QUALITY ASSURANCE REPORT - COUNTRYSIDE SURVEY:
QA and bias in vegetation recording

W.A. Scott, S.M. Smart and R. Clarke

Correspondence:
Dr. Andy Scott
E-mail: was@ceh.ac.uk

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

16 April 2008

CONTENTS

EXECUTIVE SUMMARY

1. Analyses of the QA data from Countryside Survey 2007 have suggested a possible increase in the level of under-recording of vegetation species compared to previous surveys.

2. A degree of under recording is to be expected due to the greater expertise of the QA team. Changes in the level of under-recording would, however, bias estimates of change.

3. Recording of cryptogams has always been known to be difficult and they have been excluded from analyses of previous surveys for this reason. In CS2007, recording of cryptogams appears to be substantially worse than in previous surveys.

4. Once cryptogams are removed, statistical modelling shows no significant differences in the level of under-recording of species across surveys.

5. There is, however, evidence that measures derived from the vegetation data, such as Ellenberg scores, had a different level of bias in the 1990 survey to the level in subsequent surveys.

6. It is recommended that cryptogams be excluded from analysis of CS2007 data, as with previous surveys, but that no corrections for under-recording of other species be performed.

7. Suggestions for the correction of CS1990 values of derived measures for the difference in bias to other surveys are made and the values needed for this derived.

1. BACKGROUND

Quality assurance (QA) is an important element of any major monitoring programme, allowing the validity of findings to be assessed. In Countryside Survey (CS) QA has been an aspect of the vegetation recording in each survey since 1990. A team of independent experts, the same for each survey, has been contracted to carry out repeat recording of a subset of the vegetation plots and to report on the findings. The level of agreement between the QA team and CS field teams is then a measure of botanical accuracy. In both the 1990 and 1998 surveys the QA assessors found more species than the CS field teams. This is not unexpected, given the nature of the exercise and the expertise and remit of the QA contractors, and should not be seen as a cause for concern. The scale of differences between vegetation recorders has been a subject of debate in the botanical literature for many years (see e.g. Scott & Hallam 2003) and it would be more surprising if the "experts" did not perform better that the average field surveyor. The QA reports for surveys prior to 2007 showed that the observed differences in number of species did not translate into significant differences in derived measures of species composition, such as the variables created by detrended correspondence analysis (DCA).

1.1 QA 2007

The 2007 QA report (Prosser & Wallace, 2008a) concentrated on the analysis of the full 2007 QA dataset (i.e. the QA data plus the field survey records for QA plots) but included some comparisons with previous reports. It appeared to show a decline in the quality of botanical recording compared to previous surveys, in particular a greater disparity between the number of species per plot recorded by the QA and the field survey teams. However, the report highlighted the much reduced accuracy of recording of allowable cryptogams and this will have contributed substantially to the observed differences in species richness. Previous reported CS analyses of the full dataset (not just the QA component) have excluded cryptogams because of known difficulties in identification.

Additional analyses were requested from the QA team as a consequence of the findings of the initial report. The resulting supplementary report (Prosser & Wallace, 2008b) contained two sets of analysis: 1) detailed analyses of the full 2007 QA dataset (266 plots) but with cryptogams excluded; 2) time sequence analysis of a much reduced QA dataset (referred to as the matched triplicate dataset) containing information only for those plots that were recorded in all three survey years, 1990, 1998 and 2007. This dataset contained 108 plots and was analysed both with and without cryptogams.

These supplementary analyses of the full 2007 QA dataset showed that species richness as recorded by the field surveyors was still significantly lower than that recorded by the QA team even after excluding cryptogams but that the difference was much reduced by the exclusion. In contrast, analyses of the matched triplicate dataset showed no significant difference between QA and field survey for any of the three survey years once cryptogams were excluded. However, it should be noted that the substantially reduced size of this dataset greatly reduces its power to detect such differences.

