessed by comparing attributes recorded for polygons in the same spatial location, or similarly those recorded for point or linear features. This assessment can be carried out on all attributes recorded from those at the highest level (primary attributes – such as Broad Habitat or feature type) where concordance is expected, to secondary attributes such as species which are likely to be somewhat more subjective and for which exact matching between QA assessors and surveyors is less critical. In 2007 surveyors were also asked to provide an assessment of the accuracy of the data which they took into the field with them. This resulted in data for each area, line and point which indicated whether the 1998 data was reliable in terms of BH and primary attributes (for areas) and primary attributes alone for linear and point features. The QA process can also be used to assess the decisions made in the fields by surveyors in relation to the 1998 data.

2. Extent

The QA on the mapped data covered a total of 23 1km squares. This fell short of the target number of squares but far exceeds previous mapping QA exercises and is considered an adequate representation of square types to assess overall quality of mapped data collected and to highlight issues which may have resulted in a bias in the data. As with the survey itself, refused access issues resulted in some areas of squares not being surveyed. Analyses were therefore performed on the common areas between the QA and the survey.

As described above, the mapping QA was carried out predominantly on different squares to those used for other QA exercises. Squares were selected on two criteria:

- Land Class
- Location

The squares surveyed each represented a different land class and, in order to ensure that all teams of surveyors were being adequately assessed, covered areas across the whole of GB.

3. Approach

Where possible QA teams visited squares at the same time as/or close to the times when survey teams were present. This both minimised any differences resulting from temporal changes to vegetation and ensured minimal disturbance to land-owners/managers. Unlike surveyor teams QA teams were able to map on more than one tablet and therefore divided squares by area or by ‘mapping task’ with one QA assessor concentrating on habitats and the other linear and point features. Whilst this was a more efficient use of time in the field it has resulted in difficulties in ‘checking-in’ data which and has made the analysis of QA data considerably more lengthy than it would otherwise have been. Future QA exercises will need to take this into account in their design.

Squares that were included as part of the QA were surveyed using the standard Countryside Survey tablets and software. Data was collected from each tablet, and combined into a single geodatabase for analysis. Prior to analysis, unsurveyed areas were excluded, as were areas that were refused access, this included data for linear features, areas and points. This was repeated for the survey data collected by the CS surveyors, and the two datasets aligned, to provide two comparable datasets, i.e. common surveyed areas, using a mask overlay of the two datasets.

4. Analysis

The use of digital data collection and the ability to enter data into a geodatabase has considerably enhanced the capacity for investigating comparisons between data collected by surveyors and that collected by quality assessors. A range of methods have been used to compare the data, as outlined below.

4.1 Direct comparison of aggregate areas/lengths/point for whole squares.

For the aggregate comparitors, data from each square on the extents and attributes of habitat, linear and point feature types (Fig 1) was exported to SAS for analysis. This enabled deeper analysis of the data summarised by Square, Land Class and Broad Habitat.

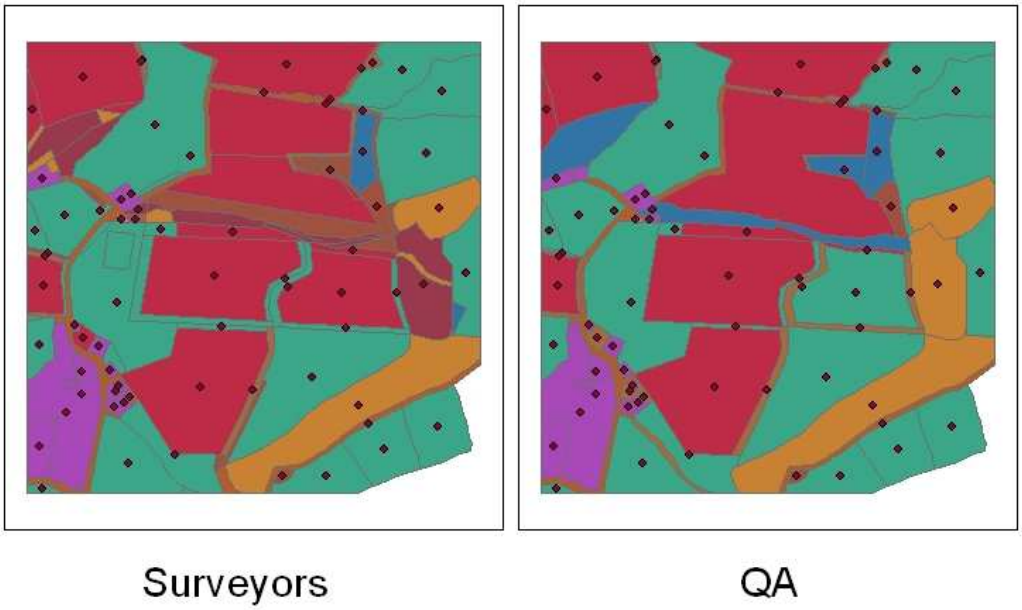

Figure 1. Illustration of square data mapped by the Surveyors and the QA team. Different colours represent different Broad Habitat types. Points indicate polygon centroids.

4.2 Comparisons of raster data

The second analysis was to convert polygon data to a raster format. In this process each 1km square was divided into 10,000 sections each measuring 10 x 10m (half of the minimum mappable unit). Each 10 x 10m unit was then assigned to a Broad Habitat on the basis of the dominant Broad Habitat in the polygon data. This process made it possible to compare both the absolute amounts of a Broad Habitat (between surveyors and the QA exercise) as well as the spatial locations of the Broad Habitats. A resolution point grid was used to sample the raster datasets (QA and surveyor respectively) to enable analysis. A range of other attributes could be sampled from the area datasets using this point feature class.

4.3 100 point sample grid overlaid to investigate commonality.

A further analysis was carried out using a 100 point sample. This involved overlaying a 10 x 10 grid of points on each dataset, performing a spatial join on the underlying attribute tables of each datasets and comparing them (Fig 2). This is of comparable resolution to the previous QA (Prosser and Wallace, 1999).

Figure 2. Illustration of a square with 100 sample points overlaying Broad Habitat data. N.B. This square is composed of few different BH's.

