



# Countryside Survey

CS Technical Report No.4/07  
**Statistical Report**

W.A. Scott

Centre for Ecology and Hydrology  
(Natural Environment Research Council)

**November 2008**

Correspondence:

**Dr. Andy Scott**

**E-mail: [was@ceh.ac.uk](mailto:was@ceh.ac.uk)**

CEH Lancaster  
Lancaster Environment Centre  
Library Avenue  
Bailrigg, Lancaster  
LA1 4AP

## **CONTENTS**

<b>EXECUTIVE SUMMARY</b>	<b>3</b>
<b>1. INTRODUCTION</b>	<b>4</b>
<b>2. OVERVIEW OF PREVIOUS CS METHODOLOGY</b>	<b>4</b>
2.1 Sampling	4
2.2 Estimation	4
2.3 Bootstrapping	5
<b>3. REASONS FOR MODIFICATIONS TO CS METHODOLOGY</b>	<b>5</b>
<b>4. CONSISTENT ESTIMATION</b>	<b>6</b>
4.1 Possible approaches	6
4.2 Modelling basics	8
4.3 Square level data	8
4.4 Plot level data	9
4.5 Model specification	10
4.6 Model testing - Broad Habitats	11
<b>5. LIMITATIONS AND IMPLICATIONS</b>	<b>12</b>
<b>6. REFERENCES</b>	<b>14</b>

## **EXECUTIVE SUMMARY**

1. This report provides a description of the statistical methodology used for the analysis of Countryside Survey data. The methods used in reporting previous surveys are summarised and a detailed description of the changes made for the analysis of the data from CS in 2007 provided..
2. Previous methods of estimation of stock and change in CS made minimal assumptions about the data and were therefore robust. However, estimates of stock were calculated using all the data from a particular survey while change was calculated from only the more limited sample of repeated measurements across pairs of surveys. This approach both failed to use all the data collected in each survey for change estimates and resulted in mis-matches between estimates, i.e. differences between stock estimates not equal to change estimates
3. An investigation using CS Broad Habitat data from previous surveys showed that consistent estimation via modelling was both feasible and reasonably robust. In general estimates derived using these methods differed from estimates obtained using the old methods by less than the inconsistencies already arising from the old methods.
4. Following the successful investigation modelling methods for consistent estimation have been adopted for the 2007 CS data.
5. The modelling approach requires additional assumptions about the distribution of data which could compromise estimates if incorrect. Care has therefore been taken to check the validity of results, especially those for small subsets of the data, by comparison with estimates produced using the previous methodology.
6. Unlike the previous methods, the new methodology does utilise all available information and hence produces more precise estimates. One consequence of this improved efficiency is that the data from each new survey, in so far as it conveys information about missing values from preceding surveys, will also produce improved estimates for previous surveys, though any such changes are likely to be small.
7. Implementation of a modelling approach has proved to be technically challenging. It requires much more computer time than previous methods since it involves the iterative fitting of a model rather than the formulaic calculation of a mean. In addition there were a number of practical issues affecting implementation that arose from the complexity of the CS sampling methodology. The successful overcoming of these difficulties has resulted in a much improved product.

## **1. INTRODUCTION.**

This technical report gives an overview of the sampling and analysis procedures used for Countryside Survey (CS) data. The previous CS statistical methodology is briefly reviewed (Section 2), before describing the reasons for, and details of, changes to estimation procedures that have been made for CS in 2007 (Sections 3 and 4). The limitations of the new methods and the implications of their introduction are discussed (Section 5). Fuller expositions of the details of previous CS sampling and estimation can be found in Barr *et al.* (1993).

## **2. OVERVIEW OF PREVIOUS CS STATISTICAL METHODOLOGY.**

### **2.1 Sampling**

CS Field Survey data comprises information collected from a stratified sample of the 1 km squares at the intersection of a 15km grid covering GB. Each selected square is mapped and detailed measurements made of selected features, for example a number of quadrats are laid out and used to collect additional information on vegetation, soils etc. Thus there are two levels of sampling. Measurements are made at both levels so that some relate to the whole square while others describe features within the square. Measurements are of varied types ranging from binary (yes/no) variables to continuous variables such as areas or lengths.