1.2 Purpose and structure of this report

The supplementary analyses contained no direct comparison of the differences in species richness between QA and field survey in the three surveys (the QA team were not requested to do this). Section 2 of this report describes detailed analyses that have been performed by the CS analysis team to rectify this. The dataset used was the full QA dataset for all three surveys, both repeated and non-repeated plots, but with cryptogams excluded. Section 3 discusses the results of these additional analyses and makes recommendations with regard to the analysis of the full CS2007 dataset.

2. ANALYSIS.

This section summarises analyses of the bias in CS vegetation measures across the three survey years for which QA results are available. Measures are divided into two groups for separate consideration, 1) species richness measures, i.e. counts of particular groups of species, which are directly dependent on the surveyors recording of vegetation, and 2) derived measures, Ellenberg scores etc. which are functions of the individual species properties and not of the number of species.

2.1 Data and statistical methods

Although the set of QA plots varied somewhat from survey to survey there were very few instances where a QA plot was not, in a particular year recorded by both the QA and field survey botanists (9 plots out of a total of 687). To simplify analysis therefore the bias of each plot in each survey year was calculated directly as the difference between the QA and CS values and the analyses reported here were calculated directly with these differences.

A mixed model with square as a random effect, individual plots as the basic units, and survey year defining the repeated measurement of individual plots was used to estimate the amount of bias in each year and changes in the level of bias from survey to survey. In addition, a model specifying a common level of bias in all three survey years was used to estimate the average bias over all three surveys and a test of whether there were significant differences in bias was performed. Use of this model utilises all the data, thus increasing the power of the analyses, and allows for possible lack of statistical independence in plots within the same survey squares

2.2 Species richness measures

Table 1 gives, for all species richness measures, the estimated bias in each of the survey years and the estimated difference in bias for each pair of surveys. Levels of bias are illustrated in Figure 1. Considering firstly the overall level of bias, there is a clear and significant bias for all species richness measures except richness of non-native species. The size of the bias, i.e. the average number of species difference between the CS and QA results varies considerably, a reflection of the probability of overlooking combined with the average number of species in each group. It is clear therefore that the CS surveyors record fewer species than the QA surveyors and that this under-recording is not confined to a small number of groups of species.

Testing for differences in the size of this bias shows, however, that there is little evidence that the extent of under-recording has changed with time. Only for grass richness was there a significant difference between survey years (p=0.045) and the level of significance is such that any correction for multiple testing within this group of measures would make this result non significant as well.

2.3 Derived measures

Table 2 gives, for all derived measures, the estimated bias in each of the survey years and the estimated difference in bias for each pair of surveys. As with the species richness measures, there is a clear and significant bias for most measures. Only Ellenberg light and R radius scores do not show significant bias. It is clear, therefore, that the difference in recording of species translates into a difference for derived measures as well.

Testing for differences in bias size gives a different picture to that found for species richness measures. Although only two measures show a significant difference between years (Ellenberg fertility and pH) the differences in Ellenberg wetness and light scores are also almost significant. There is, therefore, some evidence that the level of bias has changed with time for Ellenberg scores. Examination of the estimates of bias in individual surveys (Table 2 and Figure 2) shows that in no case is the bias from CS2000 different from the bias in CS2007. The bias for CS1990, however frequently differs from the bias in the other two surveys.

To examine this result more closely a model was fitted to the data which specified the same level of bias for 1998 and 2000 but allowed the bias for 1990 to differ. The two estimated levels of bias and the difference between them is given in Table 3 for all measures not just Ellenberg indices. None of the species richness measures show a significant difference between 1990 and the other two years, but six of the seven derived measures do. Only C radius is not significant.

Table 1. Estimates of bias and bias differences in species richness measures for QA vegetation plots.

