JNCC Report
No: 487

Analysis of harbour porpoise sightings data in relation to area-based conservation

Bravington, M., Borchers, D. and Northridge, S.

Compiled 2002
Published July 2014

ISSN 0963 8901

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
www.jncc.defra.gov.uk

This report should be cited as:

Bravington, M., Borchers, D. & Northridge, S. 2014. Analysis of harbour porpoise sightings data in relation to area-based conservation. JNCC Report No*. 480,* JNCC, Peterborough.

Summary

The analyses reported here are necessarily tentative, given the limited time available for the study, and the size and complexity of the dataset from which conclusions were drawn. In fact, it has been possible to draw only limited conclusions about areas of regular high harbour porpoise density (referred to herein as 'hotspots', for brevity). While it is not clear whether further development of the methodology used will locate such areas more reliably (or indicate areas of low density/absence of the species), it is clear that methodological development will enable more-reliable inferences to be made. Data were inadequate to examine calf-adult ratios.

Evidence of hotspots

For the most part, evidence of persistent areas of high porpoise density are lacking. The strongest suggestion of a 'hotspot' is in the south Irish Sea, with weaker indications of hotspots in the central Irish Sea and east of the Moray Firth and Firth of Forth.

Recommendations

- Investigate further the possible hotspots in the Irish Sea and off eastern Britain.
- Develop and apply methods that address the deficiencies highlighted in this analysis, to make optimal use of the data.

Contents

1 Introduction

The EC Habitats Directive requires member states to consider Special Areas of Conservation (SACs) for all species listed in the Directive's Annex II. The Annex includes bottlenose dolphins (*Tursiops truncatus*) and harbour porpoises (*Phocoena phocoena*). The aformentioned inhabit relatively discrete areas of coastal water in several parts of the European Union, but the latter are much more widely distributed, and there is little evidence of longterm 'residence' in any particular area. There are no obvious discrete areas where densities of harbour porpoises are high enough to suggest a significant proportion of the population is 'resident', unless such areas were to extend to ICES Division scale.

One of the best sources of information on the distribution of porpoises in north-western European waters is the 'Joint Cetacean Database' (JCD) – a database consisting of over 20 years of sightings data from three major sources. These are the EC funded SCANS project[1], the European Seabirds at Sea consortium, which also maintains records of cetacean sightings, and the Sea Watch foundation database of sightings collected from a wide variety of sources since 1979. Data from these three sources have been collated into a single JCD database, with a common format. Together, these datasets comprise a unique and valuable source of information about the distribution of harbour porpoise in waters around Britain. Note, however, the Sea Watch data were not available in time for inclusion in the analyses reported here.

The objective of the study reported here was to apply statistical modelling techniques to these data in order to look for areas of high porpoise density that might be suitable as SACs.

It is important to understand that although these data have been collated in a common format, the means of data collection have been quite varied. A wide variety of platforms have been used, so that observer eye height and vessel speed, for example, are highly variable throughout the database. It is also known that porpoises react aversively to vessels, and the degree of this reaction will depend on certain characteristics of the vessel, such as the amount of noise it makes. It is also known that there is considerable variability among observers in detecting porpoises. These and other factors can have order of magnitude effects on sightings rates, and they cannot be controlled in such wide-scale surveys. Whereas it might be hoped that such effects would not bias any interpretation of the overall sightings rates of porpoises, a casual examination of the data would reveal that this is highly unlikely. Many vessels and observers, for example, focus their activities on relatively small areas or repetitive transects, while some platforms or observers are only active for earlier or later parts of the survey period.

There is therefore a real concern that any detailed simple examination of the sightings data in the JCD could be significantly biased by the sightings efficiencies of different platforms or observers.

To address such issues, Bravington (2000a) developed a suite of statistical tools that are intended to model sightings rates of cetaceans spatially and with respect to covariates such as platform, observer and environmental conditions. These tools have been used here to examine data in the JCD in order to try to determine whether there are any areas within UK waters that show consistently high porpoise densities.