The strata used for square selection are defined by the ITE Land Classification. The details of the classification have changed somewhat from its original form, largely as a result of the need for separate country reporting. Originally the classification comprised 32 Land Classes. For CS2000, due to the requirement for separate reporting in Scotland, the classification was modified to contain 42 classes. For CS in 2007, as a result of modifications to the classification brought about by the requirement of Wales only reporting, the classification comprises 45 Land Classes. Effectively each country now has a separate classification, 21 classes in England, 8 in Wales and 16 in Scotland, although the classes in each of these national classifications are strongly related through their derivation from the original GB classification.

### **2.2 Estimation**

The basic procedure originally used in calculating regional or national estimates was to produce means and standard errors for the quantity of interest for each Land Class and then to combine these to produce an estimated mean or total (with standard error) for the specified region. The method of combination differs depending on whether a total or mean figure is required but in both cases involves weighting the individual land class estimates by values proportional to the area of land within the Land Class.

This procedure makes minimal assumptions about the form of the data. Estimates of means and standard errors are unbiased regardless of the distribution involved, as are the formulae for combining them. It is assumed that mean estimates for any Land Class are independent of estimates within other Land Classes and of estimates of total available land but this assumption is assured by the sampling scheme used.

### **2.3 Bootstrapping**

Testing for significance requires more information about the distribution of an estimate than just its standard error. Prior to CS2000 significance was assessed by assuming normality of estimates. In CS2000 because of concerns about the validity of this assumption, largely because of the skewness of some of the features being estimated, standard errors and confidence intervals for square level data were estimated using the bootstrap (Efron and Tibshirani, 1993).

Essentially bootstrapping involves treating sample data as a population from which to resample. Each resample produces a separate estimate of some quantity of interest, for example stock or change. A large number of resamples (typically 1000 or 10,000) then gives an approximation to the distribution of the required estimate, from which any statistic can be extracted. The main advantage of this method of estimation for CS is that it allows for non-normality in the data, without requiring details of the actual distribution. As such it provides more accurate measurements of significance.

## **3. REASONS FOR MODIFICATIONS TO CS METHODOLOGY**

The results of statistical analyses are usually reported in two forms. Point estimates, the expected or most likely value of a variable of interest, and interval estimates, the range of plausible values. Point estimates of stock and change reported from previous Countryside Surveys have been considered inconsistent by some users since reported changes in the extent of specific habitats between any two surveys have not been identical to the differences in the reported extent of those habitats in the two surveys. It must be emphasised that such discrepancies do not represent errors in the data or its analysis. The inconsistencies arise from the methods of estimation used in CS to overcome random sampling variation in the data (in particular missing information).

The reason for the discrepancies is illustrated in Figure 1. For each pair of surveys some sample squares (or plots) are not recorded in one or other of the surveys. The cause of the majority of this missing information has been the introduction of new squares as CS has developed, so that most of the unrepeated data is from the later survey in each pair, but loss of squares/plots recorded in an earlier survey, through landowner refusal for example, can also occur. Figure 1 illustrates three potential methods of estimating stock and change (others are possible):- from all squares, from repeated squares only, and from un-repeated squares only. Each method produces a consistent set of estimates. The inconsistencies in previous point estimates reported by CS arise because estimates of stock are calculated using all the data from a particular survey while change is calculated from repeated measurements only. This automatically means that only in exceptional circumstances will these estimates be consistent.

These particular estimates have been used in the past because estimating stock using all the data from a survey maximises information use from that survey, while estimating change using only repeated measurements minimises the distributional assumptions needed and hence ensures robust estimation. Difficulties arise only from the interpretation of point estimates outside the context of their standard errors and confidence intervals. If results were presented only as confidence intervals, (e.g. stock in

1990 was between  $a$  and  $b$ , stock in 1998 was between  $c$  and  $d$ , change was between  $x$  and  $y$ ), it would be clear that any inconsistency was more apparent than real.