Species Group	Estimate	Value	StdErr	P	Lower 95% CI	Upper 95% CI	Prob years differ in bias
All	90bias	2.27	0.37	0.000	1.53	3.01	
	98bias	1.88	0.37	0.000	1.14	2.61	
	07bias	2.81	0.34	0.000	2.14	3.49	
	9098biasdiff	-0.40	0.43	0.361	-1.24	0.45	
	9007biasdiff	0.54	0.43	0.206	-0.30	1.38	
	9807biasdiff	0.94	0.41	0.021	0.14	1.74	
	Average bias	2.38	0.27	0.000	1.85	2.91	0.068
sedges	90bias	0.02	0.05	0.697	-0.07	0.11	
	98bias	0.10	0.05	0.027	0.01	0.19	
	07bias	0.12	0.04	0.004	0.04	0.21	
	9098biasdiff	0.08	0.05	0.116	-0.02	0.19	
	9007biasdiff	0.10	0.06	0.063	-0.01	0.21	
	9807biasdiff	0.02	0.05	0.695	-0.08	0.12	
	Average bias	0.08	0.03	0.013	0.02	0.15	0.147
Non-native	90bias	0.06	0.04	0.091	-0.01	0.14	
	98bias	-0.01	0.04	0.848	-0.08	0.07	
	07bias	0.04	0.03	0.293	-0.03	0.10	
	9098biasdiff	-0.07	0.05	0.181	-0.18	0.03	
	9007biasdiff	-0.03	0.05	0.554	-0.12	0.07	
	9807biasdiff	0.04	0.05	0.394	-0.05	0.14	
	Average bias	0.03	0.02	0.161	-0.01	0.07	0.405
Native	90bias	2.06	0.36	0.000	1.35	2.76	
	98bias	1.91	0.35	0.000	1.22	2.61	
	07bias	2.71	0.32	0.000	2.07	3.35	
	9098biasdiff	-0.14	0.41	0.731	-0.95	0.67	
	9007biasdiff	0.65	0.41	0.108	-0.14	1.45	
	9807biasdiff	0.80	0.39	0.039	0.04	1.55	
	Average bias	2.28	0.25	0.000	1.78	2.79	0.092
grasses	90bias	0.53	0.13	0.000	0.28	0.79	
	98bias	0.45	0.13	0.001	0.20	0.70	
	07bias	0.82	0.12	0.000	0.59	1.05	
	9098biasdiff	-0.09	0.16	0.601	-0.41	0.23	
	9007biasdiff	0.28	0.16	0.081	-0.03	0.60	
	9807biasdiff	0.37	0.15	0.017	0.07	0.67	
	Average bias	0.62	0.08	0.000	0.46	0.79	0.045
forbs	90bias	1.44	0.27	0.000	0.91	1.98	
	98bias	0.94	0.27	0.001	0.41	1.47	
	07bias	1.45	0.24	0.000	0.97	1.93	
	9098biasdiff	-0.50	0.33	0.135	-1.16	0.16	
	9007biasdiff	0.00	0.33	0.992	-0.63	0.64	
	9807biasdiff	0.50	0.31	0.108	-0.11	1.12	
	Average bias	1.30	0.18	0.000	0.94	1.66	0.200

Table 2. Estimates of bias and bias differences in derived measures for QA vegetation plots.