These different methodologies are explored (as appropriate) for each of the feature types: 1) Areas, 2) Linear Features and 3) Point features. Where a different approach has been used, details are provided in the relevant section.

4.4 Analysis of attributes associated with areas, linears and points

For each area and feature type a reference ID layer was generated for comparison of QA with surveyor datasets. This meant that a comparison could be made on the species attributes of matched areas, linear and point features.

5. Surveyor efficiency

The digital system, combined with checks on the data as it was uploaded to the wiki, ensured that surveyors made effective use of the 'visit status' layers for each of the mapped feature types (areas, lines and points), although some survey teams initially used these layers more effectively than others. 'Refused access' areas and uncertainties about land ownership may have resulted in some areas being missed. The proportion of land not visited (as opposed to 'Refused access' was very small, on average 0.4% and 0% for the majority of the QA squares (Table 5.1). NB many

squares had areas of 'refused access' land, hence visited area even in squares with 0% 'not visited' does not always equal 1km.
The inclusion of mandatory recording fields in the data ensured that attributes were recorded against each feature type far more effectively than in previous surveys. Surveyors were asked to record dominant species for each area, as previously. The digital system resulted in surveyors being more likely to record a larger number of species (Table 5.2).

Table 5.1. Surveyor efficiency (area visited of available survey area) in each of the QA squares.

Square	Not Visited (m^2)	Visited (m^2)	% Not Visited
47	354.6	569714.4	0.1
63		914572.9	0.0
261		993989.4	0.0
355		1000000	0.0
359	64.0	917099	0.0
364	26088.5	973911.8	2.7
366	11733.7	988266.5	1.2
383		979131.7	0.0
434	13666.8	527779.7	2.6
488		955467.2	0.0
657	130.9	432657.2	0.0
684	2204.4	908904.1	0.2
694		979333.3	0.0
695	2703.5	992874.4	0.3
765		921320.6	0.0
773	9496.3	990502.9	1.0
920	1069.5	993040.7	0.1
935	41.6	990289	0.0
991		998306.3	0.0
1034		999394.8	0.0
1039		996344.7	0.0
1113		993825.6	0.0
1260		1000000	0.0

Table 5.2. Average number of plant species recorded per area in the last 4 surveys.

Survey year	No. of plant species per area
1984	1.80
1990	1.35
1998	0.65
2007	2.03

6. Comparison of QA and CS data - Areas

6.1 Direct comparison of aggregate areas for whole squares.

The first breakdown of the data shows the proportion of each square occupied by each of the Broad Habitats as collected by the surveyors (CS) and the QA team (Table 6.1a). This table indicates that in 81% of cases the presence of a BH in a survey square was recorded by both CS and QA teams. Further analysis of this data (Table

6.1b) indicates that the mean differences between the proportions of each Broad Habitat recorded in the QA and the CS data are less than 1% for thirteen of the Broad Habitats, rising to between 1 and 2% for a further five Broad Habitats and less than 3.5% for the remaining two Broad Habitats. These differences are relative to the extent of the Broad Habitats in the sample (see mean values for BH's in Table 6.1a), so for example, where the mean difference is 3% in a Broad Habitat present at only around 6% the issue is a 50% difference between the CS and QA surveyors for this Broad Habitat.

The standard deviations around these means are in most cases relatively small, under 5% for the majority (14) of the Broad Habitats. However, in several cases the standard deviations are over 5% and this indicates potential problems with differential allocation of these habitats between the QA team and the CS surveyors. This analysis draws attention to potential issues with Improved and Acid Grassland, Dwarf Shrub Heath, Bog, Urban and Sub-littoral sediment Broad Habitats. A negative difference implies that the CS surveyors are coding more areas to a particular BH than the QA team and a positive difference implies the reverse. The balance between the positive and negative signs for Dwarf Shrub Heath and Bog may indicate that the QA team were more likely to allocate an area to Bog, whereas the CS team would allocate the same are to Dwarf Shrub Heath.

These potential issues are explored in more detail in the following types of analyses. However, these preliminary results show that despite some differences in the coding of the Broad Habitats, the differences between the QA team and the CS team for many habitats are relatively minor.

Table 6.1a. Comparison of the summary areas of the different Broad Habitats in each square. CS = data collected by surveyors, QA = data collected by the QA team.

BH	No allocation		Broadleaved & yew woodland		Coniferous woodland		Boundary and linear		Arable & Horticulture		Improved grassland		Neutral grassland		Acid grassland		Bracken		Dwarf Shrub Heath		Fen, Marsh, Swamp	
Square	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA
47			12.08	8.75	7.54	11.75	2.48	2.26			31.40	29.36	2.24	3.60								
63			9.08	9.82			0.82	0.82	9.97	10.39	0.88	31.01	2.13	2.60								
261															46.80	85.57					16.03	0.88
355			11.17	9.72		0.18	3.50	3.34	34.70	35.30	34.44	31.88	3.46	6.32							0.36	0.28
359			1.21	1.02	4.62	4.62	3.31	2.84	35.23	35.95	32.86	40.03	12.06	4.83							0.08	
364			0.05	0.12					95.25	93.82	0.48			0.04								
366			17.70	16.95	0.13	0.13	3.53	3.53	52.30	66.16	15.46		1.54	3.14		0.75						
383			9.43	5.34	7.48	9.53	3.03	2.48	32.44	31.26	39.11	41.41	0.94	2.46								
434		0.05	4.11	4.20	0.05		0.32				26.39	24.36	20.63	22.13								
488			0.54	0.84	0.05		0.84				80.61	80.82	7.15	8.85	1.33						3.77	3.84
657			0.40	0.47							13.33	19.98	29.53	22.83								
684		0.64	7.07	11.54	38.53	33.41					0.38	3.00	2.62		27.26	25.36	1.58	2.63	2.62	2.87	2.04	2.03
694			9.23	11.06	0.63	0.13	2.93	3.27	1.03	1.03	73.11	71.23	3.49	3.97								
695			0.24				1.68	1.62	34.12	31.67	48.14	52.18	4.88	3.62								
765					45.10	45.10	2.10	1.43					0.16		4.76	6.23			18.11	1.08	4.71	0.54
773					13.20	13.20	0.76								6.32				69.55			
920		7.80	13.26	14.02		1.74	1.21	1.21				0.17			60.20	45.91	18.59	7.67	0.60	1.80	0.72	0.46
935		6.94	36.51	17.31	1.53	2.41	1.92	1.97			3.42	1.56	3.91	11.07	1.56	9.39	0.15	0.89	0.22	1.00	15.31	10.98
991			0.45		74.25	74.52	0.97	0.97							3.70				0.04	3.76		1.41
1034															1.69	1.59			65.03	72.58		
1039			6.26	3.11			1.11	1.13	25.48	23.98	30.96	10.51	21.80	15.35	0.43						0.12	0.50
1113	0.07		0.64	0.08	45.52	46.11	3.41	3.08							13.05	17.97	1.58	1.51	10.13	12.72	17.32	9.19
1260				0.30	92.92	97.07	6.82	0.34			0.26			0.36				0.40		0.24		0.29
Mean	**0.07**	**3.85**	**6.74**	**7.75**	**23.68**	**24.28**	**2.26**	**2.02**	**35.61**	**36.62**	**26.95**	**31.25**	**7.77**	**7.41**	**15.19**	**24.10**	**5.48**	**2.62**	**20.79**	**12.01**	**6.05**	**2.76**

