2: Statistical aspects

Main Woodland Survey

Sampling – woodland selection

A digital map of all woodland showing Interpreted Forest Types was derived from 1:25 000-scale aerial photography (directly for England and Wales and using the Land Cover Map of Scotland 1988). The digital map gives the extent of all woodland of ≥2 hectares. This was updated as survey work progressed by adding new planting since the photograph date (using administrative records verified in the field survey) and any woodland that had been obscured by cloud in the photo. The total area of the woodland stratum in each county or district was obtained from the digital map. Forestry Commission (FC) maps, now in digital form, were used to divide the woodland stratum into FC and non-FC ownership.

The woodland map formed the sampling framework for the Main Woodland Survey. No fieldwork for the Main Woodland Survey was undertaken outside the mapped woodland area.

From the digital map, the area of each wood (as defined in Appendix 2) was recorded and this information was used to determine the probability of each woodland being sampled.

A feature code was allocated to the wood according to its size class (see below). Any wood <2 hectares was removed from the data. This dataset was used to select those woodland parcels to be sampled. Within each 100-km tile (as per the Ordnance Survey (OS) tiles, e.g. tile SU) the woods were:

1. Stratified into size classes:
   - 2 – <100 hectares
   - 100 – <500 hectares
   - ≥500 hectares

(Note: Woodlands were split into class sizes in terms of sampling as it related to the woodland area within the OS 100-km tile; reporting was based on country level data after the tiles (and woods) were joined together).

2. Put into ascending order within each stratum, with the sampling for some woodlands split by one or more tile edges.

3. Selected according to a ratio of:

2 – <100 hectares:	1 wood in 5
100 – <500 hectares:	2 woods in 5
≥500 hectares:	all woods

For Great Britain there were about 3 000 woods of more than 100 hectares, together accounting for two-thirds of the total woodland area. There were also about 80 000 woods of 2–100 hectares, together accounting for about 30% of the woodland area. The remaining 5% of area comprised woods of <2 hectares, and these were included in the Survey of Small Woodland and Trees.

Generating the ground samples

Within sampled woodland a number of 1-hectare sample squares were to be selected for field survey work. This was a change of practice from all previous Census surveys, where whole woods had been selected for survey. The aim of the sampling scheme was to obtain information efficiently about variation over all woodlands. Particularly for large plantations, there was expected to be more variation between woods than within woods. The basic survey design, used for woods of 2–100 hectares, was to select one wood in five and apply a sampling grid with 5% density, giving a 1% sample. For larger woods, this design would have given numerous sample squares in the selected woods, and no information from woods not selected, so it was agreed that for them it would be more efficient to select a larger fraction of woods and use a less dense grid. For woods of 100–500 hectares, two woods in five were selected and a 2.5% density grid applied, giving a 1% sample. For woods of 500 hectares or more, all were selected and a 1% density grid applied, again giving a 1% sample.

Various sampling options were devised and tested on a grid (measuring 800 m east and 500 m north – 40 hectares in total) to fit on OS 1:25 000-scale maps. In the sampling grid, the squares were clustered, rather than being spread regularly or randomly over the whole area. The clusters consisted of various configurations of two to five 1-hectare squares in the 800 m x 500 m grid panels, with a random allocation of which panels contained clusters. Trials had established that clustering in this way substantially reduced travel costs for the survey teams, without significant deterioration in the standard errors of the results. Within the grid framework

finally chosen there were various clusters of 1-hectare squares (see examples in Figure 3).