Variable	Estimate	Value	StdErr	P	Lower 95' CI	Upper 95' CI	Prob years differ in bi
Ellenberg fertility	90bias	-0.002	0.020	0.934	-0.041	0.037	
	98bias	-0.076	0.020	0.000	-0.115	-0.037	
	07bias	-0.057	0.018	0.001	-0.092	-0.023	
	9098biasdiff	-0.074	0.027	0.007	-0.128	-0.020	
	9007biasdiff	-0.055	0.027	0.037	-0.108	-0.003	
	9807biasdiff	0.019	0.026	0.470	-0.032	0.070	
	Average bias	-0.046	0.011	0.000	-0.068	-0.024	0.020
Ellenberg ph	90bias	0.006	0.017	0.717	-0.028	0.040	
	98bias	-0.048	0.017	0.005	-0.082	-0.015	
	07bias	-0.038	0.015	0.012	-0.068	-0.008	
	9098biasdiff	-0.054	0.023	0.020	-0.100	-0.008	
	9007biasdiff	-0.045	0.023	0.053	-0.090	0.001	
	9807biasdiff	0.010	0.022	0.658	-0.034	0.053	
	Average bias	-0.028	0.010	0.006	-0.047	-0.008	0.049
Ellenberg wetness	90bias	-0.016	0.019	0.399	-0.052	0.021	
	98bias	0.040	0.018	0.033	0.003	0.076	
	07bias	0.038	0.016	0.022	0.006	0.070	
	9098biasdiff	0.055	0.027	0.039	0.003	0.108	
	9007biasdiff	0.054	0.025	0.031	0.005	0.102	
	9807biasdiff	-0.002	0.025	0.947	-0.051	0.047	
	Average bias	0.023	0.010	0.036	0.002	0.044	0.054
Ellenberg light	90bias	-0.023	0.014	0.117	-0.051	0.006	
	98bias	-0.003	0.014	0.825	-0.031	0.025	
	07bias	0.022	0.013	0.082	-0.003	0.047	
	9098biasdiff	0.019	0.020	0.329	-0.020	0.059	
	9007biasdiff	0.045	0.019	0.017	0.008	0.082	
	9807biasdiff	0.025	0.019	0.174	-0.011	0.062	
	Average bias	0.001	0.008	0.916	-0.016	0.017	0.054
C radius	90bias	-0.048	0.018	0.007	-0.083	-0.013	
	98bias	-0.021	0.018	0.228	-0.056	0.013	
	07bias	-0.026	0.016	0.092	-0.057	0.004	
	9098biasdiff	0.027	0.024	0.271	-0.021	0.075	
	9007biasdiff	0.022	0.024	0.359	-0.025	0.068	
	9807biasdiff	-0.005	0.023	0.822	-0.051	0.040	
	Average bias	-0.032	0.010	0.002	-0.051	-0.012	0.506
R radius	90bias	0.052	0.020	0.009	0.013	0.091	
	98bias	-0.007	0.020	0.738	-0.045	0.032	
	07bias	0.016	0.018	0.366	-0.019	0.050	
	9098biasdiff	-0.058	0.027	0.034	-0.112	-0.004	
	9007biasdiff	-0.036	0.026	0.176	-0.088	0.016	
	9807biasdiff	0.022	0.026	0.387	-0.028	0.073	
	Average bias	0.020	0.011	0.073	-0.002	0.042	0.102
S radius	90bias	0.003	0.015	0.827	-0.025	0.032	
	98bias	0.035	0.014	0.016	0.007	0.063	
	07bias	0.040	0.013	0.002	0.014	0.065	
	9098biasdiff	0.032	0.021	0.128	-0.009	0.073	
	9007biasdiff	0.036	0.019	0.061	-0.002	0.075	
	9807biasdiff	0.005	0.020	0.809	-0.034	0.043	
	Average bias	0.027	0.008	0.001	0.012	0.043	0.142

Table 3. Estimates of bias and bias differences in all measures for QA vegetation plots assuming equal bias in 1998 and 2007.