Table 6.1a. (continued)

BH	Bog		Standing open water		Rivers and Streams		Inland Rock		Urban		Supra-littoral rock		Supra-littoral sediment		Littoral sediment		Sea	
Square	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA	CS	QA
47			0.31	0.35					0.92	0.92								
63			0.05	0.05					51.89	21.13				1.26		2.49	16.64	12.22
261	36.57	8.86																
355			1.02	1.12	1.62	1.60			9.74	10.27								
359			1.41	1.41					0.94	1.03								
364			0.07						1.55	3.42								
366			1.58	1.58	0.95	0.95			5.64	5.63								
383									5.48	5.43								
434			0.04	0.05	0.57	0.57			0.66	1.39								
488			0.09						1.17	1.17								
657																		
684	8.68	8.61					0.11	0.87				0.72						
694			1.81	1.79					5.71	5.47								
695									10.24	10.19								
765	17.20	37.75																
773	9.22	85.50						0.36										
920	2.48	16.46	1.85	1.96			0.36	0.06	0.04	0.04								
935	1.18	1.49		0.09					3.08	3.49	2.28	7.02		3.26	0.72	3.21	27.24	16.99
991	20.13	19.99					0.30											
1034	32.62	25.43					0.59	0.34										
1039									5.17	5.87			8.11	40.41			0.20	
1113	3.23	4.27	1.60	1.40	2.34	2.34	0.49	0.01						0.18				
1260																		
Mean	**14.59**	**23.15**	**0.89**	**0.98**	**1.37**	**1.36**	**0.37**	**0.33**	**7.30**	**5.39**	**2.28**	**3.87**	**8.11**	**11.28**	**0.72**	**2.85**	**14.69**	**14.6**

Table 6.1b. Mean difference between the proportions of each BH recorded in the QA and CS surveys and the standard deviation (S.D.)

Broad Habitat	Mean difference (QA-CS)	S.D.
No allocation	*0.67*	*2.12*
Broadleaved & yew woodland	*-1.08*	*4.30*
Coniferous woodland	*0.36*	*1.74*
Boundary and linear	*-0.45*	*1.35*
Arable & Horticulture	*0.39*	*3.02*
Improved grassland	*0.27*	*8.82*
Neutral grassland	*-0.23*	*3.15*
Acid grassland	*1.12*	*9.11*
Bracken	*-0.38*	*2.31*
Dwarf Shrub Heath	*-3.05*	*15.06*
Fen, Marsh, Swamp	*-1.31*	*3.66*
Bog	*3.35*	*17.83*
Standing open water	*0*	*0.06*
Rivers and Streams	*0*	*0*
Inland Rock	*-0.01*	*0.23*
Urban	*-1.16*	*6.47*
Supra-littoral rock	*0.24*	*0.99*
Supra-littoral sediment	*1.61*	*6.73*
Littoral sediment	*0.22*	*0.72*
Sea	*-0.65*	*2.29*

6.2 Comparisons of raster data

The raster data was compared by counting the number of raster squares in the surveyors' data which matched the QA data in terms of Broad Habitat (Table 6.2a). For many squares the % agreement is good, with an overall average of 76% and some squares reaching very high levels of concordance (e.g. squares 364 and 488). In contrast, squares 261 and 773 show poor levels of agreement.

Table 6.2a Number of raster squares, matches and overall % agreement between the raster data for the QA and the surveyor data.

Square	Total number of raster squares	Total number of matched raster squares	% Agreement
47	5698	4860	85
63	9155	5395	59
261	9941	4864	49
355	10000	8882	89
359	9171	8348	91
364	9731	9533	98
366	9886	8267	84
383	9793	8507	87
434	5273	3488	66
488	9556	8949	94
657	4326	2843	66
684	9098	7702	85
694	9797	8853	90
695	9925	8274	83
765	9217	6642	72
773	9903	2236	23
920	9930	6867	69
935	9905	5127	52
991	9985	9177	92
1034	9994	7686	77
1039	9963	6013	60
1113	9937	7756	78
1260	10000	9076	91

Data showing the % agreement by Broad Habitat for each square reveal where there differences arose between the QA team and the surveyors (table 6.2b). These differences are discussed at the end of section 6c below.

Table 6.2b. Agreement by Broad Habitat (%) between the raster layer based on the QA data and that based on the surveyor data for each QA square.