For CS in 2007 the reporting emphasis has changed, from describing the current survey and changes since the immediately preceding survey, to timelines spanning the interval from the first survey to the present. This change of emphasis highlights the apparent inconsistencies and could also introduce additional ones since, using the same methods of analysis, estimated changes between adjacent surveys would not sum to estimates of change between non-adjacent surveys. To overcome this problem the feasibility of producing consistent estimates was examined and in consequence new methods of estimation have been introduced for the reporting of results from the 2007 CS.

## 4. CONSISTENT ESTIMATION

### 4.1 Possible approaches

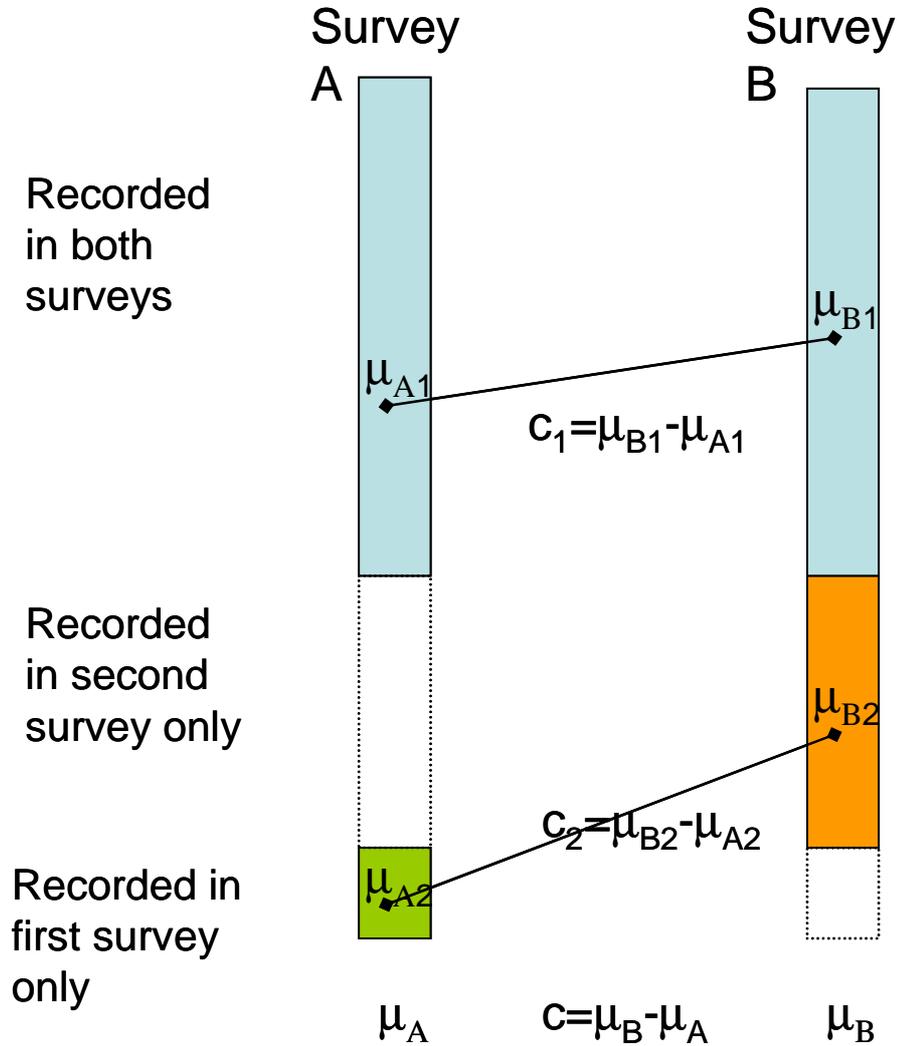
A number of approaches could be used to ensure consistency. Figure 1 illustrates three sets of consistent estimates that can be derived from a pair of surveys. Each of these three sets estimate the same values and could, in principle at least, form the basis for reporting. If almost all squares were measured on every sampling occasion then the unrepeated squares could be discarded with little loss of precision. This is not a practical approach for CS as a substantial number of extra squares has been added with each survey. Equally using just the unrepeated squares is clearly not sensible for CS since most squares are re-surveyed.