Species Group	Estimate	Value	StdErr	P	Lower 95%	Upper 95%
All	90bias	2.325	0.374	0.000	1.586	3.064
	98&07bias	2.407	0.291	0.000	1.826	2.988
	difference in bias	0.082	0.381	0.830	-0.667	0.831
sedges	90bias	0.020	0.046	0.659	-0.071	0.112
	98&07bias	0.114	0.036	0.002	0.043	0.185
	difference in bias	0.093	0.048	0.055	-0.002	0.188
Non-native	90bias	0.062	0.038	0.099	-0.012	0.136
	98&07bias	0.016	0.025	0.516	-0.033	0.066
	difference in bias	-0.046	0.044	0.299	-0.133	0.041
Native	90bias	2.100	0.355	0.000	1.399	2.801
	98&07bias	2.364	0.277	0.000	1.813	2.916
	difference in bias	0.264	0.362	0.466	-0.446	0.975
grasses	90bias	0.554	0.130	0.000	0.298	0.810
	98&07bias	0.656	0.095	0.000	0.467	0.846
	difference in bias	0.102	0.143	0.475	-0.179	0.383
forbs	90bias	1.460	0.270	0.000	0.928	1.991
	98&07bias	1.228	0.200	0.000	0.829	1.626
	difference in bias	-0.232	0.291	0.425	-0.803	0.339
Ellenberg fertility	90bias	-0.001	0.020	0.951	-0.040	0.038
	98&07bias	-0.065	0.013	0.000	-0.092	-0.039
	difference in bias	-0.064	0.024	0.007	-0.111	-0.018
Ellenberg ph	90bias	0.007	0.017	0.698	-0.027	0.041
	98&07bias	-0.043	0.012	0.000	-0.066	-0.020
	difference in bias	-0.049	0.020	0.016	-0.089	-0.009
Ellenberg wetness	90bias	-0.016	0.019	0.399	-0.052	0.021
	98&07bias	0.039	0.012	0.002	0.015	0.063
	difference in bias	0.054	0.022	0.015	0.010	0.098
Ellenberg light	90bias	-0.023	0.014	0.114	-0.051	0.005
	98&07bias	0.011	0.010	0.253	-0.008	0.030
	difference in bias	0.034	0.017	0.046	0.001	0.067
C radius	90bias	-0.048	0.018	0.007	-0.083	-0.014
	98&07bias	-0.024	0.012	0.043	-0.048	-0.001
	difference in bias	0.024	0.021	0.251	-0.017	0.066
R radius	90bias	0.052	0.020	0.009	0.013	0.091
	98&07bias	0.006	0.013	0.653	-0.020	0.032
	difference in bias	-0.046	0.024	0.050	-0.092	0.000
S radius	90bias	0.003	0.015	0.831	-0.025	0.032
	98&07bias	0.038	0.009	0.000	0.019	0.056
	difference in bias	0.034	0.018	0.050	0.000	0.069