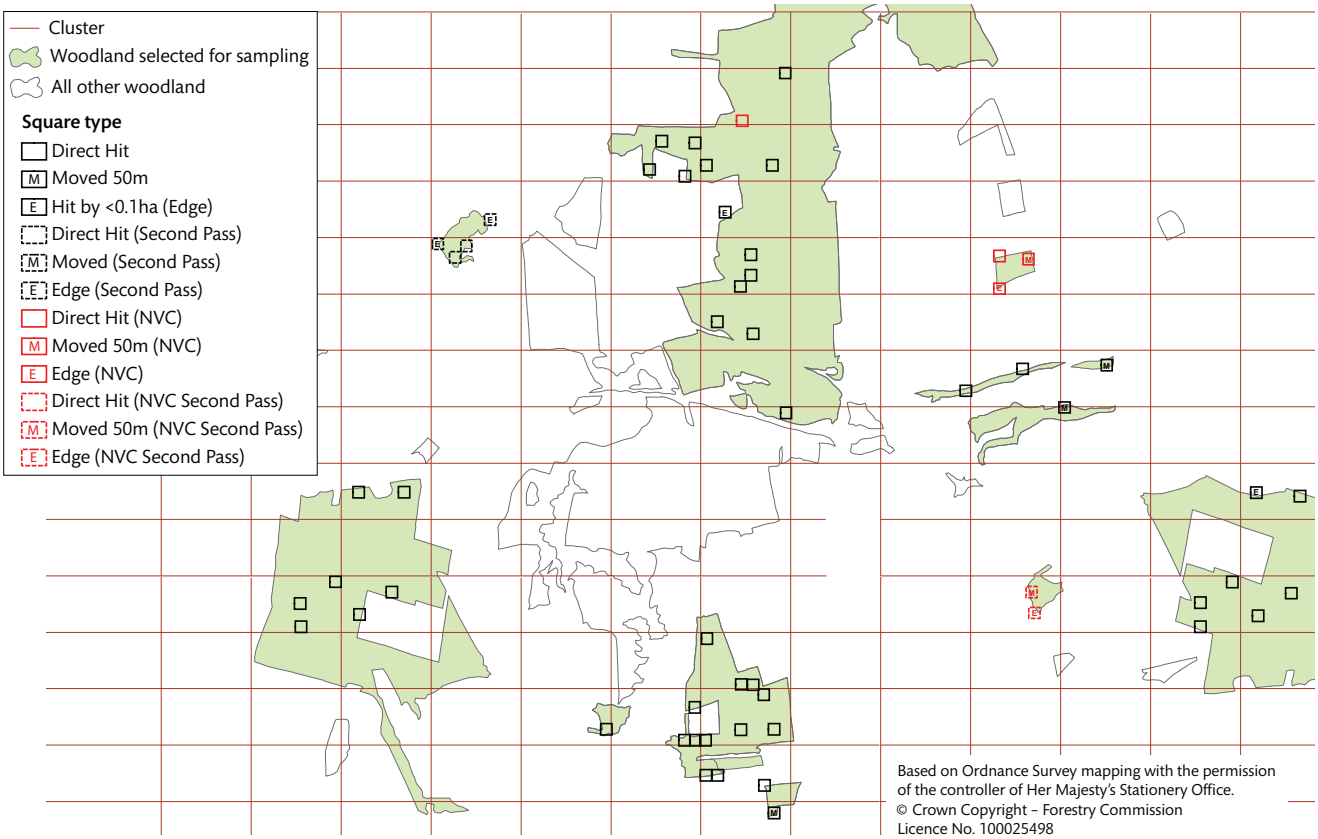
The 1-hectare sample squares could fall in a number of places and were dealt with in different ways:

- 1-hectare squares falling completely outside the woodland were ignored (and not recorded).
- 1-hectare squares with ≥10% of the area falling within woodland were sampled.
- Squares containing <10% woodland were recorded as edge squares, but were ignored by the ground survey.
- If the only squares selected for a wood were edge squares (0.1–0.5 hectare) then one of those squares was randomly selected and moved 50 m towards the centre of the wood. Moved squares were treated as normal during the ground survey, but data from them were given a reduced weight in the aggregation to scaled-up results, roughly balancing the increase in area resulting from the movement. (Note: The idea of the 'move' was to increase the amount of information gathered for the travel costs associated with the survey, but squares were only moved under particular circumstances, i.e. where all the squares for a sampled woodland were 'edge' squares and even then only one square was moved 50 m towards the centre of the woodland, adding up to approximately 0.5 hectare of woodland).