The first potentially usable approach is to use the stock estimates for all measurements, as is done at present, but to estimate change as the difference between stock estimates. The statistics relevant to this approach were described in the CS1990 main report (Barr *et al.* 1993, p171) but the method has not so far been used as a general approach in CS. It has a number of advantages. Estimates are consistent and robust and estimates from one survey do not change following the implementation of later surveys. The methodology is easy to implement and quick to run, and standard errors and confidence limits can be estimated with the bootstrap. There are however a number of disadvantages as well. Most importantly, the method is only efficient for measuring change when the covariance (or correlation) between measurements in successive surveys is not large. For many CS measurements estimating change as the difference between stock estimates would produce change values with larger standard errors than by estimating change just from repeat squares. Furthermore the approach is not directly applicable to plot level data. Thus there would be inconsistencies in methodology between estimates of square level and plot level data.

An alternative is to use modelling techniques to estimate stock and change. This approach can be applied to both square and plot level data and produces consistent as well as efficient, i.e. more precise, estimates of both stock and change. As with the previous approach, however, there are a number of drawbacks. To fully explain these the modelling approach is described in detail in the remainder of this section.

**Figure 1. Previous reporting of stock and change**

a) Survey structure



b) Estimates: (stock  $\mu$ , change  $c$ )

- $\mu_A, \mu_B, C$  all squares,
- $\mu_{A1}, \mu_{B1}, C_1$  repeat squares only
- $\mu_{A2}, \mu_{B2}, C_2$  unrepeated squares only

Reported stock  $\mu_A$  and  $\mu_B$ , reported change  $c$

## 4.2 Modelling basics

The discrepancies between estimates of stock and change arise, as explained above (Section 3, Figure 1), because missing information means that stock and change are estimated from different sets of data. Effective statistical methods for dealing with incomplete data were only devised in the late 1970's (Dempster *et al.*, 1977) and it was some time before their regular use spread from the statistical to the user community. At first a computational slow and demanding set of techniques, their practical utility and computational efficiency has been gradually increased by the introduction and theoretical justification of more effective estimators and algorithms over the last two decades (see e.g. Scott, 2002). Many statistical models fitted by proprietary software can now cope with incomplete data. Such techniques do, however, require assumptions about data distributions for implementation. Thus ensuring consistency of CS estimates of stock and change involves making additional assumptions about the data which may in some instances not be met.

The way in which incomplete data techniques work can be illustrated by considering ways in which missing information might be replaced. Change and stock estimated from the completed dataset would then be automatically consistent. Two extremes are possible, depending upon whether stock or change values are used to replace the missing data. If missing values are replaced by the appropriate survey mean then stock estimates are unchanged but a new value for change is found. Alternatively if the average change found from repeated measurements is used to predict the missing values then the change estimate from the completed dataset is the same as the change from the repeated measurements but the stock estimates will change. In reality, of course, these are extremes and a procedure somewhere between will be most appropriate. Missing information techniques in effect use the correlation structure from the repeated measurements to judge where between these two extremes the most appropriate estimates lie. In practise the techniques work directly with the observed data and not by filling in missing values.

For CS, because of its hierarchical sampling scheme, implementation of consistent estimation via modelling requires fitting a mixed effects and/or repeated measures statistical model to data from all surveys. Such models contain two types of parameter: fixed effects parameters are functions of stock and change values while random effects parameters have a specification that reflects the random variation in the data as determined by the sampling structure. After model fitting the estimates of the fixed effect parameters are then transformed to estimates of stock and change.