Figure 1. Estimated bias in species counts of vegetation plots

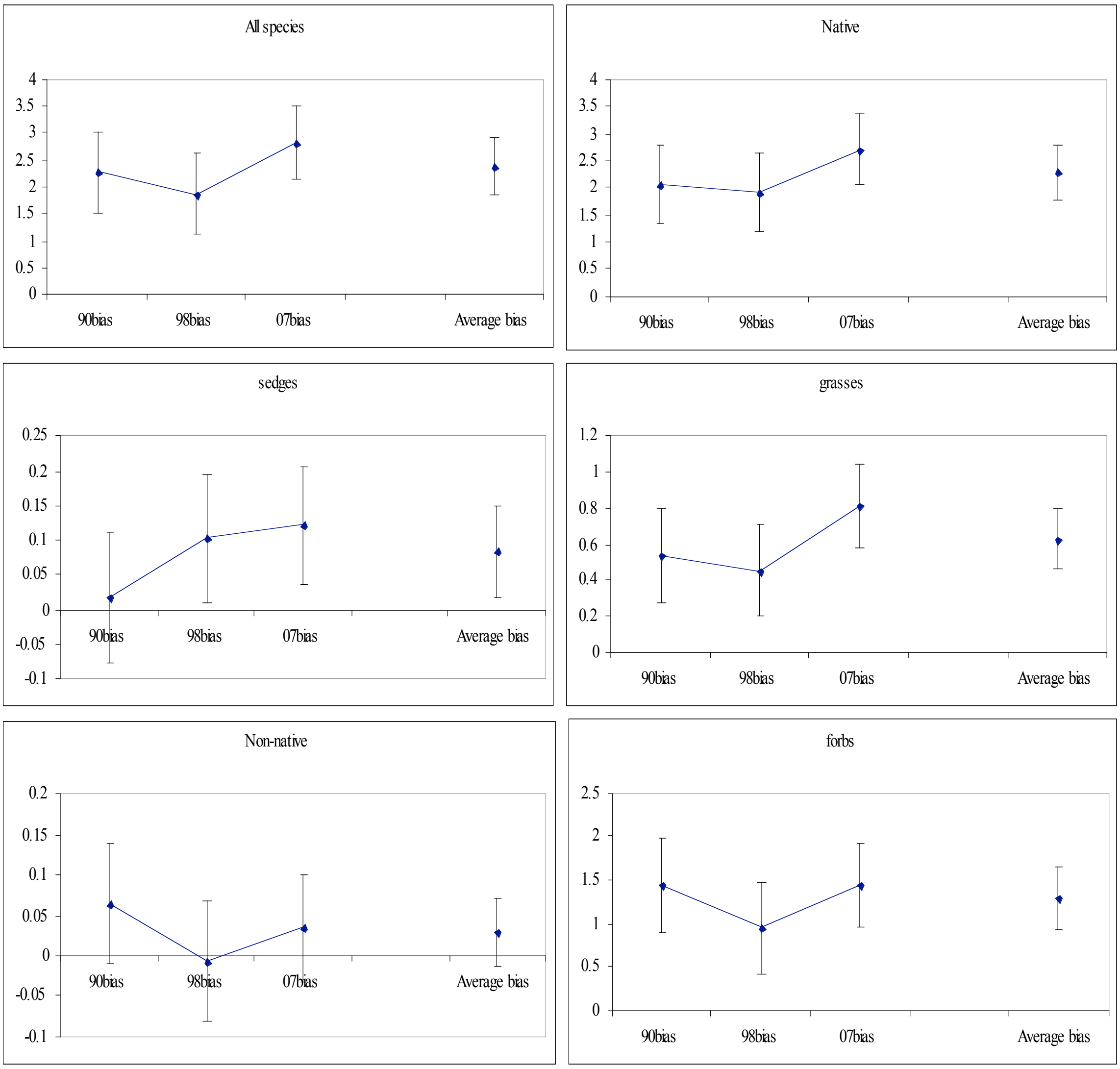
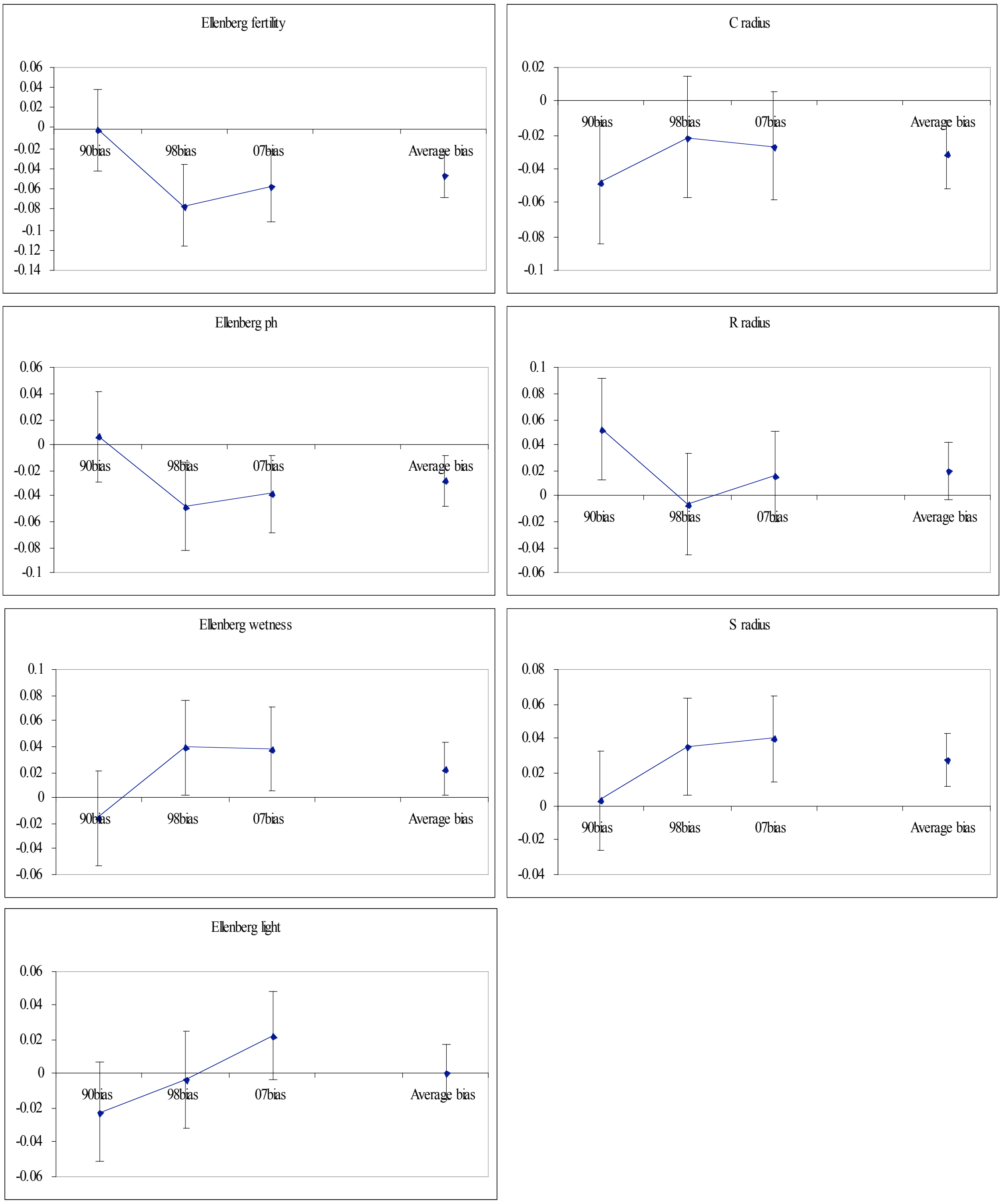