Apart from the under-representation of woodland edge, as a result of moving squares away from the edge and omitting squares, the sampling procedures described up to this point were thought to give an unbiased sample amounting to ≈1% of all woodland of ≥2 hectares. The pilot study indicated that the procedures were generally satisfactory, but there was concern that the cluster grid pattern resulted in many of the selected sample woods of 2–100 hectares not including any squares. This meant that the initial results for woods of 2–100 hectares were based on data from markedly less than the initial selection of one wood in five, because of closer geographical grouping of squares than expected and hence worse precision for the given sample size.

It was therefore decided to boost the sample by a second pass of the sampling grid, which ensured that each of the one-in-five selected woods of 10–100 hectares included at least one sample square. These additional sample squares were treated as normal during the ground survey, but all data from the size class were given a reduced weighting in the aggregation of the results, roughly balancing out the increase in sample area from the additional squares. This resulted in estimates that were still unbiased, but with improved precision from the additional squares and the distribution of squares over more woods.

Figure 3 Example of the distribution of 1-hectare squares on selected woodland.



An initial estimate of the area of the woodland stratum was obtained from the sample results, rating up all areas either by 100 (which assumes that the original data represents 1% of the woodland area) or by an alternative lower weighting to take into account the moved or added squares (as described above). By chance, these sample results could be above or below the actual Forestry Commission and non-Forestry Commission areas for each geographic unit (county or former Scottish district), as measured from the original digital map. The rated sample results for each geographic unit were scaled up or down, to match the areas from the digital map. The National Inventory therefore has an internally consistent set of areas from the digital map and final rated sample data.

The woodland stratum could include some areas incorrectly classified as woodland, e.g. the land cover could be gorse. In addition, some areas could have been felled and permanently converted to another land use since the date of the aerial photography or doubt could have been expressed by the photo interpreter using a non-forest classification. While it was not possible to identify all such areas from field survey, as not all woodland areas were visited, an unbiased estimate was obtained by recording such sections of sample squares as 'non-woodland', leaving them to be rated up in the same way as data from any other sections. The rated-up areas were subtracted from the total areas of the woodland stratum, to produce the final main woodland results for publication. Processing of the Scottish data initially followed a different approach, trying to identify all areas of 'non-woodland' and removing them from the digital map. This may have resulted in a slight overestimate of the true woodland area.

Standard errors and precision

Approximate standard errors were calculated for the main report. Each wood was allocated to one of eight categories, based on the mix of Interpreted Forest Types (see Appendix 1) of all polygons making up the wood in the aerial photographic interpretation. Within each of these categories, the data were treated as if they came from a simple random sample. For any variable of interest (e.g. species), the area and its variance were estimated for each category, added up over the categories, and the overall standard error calculated. In this calculation, the sample size for a category was the total number of sample squares in that category, implicitly including zeros for the many squares that had no area of a given species. Within a sample square it made no difference whether the area was in a single section unit or in more than one – it was treated as a single total for each sample square.

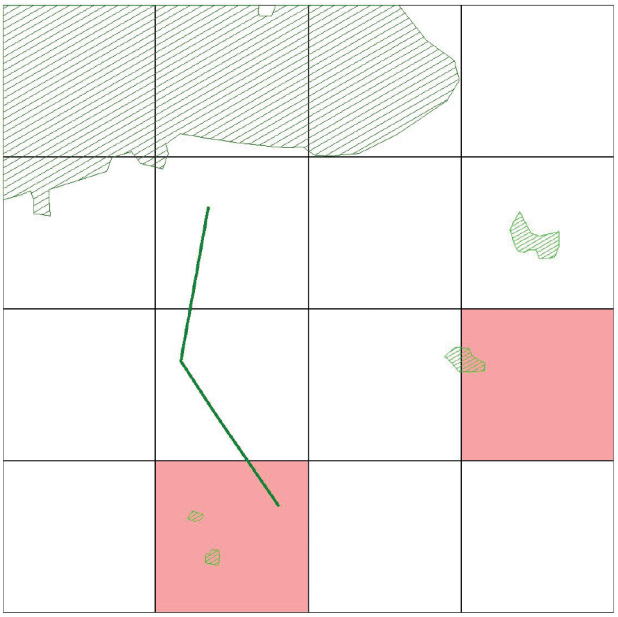
This calculation is likely to underestimate the standard errors, because it ignores the more complicated aspects of the survey design. Sample squares tended to be grouped more closely together than in a random design because of the use of a cluster grid, and also because of the two-stage sample design in which the first stage was the selection of woods and the second stage the selection of sample squares within woods. Also, although rating-up all results to equal the areas from the digital map produced more accurate totals, it produced a small loss of precision in attribute estimates.