Such models require more assumptions than the current methods, which, following the introduction of the bootstrap, essentially only require calculation of means. In essence they require calculation of variances and covariances as well as means and specification of the distributional form of the random and repeated effects. Models appropriate to measurements made, or summarised, at the square and plot level are described below.

## 4.3 Square level data

For measurements applicable to complete 1 km squares the CS dataset can be considered as made up of a random sample of squares within each Land Class, each square providing a value on each survey (apart from missing observations). Statistically

the appropriate model for this form of data is a repeated measures model. Such a model comprises two separate model components, one for fixed effects and one for random elements (hence the generic name mixed model).

The fixed effect component is just a standard regression model. For CS square level data the simplest fixed effect model treats the mean value within a Land Class on each sampling occasion as a fixed effect to be estimated, i.e. a simple regression of the variable of interest on year (or, equivalently, survey) treated as a categorical variable. The fitted effects are then, when scaled by the land class area, just the required estimates of stock in each survey. Estimates of change are just the differences between fitted stock estimates and hence are automatically consistent. More complex models, with additional explanatory variables can be used to break down the stock, or change, estimates into additional categories.

The random effects component of the overall model describes the variation of individual recorded measurements about the fitted fixed effects. Standard regression models specify one random element per observation, usually referred to as a residual, and all residuals are assumed to be independent. A mixed model differs from a standard regression model in including parameters describing the structure of the residuals. For CS square level data each square within a land class is assumed to have a constant random difference from the land class average. Measurements from the same square in successive surveys vary about this square level residual, and these survey deviations from the square level residual are allowed to be correlated.

#### **4.4 Plot level data**

CS measurements are made not just at the whole square level but also within squares. Vegetation and soil data, for example, are recorded for a number of plots within each sample square. In previous surveys the full hierarchical nature of the plot level data was not explicitly dealt with. A variety of approaches were adopted for different analyses. In some, measurements were summarised at the square level prior to analysis. This approach is robust but clearly does not make full use of the data and hence will generate standard errors that are larger than necessary. In other analyses plots were treated as independent observations within a land class and the square level variation not allowed for. This approach is efficient if the variation among plots within squares is the same as their variation across squares but can produce biased results, or incorrect standard errors, if this is not true. In CS2000 mixed models that allowed for square level variation but ignored the sampling structure in terms of land classes were used for some plot level data. In addition, because the bootstrapping macros written for CS2000 were produced for square level data only, results for plot level data had standard errors calculated from, possibly incorrect, distributional assumptions rather than from bootstrapping.

The model described above for square level data can be extended by the inclusion of a plot level residual, or random effect, in addition to the square level random effect. The correlated survey residuals now vary about the average level for the plot, not the square. Both forms of model can be embedded within bootstrapping procedures.

#### 4.5 Model specification

Exposition of the proposed models and the assumptions on which they are based requires at least some mathematical specification for clarity. Let  $y_{ijk}$  represent an observation in survey  $k$  from square  $j$  in land class  $i$ . Then a general model for square level data can be written as

$$y_{ijk} = a_{ik} + s_{ij} + e_{ijk}$$

where the  $a$  parameters (the fixed effects) represent land class means in successive surveys, the  $s$  values are the square random effects and the  $e$  values are the repeated measures effects. To complete the model requires specification of the distribution of the random and repeated effects. The  $s$  values are assumed to be normally distributed with mean zero and standard deviations  $\tau_i$  which differ across land classes. The  $e$  values are also normally distributed with zero mean, standard deviations  $\sigma_{ik}$  which vary across land classes and surveys, and covariances, for individual squares, which vary across land classes and pairs of surveys.