Figure 2. Estimated bias in summary properties of vegetation



3. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

It is clear from the results presented above (and the other analyses of the QA data) that there is a degree of under-recording of species in CS vegetation monitoring and that this translates into differences in measures derived from species composition such as Ellenberg scores. The difference in species richness measures is not unexpected and should not be taken as a cause of concern with regard to CS results. The QA surveyors are experts in their field with decades of experience and it would be surprising indeed if they did not perform better than the CS field surveyors who are, in general, of average competence and are under more direct time pressure than the QA surveyors. The CS field surveyor results may be more comparable to those of other large-scale vegetation surveys than would be the QA results.

The particularly poor recording of cryptogams has been noted in previous surveys and has previously been dealt with by excluding them from published CS analyses. The reason why it is substantially worse in the 2007 survey is not yet clear.

Differences in the level of under-recording across surveys is of more concern. The primary focus of CS vegetation monitoring is change in vegetation, and changes in the level of under-recording could affect results and interpretation. The results presented above show no significant change across surveys in the level of bias in species richness measures, once cryptogams are excluded, but a significant difference in the bias of derived measures, in particular Ellenberg scores, in CS1990 compared to the two later surveys.

As a result of these findings the following recommendations are made:-

1) Cryptogams should be excluded from the analysis of CS2007 data as was done for previous surveys.

2) No adjustment of CS results to correct for the bias between CS and QA results should be made.

3) For derived measures only (and possibly only Ellenberg scores) an adjustment to the 1990 survey results should be made to correct for the *differences* in bias between this survey and the others. The values for these differences in Table 3 can be used for this purpose.

4) The adjustment should take the form of a static adjustment to the derived measure values prior to full analysis.

5) If thought necessary an adjustment to the standard errors of 1990 estimates can be made post analysis using the standard errors in Table 3 and formulae for the standard error of a sum of variables.

4. REFERENCES

Prosser, M. and Wallace, H. (2008a). *Countryside Survey 2007 Quality Assurance Exercise*. Ecological Surveys (Bangor).

Prosser, M. and Wallace, H. (2008b). *Countryside Survey 2007 Quality Assurance Exercise: Additional Analyses*. Ecological Surveys (Bangor).

Scott, W.A. and Hallam, C.J. (2003). Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology*. 165(1): 101-115.