% Agreement	Broadleaved & yew woodland	Coniferous woodland	Boundary and linear	Arable & Horticulture	Improved grassland	Neutral grassland	Acid grassland	Bracken	Dwarf Shrub Heath	Fen, Marsh, Swamp	Bog	Standing open water	Rivers and Streams	Inland Rock	Urban	Supra-littoral rock	Supra-littoral sediment	Littoral sediment	Sea
Square	BH1	BH2	BH3	BH4	BH5	BH6	BH8	BH9	BH10	BH11	BH12	BH13	BH14	BH16	BH17	BH18	BH19	BH21	BH22
47	87	63	100		98	42						71			100				
63	85		100	96	2	41						100			100				100
261							50			58	38								
355	93		91	98	89	14				100		90	100		95				
359	100	100	100	98	82	95						100			88				
364	46			100											45				
366	100	100	100	79		49						100	100		100				
383	86	73	98	90	89	27									97				
434	93				69	59						80	100		43				
488	47				97	68				86					86				
657	88				46	82													
684	59	97			12		100	33	54	94	100			11					
694	78		88	100	96	10						99			96				
695			79	92	81	1									98				
765		100	100				20		100	100	45								
773		100									11								
920	93		100				96	92			9	94			100				
935	86	38	97			16		6	22	75	55				87	29		17	100
991		99	100						1		89								
1034							100		79		69			100					
1039	71		95	100	100	60				22					85		21		
1113		96	95				49	64	61	77	43	99	84						
1260		93																	
Mean	**80.8**	**87.2**	**95.9**	**94.8**	**71.8**	**43.4**	**69.2**	**48.8**	**52.8**	**76.5**	**51.0**	**92.6**	**96.0**	**55.5**	**87.1**	**29.0**	**21.0**	**17.0**	**100.0**

6.3 100 point sample grid - Primary land cover codes and habitats

The 100 point grid was used to determine the proportion of Broad Habitats which had been mapped at the same location in both the CS and the QA squares at an equivalent resolution to that used in previous QA exercises. Broad Habitats which agreed at one of the 100 points were given a positive score; Broad Habitats which didn't were given a negative score. Where the area surveyed within a square was less than 100% (usually due to 'refused access' land) the number of points used reflected the survey area common to both the CS and the QA exercise. For each square, the frequency of occurrence of matching Broad Habitats for the surveyed and the QA areas were aggregated. The results show good agreement for the majority of squares, but highlight some squares where there were a number of mis-matches between the QA assessors and the surveyors (Table 6.3a). Squares with particular issues are square 773 where only 19% of the point samples matched and square 261 with a 41% match.

Table 6.3a. Concordance (+) and discrepancy (-) between QA & surveyors for Broad Habitats

Square	+	-	% Concordance
47	41	11	78.85
63	50	43	53.76
261	41	59	41.00
355	84	16	84.00
359	84	7	92.31
364	90	4	95.74
366	83	16	83.84
383	78	22	78.00
434	30	23	56.60
488	91	5	94.79
657	31	17	64.58
684	76	15	83.52
694	91	9	91.00
695	84	16	84.00
765	70	23	75.27
773	19	80	19.19
920	60	40	60.00
935	43	56	43.43
991	77	23	77.00
1039	73	27	73.00
1113	56	43	56.57
1260	76	23	76.77

A matrix of Broad and Priority Habitat agreement is presented in Table 6.3b. This shows the number of matches/mismatches for each of the Broad and Priority Habitats based on the 100 point grid comparison.

Table 6.3b Matrix of Broad and Priority Habitat codes for the 100 point analysis.

	Broadleaved Mixed and Yew Woodland	Coniferous Woodland	Boundary and Linear Features	Arable and Horticulture	Improved Grassland	Neutral Grassland	Acid Grassland	Bracken	Dwarf Shrub Heath	Fen, Marsh, Swamp	Bog	Standing Open Waters and Canals	Rivers and Streams	Inland Rock	Urban	Supra-littoral Rock	Supra-littoral Sediment	Sea	(ph) Upland Mixed Ashwood	(ph) Wet Woodland	(ph) Lowland Mixed Deciduous	(ph) Fen	(ph) Purple Moor Grass Rush Pasture	(ph) Blanket Bog	(ph) Sand Dune	(ph)Strandline/Coastal Vegetated Shingle	(ph) Coastal Saltmarsh	(ph) Northern Birchwood	
	CS BH/PH																												
QA BH/PH	1	2	3	4	5	6	8	9	10	11	12	13	14	16	17	18	19	22	25	26	28	35	36	38	42	43	44	45	Grand Total
1	68	7	2	2	1	4	10					1			1				2	1	4	4			2			1	110
2	7	323					1	1	2					1										1					336
3		4	35								1																		40
4				301	6	2									3														312
5	5	1		18	351	42									1										22				440
6					42	60									2		2								9				115
8						1	118	4	1		12											5		19					160
9	2	3						9	3													3							20
10		1					2		71		2													90					166
11					1	7	20			3	5											11		2					49
12							4		18	1	14													23					60
13					1							6																	7
14							1						10																11
16									1																				1
17	2				27										70														99
18																2													2
22						1										4		29								3	6		43
24																					6								6
25																			2										2
26	2																			2		1							5
28	6	5																	4		3								18
35							7			1			1									4		1					14
36																							3						3
38		1					27		4		3													31					66
42																									6				6
44																											1		1
Grand Total	92	345	37	321	429	117	190	14	100	5	37	7	11	1	77	6	2	29	8	3	13	28	3	167	39	3	7	1	2092

Tables 6.2b and 6.3b indicate that for the majority of Broad and Priority Habitats there is a good match between the QA and the surveyors (76 and 73% respectively). Discrepancies are relatively minor in most cases, e.g. those between Broad and Priority Woodland Habitats. In some cases there are more substantial inconsistencies, such as in the recording of: 1) Neutral and Improved Grassland, 2) Broadleaf and Coniferous woodland, 3) upland habitats including Acid Grassland and Fen, Marsh and Swamp, 4) Urban and Improved Grassland and 5) Sand Dune Priority Habitat, explored below.

1) Inconsistencies in the recording of Neutral and Improved Grassland are a recognised issue in the survey as these Broad Habitats overlap in terms of species present. Where they do so the decision to record as one or the other may depend on perception of species cover or factors such as the amount of these Habitat types already encountered and their species composition and cover. This issue arose in a number of squares but there was no bias in the Broad Habitat used by surveyors or the QA team (e.g. for example, the QA team were not more likely to have assigned polygons to Improved grassland than the surveyors). Table 6.3b shows a good balance between the numbers of areas which were potentially 'wrongly' allocated to either Improved or Neutral (i.e. 42 in both cases).