For  $K$  successive surveys this general model includes, for each land class,  $K$  fixed effect parameters but  $1+K(K+1)/2$  random parameters (the variances and covariances of the random and repeated values). Thus the number of random effect parameters is greater than the number of fixed effect parameters and this imbalance increases with the number of surveys. Unfortunately estimates of variances and covariances are much less precise than estimates of means, which the fixed effect parameters effectively are. Because of the large number of land classes used for CS sampling there are relatively few sample squares in each class. The result is that the full model tends to be unstable and difficult to fit, increasingly so as the number of surveys increases. An additional technical complication is that the computer time for model fitting also increases with the number of parameters.

To make consistent estimation via modelling practicable, therefore, it is desirable to reduce the number of random effect parameters. A helpful property of mixed effect models is that estimation of the fixed effects parameters is relatively robust to mis-specification of the distribution of the random values. Thus the number of random effect parameters can often be reduced considerably without substantially affecting the accuracy or precision of the fixed effect parameters. Reducing the number of parameters can be done in a variety of ways, giving a choice of models to fit. It is not usually sensible to set random parameters to zero, the usual method of reduction for regression or fixed effect parameters. The alternative is to assume certain sets of parameters are equal or can be specified as functions of a smaller number of parameters.

One possibility is to assume that variance and/or covariance parameters do not vary with land class. However for many CS variables this is demonstrably not true, variability is very different across land classes. A more realistic assumption is that random effect parameters do not vary across surveys. Thus it can be assumed that the standard deviations,  $\sigma_{ik}$ , take a common value,  $\sigma_i$ , for all surveys. This assumption reduces the number of repeated measures variance parameters per land class to one. Many theoretical structural models have been proposed for covariances. A particularly effective model is the autoregressive model of order one which assumes that the covariance between repeated measure values in successive surveys is constant and that non-adjacent survey values are conditionally independent given the values of

intervening surveys. This assumption reduces the number of repeated measures covariance parameters per land class to one. Using both of these assumptions (giving a model that will be referred to as the AR1 model) reduces the total number of random effect parameters to three per survey, regardless of the number of surveys.

Although estimation of fixed effects is relatively robust to mis-specification of distributional assumptions this is not the case for variance and covariance parameters. Thus parametric calculation of standard errors may produce erroneous values and this applies to the standard errors automatically output by the modelling software. However bootstrap estimation, which requires only the fixed effect values, will also be robust. The AR1 model with bootstrap estimation of standard errors was therefore investigated in detail as a means of providing consistent estimation with CS data.

#### **4.6 Model testing - Broad Habitats**

One of the main outputs of previous CS surveys (Barr *et al.*, 1993; Haines-Young *et al.*, 2000) has been an assessment of the stock of, and change in, acreage of a variety of habitats. In CS2000 standard Broad Habitats were used. Broad Habitat information is recorded at the square level as the proportion of rural land within the square that falls into each category. Information on Broad Habitats is available from the 1984, 1990 and 1998 (CS2000) surveys. Habitat information from the 1978 survey, long before the definition of Broad Habitats, was coded differently so is not directly comparable. To investigate the application of modelling methods prior to their adoption, data from the 1984, 1990 and 1998 surveys for seven Broad Habitats were analysed.

Using the old methodology two forms of change estimate were calculated for each pairs of surveys, change estimates from repeated squares and the differences between stock estimates. The inconsistencies between stock and change from repeated squares, evident from previous reports, were clear. Differences between the stock estimates for any pair of surveys did not equal the corresponding change estimates. Additional discrepancies, not obvious from previous surveys because of the reporting structure used, could also be seen. Using only repeated squares, estimates of changes from 1984 to 1990 and from 1990 to 1998 did not sum to the estimates of change from 1984 to 1998. Ratios of the standard errors for the two methods of calculating change were also calculated. Almost all of these ratios were greater than one and many substantially higher, emphasising the fact that estimating change from stock values gives less precise estimates in general than estimating change from repeat squares. This is the reason that CS has used the methods that it has to date.