Survey of Small Woodland and Trees

The land area of Britain was stratified into coastal and inland 1 km x 1 km squares. A systematic sample of 1 km x 1 km squares was then selected, representing around 1% of the land area of each county or former Scottish district (Figure 4). This was preferred to a regular grid (e.g. choosing the southwest 1 km square from each 10 km x 10 km tile) because it reduced the chance variation in sampling intensity of coast and of each local authority. 1:25 000-scale aerial photo interpretation (API) was then used to identify features in each sample square – i.e.:

- small woodland (0.1 – <2 hectares)
- linear features.

Figure 4 Sample grid square, 1 km x 1 km. This diagram shows an example with one linear feature (green line) and four small woodlands. The large area at the top left is outside the scope of the survey either because it is woodland >2 hectares (and therefore in the scope of the Main Woodland Survey), an urban area (where all woodlands were excluded), an orchard or an area of water.



Each 1 km square was then divided into 16 parts, and two of these (shaded red in Figure 4) were selected for field data collection. The method of selection ensured that the two parts were not adjoining, to limit the effects of clustering of data within the square. The first was selected at random from 1–10 (counting row by row, left to right and starting from the bottom left), and the second added 6 to the count. Although this gives parts 7–10 a higher probability of selection, there was no relationship found between the precise location of grid squares and wood/tree features, so this did not bias the results.

Ground sampling looked at four features and was undertaken to collect information about each feature located in the two selected sub-squares:

1. Small woodlands (0.1 – <2 hectares).
2. Linear features – ≥25m in length and four times as long as wide, separated into:
   - 2.1 narrow linear features (with a width of 16m or less)
   - 2.2 wide linear features (with a width greater than 16 m).
3. Groups – two or more trees with an area <0.1 hectare.
4. Individual trees – a tree with a crown that has no contact with any other tree crowns and which is at least 2m tall, separated into:
   - 4.1 boundary tree (an individual tree on any boundary)
   - 4.2 middle tree (an individual tree not on a boundary).

For small woodlands and wide linear features, each of which had an area of at least 0.1 hectare, this survey collected a similar range of data to the Main Woodland Survey sample, including species, age and stocking. This enabled the addition of the results to those from the Main Woodland Survey, to produce statistics for all woods over 0.1 hectare. A more limited range of data were collected for narrow linear features, groups and individual trees.

The area estimates for small woodlands and wide linear features were rated-up to regional totals by a combination of factors:

- The first factor converted sample results from the two sub-squares to estimated totals for the whole 1 km x 1 km square. For small woodlands, this factor was the total number of small woodlands in the API for the square divided by the number of small woodlands in the sample field survey data. For wide linear features, this factor was the ratio of the total length of all linear features (wide and narrow) in the API divided by the total length of all linear features (wide and narrow) in the sample field survey data.
- The second factor converted estimates for the 1% sample of 1 km x 1 km squares into estimates for the whole region. This was calculated as the total land area of the region divided by the total land area of all sample 1 km x 1 km squares in the region.

Approximate standard errors for the counts of the number of these features and for their total areas were calculated in a similar way to the Main Woodland Survey. The sample size was the total number of 1 km x 1 km sample squares in that region, implicitly including zeros for the many sample squares that contained no small woodlands or wide linear features. As with the Main Woodland Survey, this is likely to underestimate the standard errors because it ignored the complication of the two-stage sample survey, in which the first stage was the selection of 1 km squares and the second stage selection of the two parts for field survey.