2) The definition of woodland used in the key stated that it contained 'Vegetation cover consisting of over 25% cover of trees or shrubs over 1m high' and Coniferous Woodland should contain 'More than 20% coniferous in canopy (excluding yew but includes juniper)'. The Broad Habitat definition for Coniferous Woodland states that the cover of coniferous trees should exceed 80%. Unfortunately despite the fact that the key was widely circulated prior to CS we failed to pick up this issue. In the majority of cases this will not have affected the woodland data as most woodlands are predominantly either Broadleaved or Coniferous. As with the Neutral/Improved Grassland issue Table 6.3b shows that there was a good balance between the numbers of areas that were potentially 'wrongly' allocated to either Broadleaf or Coniferous woodland. Further issues with woodland arise because in previous surveys surveyors in the field did not directly allocate areas to Broad Habitats but rather areas were assigned to Broad Habitats on the basis of attributes recorded using an allocation matrix. The uncertainties resulting from this and how back-allocation and the new approach to allocating areas to woodland types relate to it are yet to be fully explored.

3) Upland habitats are variable at the small scale and inconsistencies may reflect a spatial difference in the mapping of the location of these areas of Broad Habitat. In many cases upland habitats were mapped as mosaics comprising of a number of Broad Habitats. In order to carry out this analysis the mosaics have been disaggregated into their component Broad Habitats which will have resulted in mis-matches, where 'like' BH's within the same mosaic have been matched with 'different' BH's. This was a particular issue in squares with large amounts of mosaic mapped (e.g. squares 261, 765 and 1113). This is as much of an issue with the methods used for comparison as with the data. It illustrates the difficulties of mapping precisely in a continuum of Broad Habitats which grade into one another.

4) The inconsistency in the recording of some areas of Urban as Improved Grassland potentially reflects a lack of clarity in the recording of Urban areas. In the field handbook advice is given that ‘Amenity grass>1Ha should be recorded as Improved Grassland’. Further analysis on this issue is ongoing.

5) The differences in the recording of Sand Dune vegetation are down to a QA square in Scotland containing machair habitats. The machair is an uncommon habitat with a localised distribution; hence this issue is not likely to have had a major effect on survey results. Whilst the machair occurs on sand dunes its vegetation is essentially that of the Neutral and Improved Grassland Broad Habitats. This anomaly highlights the potentially confounding issue of location and landscape, alongside vegetation, on habitat recording.

The most substantial discrepancies occurred in the recording of Blanket Bog. This habitat, which is particularly important in Scotland, has been the subject of debates about definitions for the organisations responsible for maintaining its extent and quality. The key to vegetation types and Habitats which formed part of the Field Handbook for the Survey was revised and updated for 2007 and included revisions, resulting from the debates between the experts, to the description of Blanket Bog. The matrix of habitats in table 6.3b indicates that the main mis-matches were between Dwarf Shrub Heath and Blanket Bog. This particularly affected square 773 in which concordance between surveyors and the QA team was very low. The different nature of these two habitats makes it possible that they can occur in the same location with the Dwarf Shrub Heath overlying the Blanket Bog. Guidance in the Field handbook advises that where this occurs ‘wet heath is differentiated from blanket bog by the absence of species of wet/deeper peats’. This particular square was covered by a team of surveyors early in the survey who were unused to these habitats having largely worked in lowland squares previously. It is apparent that experience and training are likely to have helped minimise the differences between the QA team and the surveyors. However, there was also an issue with Bog and Dwarf Shrub Heath affecting a number of polygons in squares 991 and 1034, emphasising the difficulties of habitat definition in upland areas.

The Blanket Bog Habitat was a difficult one to make a judgement on for the QA assessors as well as the surveyors. Strict adherence to the key often resulted in areas being allocated to different habitat types when intuitively (on the basis of location and landscape structure) Blanket Bog would have been the chosen Broad Habitat. Given the importance of the Blanket Bog habitat these discrepancies are of concern and further work on definitions using the CS data (both mapped and plots) is proposed.

6.4 Polygon analysis of species attributes

Species recording

In each of the polygons surveyed, alongside vegetation types (nested within Broad Habitats), surveyors also recorded 2-3 dominant species. The polygon datasets comprise of 1081 polygons in the QA dataset and 998 polygons in the CS data. After spatial matching 45% of the QA polygons had at least one listed species which was present in the location matched CS polygon (excluding polygons which were a mosaic of Broad Habitats).

Table 6.4 summarises the data for the listed species from a joined polygon dataset, i.e. where polygons in both CS and QA datasets are matched in terms of position and presence of at least one of the same species. The analysis has been carried out by comparing the list of species selected by the QA team to the CS list of species. Results are presented by different Broad Habitat types to show differences between the Broad Habitats in terms of commonality of species in the CS and QA datasets.

Table 6.4. Number of polygons with matching species for matched and unmatched Broad Habitats.

	BH Match?				
Broad Habitat	No match	%	Match	%	Total
Broadleaved & yew woodland	15	7.5	184	92.5	199
Coniferous woodland	14	11.4	109	88.6	123
Boundary and Linear features	2	100.0			2
Arable and Horticultural	4	44.4	5	55.6	9
Improved grassland	65	18.4	289	81.6	354
Neutral grassland	50	46.7	57	53.3	107
Acid grassland	33	27.0	89	73.0	122
Bracken	16	44.4	20	55.6	36
Dwarf shrub heath	31	24.2	97	75.8	128
Fen, Marsh and Swamp	54	45.0	66	55.0	120
Bog	10	9.3	97	90.7	107
Standing open waters and canals	2	100.0			2
Inland rock	2	100.0			2
Littoral sediment			1	100.0	1
Sea	8	100.0			8
Grand Total	306	23	1014	77	1320

Further analyses of these results will make it possible to understand how species composition affected Broad Habitat allocation by QA and CS teams. In-depth analyses of this sort are not within the scope of this report but may well contribute to refining BH definitions for possible future surveys and beyond. The proportion of polygons, matched on location, with matching species (45%) is somewhat lower than expected but may in part result from the assignment of polygons to different BH's. The choice of Broad Habitat is likely to influence the choice of 'dominant' species as surveyors will tend to choose the species that support the chosen Broad Habitat. Whilst it is possible to compare this figure unfavourably with rather higher figures for concordance for species in previous QA exercises, the scale and level of detail in this QA exercise far exceeds previous exercises and concordance across a far broader range of habitat types than covered previously is likely to be lower.