Estimates of stock and change, with their standard errors, were also obtained from fitting mixed effect/repeated measures models to the data from the three surveys. Separate models were fitted for each land class. The form of model used (denoted AR1) assumed constant within land class variance of each variable across surveys with correlation between surveys represented as a first order autocorrelation process. These estimates, by definition, did not exhibit the discrepancies of the old methodology. Each change estimate was equal to the difference between the corresponding stock estimates and change estimates from consecutive inter-survey periods summed to the change estimate for the change over the whole period.

The effect of moving from the current methodology to a modelling approach was summarised by the difference in estimates from the old and new methods as a percentage of the standard error of the estimate using the old CS method of analysis. For stock only one estimate changed by as much as a standard error (neutral grassland in England & Wales in 1990) and most changes were much less than this. For change estimates also only one estimate has altered by more than a standard error (the change in Coniferous woodland in England & Wales from 1984 to 1990). Overall none of the estimates obtained using the new modelling approach were outside the error bounds of the estimates from the old methodology and most were well within them.

This particular form of comparison of old and proposed methods emphasises the lack of significance of the reported differences. However it can appear to exaggerate the actual alterations that occur. The estimated stock of Improved Grassland in GB in 1984, for example, changed by a fifth of a standard error when modelling was used in place of the older methods but the actual change in the estimate was less than one percent. The change was only substantial in terms of the standard error because the extent of this broad habitat is large and is well estimated with a relatively small standard error

As a further check, the differences between old and new estimates of change were compared to the discrepancies under the old methodology (i.e. the differences between changes derived from two stock estimates and those estimated directly from repeated squares). In general the estimates derived using consistent estimation methods differed from estimates obtained using the old methods by less than the inconsistencies already arising from the old methods.

In addition to the Broad Habitat data consistent analyses were undertaken for CS soil data collected in 1978 and 1998, a plot level dataset. The results confirmed the feasibility of producing consistent estimates at this level as well as at square level. This not only makes such estimates numerically consistent but would also produce a consistency of approach across plot and square level data, something not achieved in previous surveys. It was therefore decided to adopt the new methodology for the analysis of data from the 2007 CS.

## **5. LIMITATIONS AND IMPLICATIONS**

Implementation of model based analysis within a bootstrapping envelope for square level data, although computationally challenging at times, proved to be reasonably straightforward. Initial experimentation with a variety of models confirmed that estimates of fixed effect parameters, e.g. stock and change, were robust to model variation. Fully parameterised models were extremely slow to fit, to the extent that use of this model is impractical for the analysis of large numbers of variables. However, the AR1 model, although taking substantially longer to run than the old methodology, was not sufficiently slow as to suggest that extension of the technique to the large number of analyses required for the complete survey was impractical.

Choice of a suitable model is clearly an important part of ensuring estimates are accurate. The AR1 model has many desirable properties; it is stable, relatively quick to

fit, has a small number of random parameters that does not increase with the number of surveys, and appears to give estimates of fixed effects that are robust to distributional mis-specification. When producing large numbers of analyses, as for Countryside Survey, it is clearly not possible to spend substantial amounts of time on model selection and checking. The need is for a standard model that can be applied in an automated manner to a large number of variables to produce robust results. The AR1 model appears to meet these criteria. When adopting it for CS2007, however, it was thought prudent to implement checks on performance and accuracy. The analysis programs for CS have therefore been written to produce estimates using both the new and old methods. Differences are checked to be small and comparable to the discrepancies between stock and change arising from the old methodology.

In addition to model structure, defined by the chosen parameter set, the distributional assumptions of the model will affect estimation. For the models described here, the effect of treating the distributions of random effects as normal when they are not does not appear to markedly affect fixed effect estimation. However for more complex models or for very non-normal distributions this may not be true.

For very non-normal data a standard approach to non-normality is to transform data prior to analysis. For CS, however, it is important to present results on the original scale of measurement. Analysis could be performed on some transformed scale but it would then be necessary to convert fitted parameter values to measures on the original scale of measurement. Such conversions almost always involve random as well as fixed effects and so are susceptible to the less precise estimation of these parameters.

During the analysis of CS freshwater data it became clear that there were major differences between the estimates for the extent of standing water using the old and new methodology. On investigation this was judged to be due to the very non-normal distribution of this variable causing the new methodology to converge to a local maximum of the likelihood function. Attempts to remedy this situation were unsuccessful so the estimates presented in the CS report are those given by the old methodology.

There are other reasons than just consistency for adopting a consistent methodological approach to estimation and analysis. Although robust, previous methods of analysis were not always fully efficient in that they did not utilise all the available information in producing individual estimates and did not always incorporate the hierarchical structure of the data. The modelling approach does utilise all available information as well as correctly representing the data hierarchy and hence, assuming of course that the distributional approximations are sufficiently reasonable not to bias the analyses, should produce more precise estimates.

Adoption of this approach has other implications for results. Because analyses involve data from all surveys then estimates for any one survey are influenced by information from all others. A consequence of this is that estimates can not be made consistent across reporting occasions since the introduction of additional data with each new survey will produce updated estimates for previous surveys. Such updating is conceptually different to the inconsistencies currently present in the reporting from previous surveys. The latter arise from not fully utilising available information. In

contrast it does not seem unreasonable for the acquisition of new information to be expected to produce small revisions to previous findings.

## 6. REFERENCES

Barr, C.J., Bunce, R.G.H., Clarke, R.T. , Fuller, R.M., Furse, M.T., Gillespie, M.K., Groom, G.B., Hallam, C.J., Hornung, M., Howard, D.C., and Ness, M.J. (1993). *Countryside Survey 1990 Main Report*, DETR, London.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977).  
Maximum likelihood from incomplete data via the E.M. Algorithm.  
J.R. Statist. Soc. B. 39: 1-38.

Efron, B. and Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman and Hall; London.

Haines-Young, R.H., C.J. Barr, H.I.J. Black, D.J. Briggs, R.G.H. Bunce, R.T. Clarke, A. Cooper, F.H. Dawson, L.G. Firbank, R.M. Fuller, M.T. Furse, M.K. Gillespie, R. Hill, M. Hornung, D.C. Howard, T. McCann, M.D. Morecroft, S. Petit, A.R.J. Sier, S.M. Smart, G.M. Smith, A.P. Stott, R.C. Stuart, R.C. & J.W. Watkins. (2000). *Accounting for nature: assessing habitats in the UK countryside*, DETR, London.

Scott, W.A. (2002).  
Maximum likelihood estimation using the empirical Fisher information matrix.  
J. Statistical Computation and Simulation. 72(8): 599-611.



This report is the Copyright of the Natural Environment Research Council. Copyright enquiries should be addressed to: Knowledge Transfer Team, Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford OX10 8BB

This publication, excluding logos, may be reproduced free of charge in any format or medium for research, private study or the purposes of internal use within an organisation. This is subject to it being reproduced accurately and not being subject to any treatment that could be considered derogatory. The reproduced material must be acknowledged as NERC Copyright (except where otherwise stated) and the publication named in full.

#### **Disclaimer**

Any decisions or actions informed by this Technical Report are taken entirely at your own risk. In no event shall NERC be liable for any damages, including loss of business, loss of opportunity, loss of data, loss of profits or for any other indirect or consequential loss or damage whatsoever arising out of the use of or inability to use the material presented in this report.

For further information on Countryside Survey see [www.countrysidesurvey.org.uk](http://www.countrysidesurvey.org.uk)

*Or contact:*

Countryside Survey Project Office,  
Centre for Ecology and Hydrology,  
Lancaster Environment Centre,  
Library Avenue, Bailrigg, Lancaster LA1 4AP

Telephone: 01524 595811; Email: [countrysidesurvey@ceh.ac.uk](mailto:countrysidesurvey@ceh.ac.uk)

*Countryside Survey in 2007 was funded by a partnership of government-funded bodies led by the Natural Environment Research Council (NERC) and the Department for Environment, Food and Rural Affairs (Defra).*