The results presented in Table 6.4 indicate that for 77% of polygons in which there was a species match between the QA and CS data there was also a match in the chosen BH. For some Broad Habitats, the proportion of polygons with a matched species was relatively high, e.g. woodlands and Improved Grassland. For these habitat types, there are often conspicuously dominant species. Grassland types other than Improved Grassland tend to be composed of a larger number of species with lower coverage, making it less likely that both CS and QA surveyors would choose the same species. For Bracken the relatively low percentage of concordance in Broad Habitat despite identical species choice is not unexpected. The definition of Bracken is a cover of >95% Bracken, differences in the assessment of cover of even 5% will result in the assignment of a polygon to different Broad Habitats. Upland habitats were often contained within mosaics representing a range of different Broad Habitats making comparison between polygons more complex. Polygons with Broad Habitats in mosaics are excluded from this analysis.

The results of the analyses on the listed species within spatially matched polygons are presented in Table 6.5. Overall the concordance in this list of the most commonly recorded species was around 40%. The QA team tended to record higher numbers of species. For the most commonly recorded species in the survey as a whole, *Lolium perenne*, concordance between QA and CS surveyors was higher (66%) reflecting ease of identification and dominance of this species in swards where it is present.

Table 6.5 Comparisons of listed species within spatially matched polygons

Species	Common	Surveyors only	QA only
Lolium perenne	212	41	66
Juncus effusus	132	58	91
Calluna vulgaris	123	117	118
Molinia caerulea	103	42	49
Holcus lanatus	78	102	80
Agrostis capillaris	76	66	78
Eriophorum vaginatum	58	27	60
Trifolium repens	57	26	86
Pteridium aquilinum	54	16	33
Deschampsia cespitosa	49	41	37
Nardus stricta	43	49	25
Cynosurus cristatus	24	21	12
Erica cinerea	23	9	35
Trichophorum cespitosum	19	20	18
Eriophorum angustifolium	17	23	12
Rhododendron ponticum	15	4	11
Juncus articulatus/acutiflorus	15	28	13
Dactylis glomerata	12	17	18
Anthoxanthum odoratum	11	57	9
Myrica gale	9	3	12
Birch	67	23	63
Larch	50	25	36
Spruce - Sitka	49	29	33
Oak	32	36	25
Pine - Scots	26	24	12
Ash	16	42	24
Alder	14	6	19
Sycamore	12	24	1
Willow	12	20	10
Hazel	9	29	8
Total	**1417**	**1025**	**1094**

7. Comparison of QA and CS data – Linears

7.1 Direct comparison of aggregate linear features for whole squares.

A comparison of the total lengths of features in each square (Table 7.1, Fig 3) showed minor differences between the QA and the CS data in most squares. Linear data was collected under a range of land use themes within which different feature types were nested. Land use themes included: Banks (B), Fences (F), Forestry e.g. belts of trees (FO), Inland Physiography (IL), Inland water (IW), Transport (TR), Walls (W), Woody linears in which trees take their natural shape (WNS), Woody linears in which trees take an unnatural shape (WUS) and other un-reported linear feature types. Data on the lengths of linears grouped under the different land use themes is included in Table 7.1.

Table 7.1. Comparison of the total lengths and the lengths of linear features grouped by Land Use Theme (as above) per square (m) in the QA and the CS data.

	CS Linear feature												QA Linear feature													
Square	AN	B	F	FO	GS	IL	IW	TR	US	W	WNS	WUS	AN	B	F	FO	IL	IW	ST	TR	US	W	WNS	WUS	**Total QA**	**Total CS**
47		2900	5554	65			4168	2534			1331	3441		2979	5526	65		4278		2317	7		1765	3247	19993	20185
63			1279	264			194	2363			658	478		228	1946			194		2974			683	599	5236	6623
261			862			220	9735						220		265		220	6962							10817	7666
291		563	10515			70	4123	5512			2523	3649		616	10709	301	70	3078		4246			2829	3565	26955	25415
355			6296	81			782	1618		623	3369	2465			6780	984		1156		2664		618	1137	2026	15234	15365
359			10736				8552	2249			2253	9767			10089			8096		1692			2086	8535	33557	30497
364			369				4668	1333			1524	1948			280			4673		1334	509		1859	1999	9842	10653
383		437	11215		67		4447	4590			1229	4228		469	11517			4513		3684			1561	7018	26213	28763
434	12		3672				795	808	35		1358	881		842	7091			1977		1232	13		3625	2185	7561	16964
488		3911	12486	215	1076		5187	428	155		5090	5888		5092	11285			6332	105	327	13		4530	4313	34436	31996
657		1429	12061	214			2851	2562	27	1116	2753	5958		1371	11914			3000		2560		1103	2818	5841	28971	28606
684			2389				2015	1854		2749					2490			1866		1713		2577			9007	8646
694			10610	44			1556	2921	10	3358	1072	1478			10485	105		1576		2997	11	2869	1714	1540	21049	21297
695			11432				365	1426	67	208	4201	6938			14772			408		592	40	392	3336	8156	24637	27696
765							4394											4352		78					4394	4430
773			2064				2787			486					2063			2647				484			5337	5194
920			2181			629	4383	881							1947		629	5162		952					8074	8690
935		102	6347				2289	485		2115	736				6736			2183		394	185	2123	1071		12074	12692
991			1200				2298	1038							1200			2421		1038					4536	4659
1039			6293	33			2241	814		475	376	130			6508	33		2239		813		474	375	130	10362	10570
1113			2765			91	5038	1923	8	193	66			564	3105		119	7523		858		193	679		10084	13040
1260		453	2256				999	808	458			1024			163			1136							5998	1299

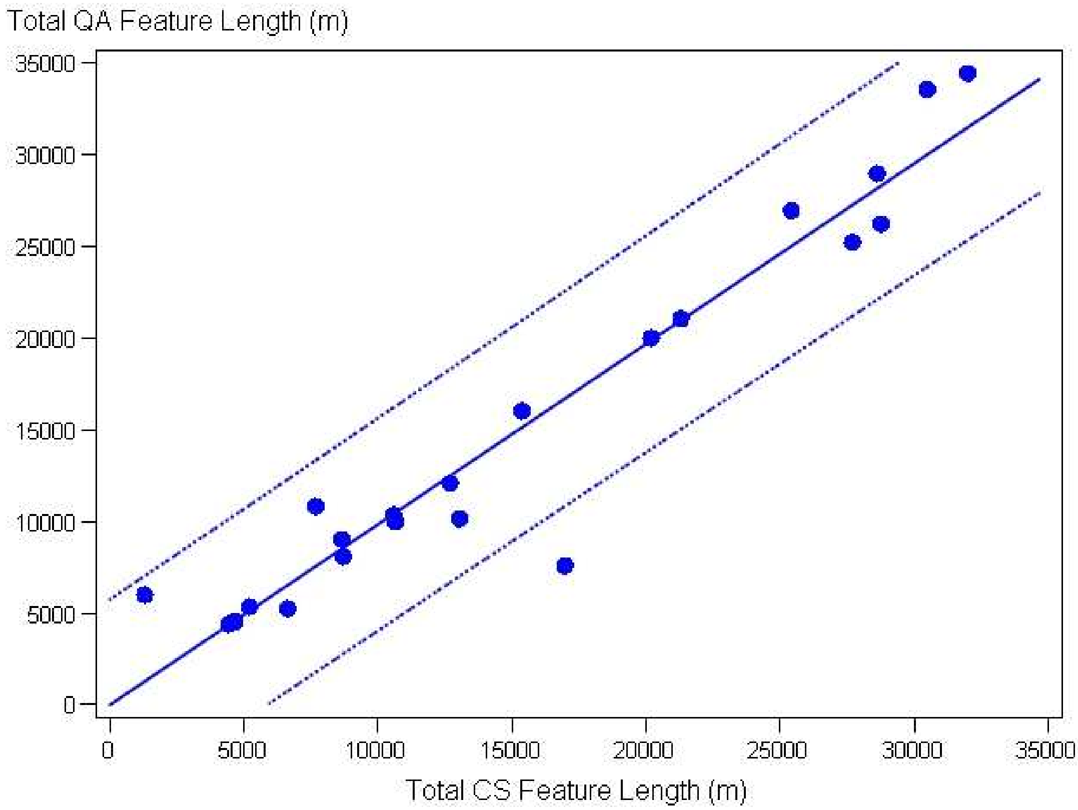

Figure 3. Aggregate lengths of features per square for countryside survey data plotted against the QA data with 95th percentile confidence limits (r^2=0.93).

7.2 Land use themes for linear features

Due to the fact that a 100 point analysis is unlikely to target linear features adequately, this analysis was carried out using a different method. Linear features in the QA dataset were buffered with a small 5m buffer and matched to linear features in the CS dataset. This enabled a comparison of the land use themes of the QA linear features with those of the co-located CS linear features. This method fails to pick up on the few features recorded in the CS data but not in the QA data, hence there is no 'no match' column under the listed QA Linear features.

Table 7.2. Matrix of Linear feature land use themes (as in 7.1) for the joined QA and CS linear datasets.

Count of Match	CS Linear feature											
QA Linear feature	B	F	FO	IL	IW	TR	US	W	WNS	WUS	No Match	Grand Total
AN											1	1
B	93										1	94
F		1289									1	1290
FO			8								5	13
GS											4	4
IL				15								15
IW					591							591
TR						267					2	269
US							18				2	20
W								119				119
WNS									374			374
WUS										474	1	475
Grand Total	93	1289	8	15	591	267	18	119	374	474	17	3265

This analysis shows a very high level of consistency between the QA and the CS surveyors in choice of land use themes for linears and a very limited number of cases where the QA surveyors had recorded a feature and the CS surveyors had not. For most of these features, this level of consistency would be expected. It is, for example, pretty unlikely that the QA team would have chosen a fence and the CS surveyors, a wall, to represent the same feature in the same location. There is more potential for confusion between forestry features like belts of trees (FO) and woody features in which trees do, or do not take their natural shape (WNS, WUS), but there is no evidence of problems with these features in the comparison above.

8. Comparison of QA and CS data - Points

8.1 Direct comparison of aggregate points for whole squares.

A comparison of the total number of points in each square (Table 8.1) showed very minor differences between the QA and the CS data in most squares. Point data was collected under a range of different land use themes which included: Forestry (FO), Inland Physiography (IL), Inland Water (IW), Structures (ST) and Veteran Trees (VT). Data on the numbers of points grouped under the different land use themes is included in Table 8.1.

Table 8.1 Comparison of the total number and numbers of points grouped by land use theme (as above) per square (m) in the QA and the CS data.

	CS						CS Total	QA						QA Total
BLK	FO	FOF	IL	IW	ST	VT		FO	FOF	IL	IW	ST	VT	
47	15						15	15						15
63	3						3	3						3
261	3		4				7	3		4				7
355	7					4	11	8						8
359	100			3	2	2	107	96	1		2	2	3	104
364	19					1	20	20						20
383	58			1		3	62	58	1		1		1	61
434	1						1	2						2
657	14						14	13					4	17
684	18						18	18						18
694	30				3	1	34	31				3	2	36
695	25					1	26	30					3	33
765			2				2			2				2
920	2				1		3							
935	24				1		25	33			1	1		35
991	13						13							
1039	3						3	6						6
1113	3						3	5						5
(blank)														
Grand Total	338	0	6	4	7	12	367	341	2	6	4	6	13	372

The theme comprising the largest number of points is forestry which includes individual trees alongside clumps of trees, patches of scrub etc. ‘Veteran trees’ was the land use theme with the greatest disparity between the CS and the QA datasets. Given that surveyors’ instructions were to select up to two relevant trees (on the basis of tree size and species) per species, and they may have had the choice of many, it is not surprising that concordance between veteran trees is low.

b) Habitat codes for point features

As for linear features, a 100 point analysis was considered inappropriate for the analysis of point features. Point features in the QA dataset were buffered and matched to point features in the CS dataset. This enabled a comparison of habitat codes of the QA point features with those of the co-located CS point features. This method fails to pick up on the features recorded in the CS data but not in the QA data as well as on points which are not located in quite the same proximity, but it does allow a cross check on the use of habitat codes for points.

Table 8.2. Matrix of Point feature habitat codes for the joined QA and CS linear datasets.

Count of Match	CS Point feature							
QA Point feature	Scattered trees	Woodland	Scattered scrub	Timber prod'n	Regrowth Cut stump	Area of small water bodies	Playing field	Total
Scattered trees	176	2	2		6			186
Woodland	4	6						10
Scattered scrub		1	12		6			19
Timber prod'n	1		1	6				8
Regrowth Cut stump					2			2
Dead tree	1							1
Windblow	1							1
Area of small water bodies						5		5
Playing field							2	2
Total	183	9	15	6	14	5	2	

In almost all cases surveyors chose the same habitat codes for the same points as the QA team. In a few cases there was confusion in the use of a 'woodland/forest' code which is likely to result from the inappropriate use of this code in previous surveys and its retention in the data.

9. Assessment of change recording

In the 2007 mapping exercise, as in previous surveys the surveyors were asked to record where change had occurred. In 2007, with digital recording, it was possible to ask the surveyors to provide more detail on the kind of change being recorded than they had done previously. Change codes in 2007 included:

- Real change – a physical change in the area/point/linear feature recorded in the data previously
- Error Change – a change to the data because of a wrong assignment in the previous data, e.g. a feature that was an established hedge (in 2007) had been present in the data as a fence (in 1998)
- No change - the feature in 2007 is the same as that recorded in 1998

The inclusion of an 'error change' field provided the potential to update the previous data with the correct code, using this field in combination with others.

This process was quite complex and it is known that some surveyors struggled with it initially (the Quality Control exercise early in the survey was used to help clarify the issues). An analysis of the use of the 'change' field for areas, linear and points has been carried out to assess the extent to which surveyors and the QA team agreed on the correct 'change' field to use (Tables 9.1-9.3).

Table 9.1. Comparison of assignment of 'change' between the CS and QA surveyors by Broad Habitat, (+) agreement on assigning change, (o) QA noted change – surveyors did not, (-) Incorrect change assigned by surveyors.

Broad Habitat	+	% +	o	% o	-	% -
Broadleaved Mixed and Yew Woodland	117	87	13	10	4	3
Coniferous Woodland	107	93	6	5	2	2
Boundary and Linear Features	47	98	1	2	0	0
Arable and Horticulture	39	89	5	11	0	0
Improved Grassland	141	92	10	6	3	2
Neutral Grassland	59	74	20	25	1	1
Calcareous Grassland	79	64	25	20	20	16
Acid Grassland	20	63	8	25	4	13
Bracken	48	87	5	9	2	4
Dwarf Shrub Heath	42	86	4	8	3	6
Fen, Marsh, Swamp	51	82	7	11	4	6
Bog	10	71	3	21	1	7
Standing Open Waters and Canals	9	100	0	0	0	0
Rivers and Streams	2	67	1	33	0	0
Montane	84	94	4	4	1	1
Inland Rock	12	92	1	8	0	0
Urban	31	100	0	0	0	0
Littoral Sediment	8	100	0	0	0	0
Sea	5	100	0	0	0	0

This table (9.1) indicates that for most Broad Habitats there was broad agreement between the surveyors choice of 'change' code and that used by the QA team.

Table 9.2. Comparison of assignment of 'change' between the CS and QA surveyors by linear feature type, (+) agreement on assigning change, (o) QA noted change – surveyors did not, (-) Incorrect change assigned by surveyors.

Linear feature type	+	% +	o	% o	-	% -
WLF natural shape	33	100	0	0	0	0
WLF unnatural shape	41	95	2	5	0	0
Roadside ditch, Sampled	5	100	0	0	0	0
River Sampled	0	n/a	0	n/a	0	n/a
Clump of trees	1	100	0	0	0	0
Mortared wall	25	100	0	0	0	0
Other wall	6	100	0	3	0	0
Fence - iron only	32	97	1	0	0	0
Fence - wire on posts	1	100	0	2	0	0
Other fence	86	98	2	2	0	0
Cliff 5-30m high	2	100	0	0	0	0
Grass strip	0	n/a	1	100	0	0
Rocky/boulder shore	3	100	0	0	0	0
Unconstructed track	15	100	0	0	0	0
Footpath (exclusive)	10	100	0	0	0	0
Surface boulders	6	75	2	25	0	0
Roadside ditch	45	100	0	0	0	0
Other ditch	16	100	0	0	0	0
Spring	18	100	0	0	0	0

As for the assignment of habitat codes to linears, there was a very high level of agreement between surveyors and the QA team on the change codes used for linear features.

Table 9.3. Comparison of assignment of 'change' between the CS and QA surveyors by point feature type, (+) agreement on assigning change, (o) QA noted change – surveyors did not, (-) Incorrect change assigned by surveyors.

Point feature type	+	% +	o	% o	-	% -
Scattered trees (201)	160	98	2	1	1	1
Scattered scrub	10	100	0	0	0	0
Woodland	11	100	0	0	0	0
Timber prod'n	6	100	0	0	0	0
Area of small waterbodies	5	100	0	0	0	0
Grand Total	968	87	105	9	41	4

Use of change code on point features was highly consistent between surveyors and the QA team.

Overall the coding of change appears to have been carried out according to the protocols. Change is at times difficult to assess, particularly in relation to habitats and this is reflected in the lower agreement between surveyors and the QA team for 'change' on Broad Habitats in comparison to that on linear and point features. The shift to a digital system provided an opportunity for surveyors to 'tidy' up the data as they were in the field providing final habitat maps (as had never been done before, given the previous method of surveyors recoding on paper maps and digitising taking place in the office). In the main, this process of getting surveyors to both provide new data on change and verify the previous maps worked well, despite the complexity it added to the survey.