[1] Distribution and Abundance of the harbour porpoise and other small cetaceans in the North Sea and adjacent waters. LIFE 92-2/UK/027

2 Methods

2.1. Data

The JCD was initially compiled in 1999, and was based on the existing SCANS data set, the ESAS 2 data set (comprising European Seabirds at Sea data up to 1998) and a version of the Sea Watch Foundation (SWF) database.

It was decided early in the project to try to work with the latest data available, and accordingly the ESAS-3 dataset was successfully updated to JCD format with data up to the year 2000. Consultations with the Sea Watch Foundation indicated that there were some serious inconsistencies within the SWF database as supplied to the JCD, but that these were being addressed and corrected. It was initially anticipated that the SWF database would be corrected and updated within the JCD by January 2002.

It is important to understand that whereas errors in a sightings database may not be significant for generalised or large-scale geographical overviews, the statistical tools utilised have a far smaller tolerance of errors. Small errors in time or location data can seriously affect the analysis, either causing the programme to crash, or producing nonsensical results. Consequently it can be very time-consuming for the the analyst to try to identify problem records or batches of records. Given the short length of time available for analysis it did not make sense for the project team to proceed with an attempted analysis of a dataset which was known to have a large number of inconsistencies. The team therefore opted to wait for the revised SWF dataset, and to proceed in the interim with the analysis based solely on the other datasets. Unfortunately, unforeseen problems with the SWF database meant that an updated version of these data was not received until after the termination of the contract, so that no analyses were been performed on the SWF part of the JCD.

Working with the remaining data, it was found that even here there were a number of unforeseen data problems that hindered analysis. Two major data error types within the updated JCD were apparent. First, 478 out of 355,144 records were found to have effort start times that were later than their end times. Secondly, some 659 records were found to have start times that preceded the time at which the previous effort leg was supposed to have ended. The origin of these data anomalies could not be determined, but some were simple typographical errors at data entry, while others, especially of the second type, may have been introduced during data transformation from ESAS to JCD format.

It was not possible to review these 1137 records to try to correct them all manually, so that the analysis had to proceed by removing all such records and adjacent ones on the same sighting trip during the same day. This meant that 15% of the total data was excluded from the analysis.

This problem highlights the issue of data maintenance. The JCD in total represents an enormous dataset, which is also a very rich source of information. However, it also represents an enormous house-keeping task to ensure that all of the hundreds of thousands of records are error-free. Errors may be introduced at any stage to such a database, and even very thorough checks will inevitably miss some types of error. This is especially a problem where a new type of analysis is being attempted, and where errors that had been missed as unimportant to the working of previous analyses suddenly become highly significant. Considerable analytical time was therefore spent in trying to uncover and remove these errors in timing which had not previously been picked up in other analyses. This also helps to explain why the project team were reluctant to attempt any analyses on the SWF part of the dataset before it had been corrected to the satisfaction of the SWF data manager.

Despite these data problems the statistical analyses were still run with over 300,000 effort records (typically 10 minutes of observation time, or over 50,000 hours of observation in all).

2.2 Statistical analyses

2.2.1 Pseudotime

The analysis followed that of Bravington (2000a, 2000b). With this approach, the amount of effort spent searching before detecting an animal or group of animals is the key to estimation of local density. For a given level of search effort, if you searched for a long time before detecting an animal, the local density is estimated to be low. Conversely, if little time was spent searching before making the next sighting, the local density is estimated to be high. Time spent searching is therefore a key explanatory variable in the analyses. Because there are breaks in searching, the total time spent searching is not the same as the time since the start of the survey or trip. It is therefore useful to work in terms 'pseudotime', which is just the sum of all time spent searching up to that point (see Figure 1). By analogy, if each section of uninterrupted effort is compared to a scene in a feature film, then the film plays in pseudotime.

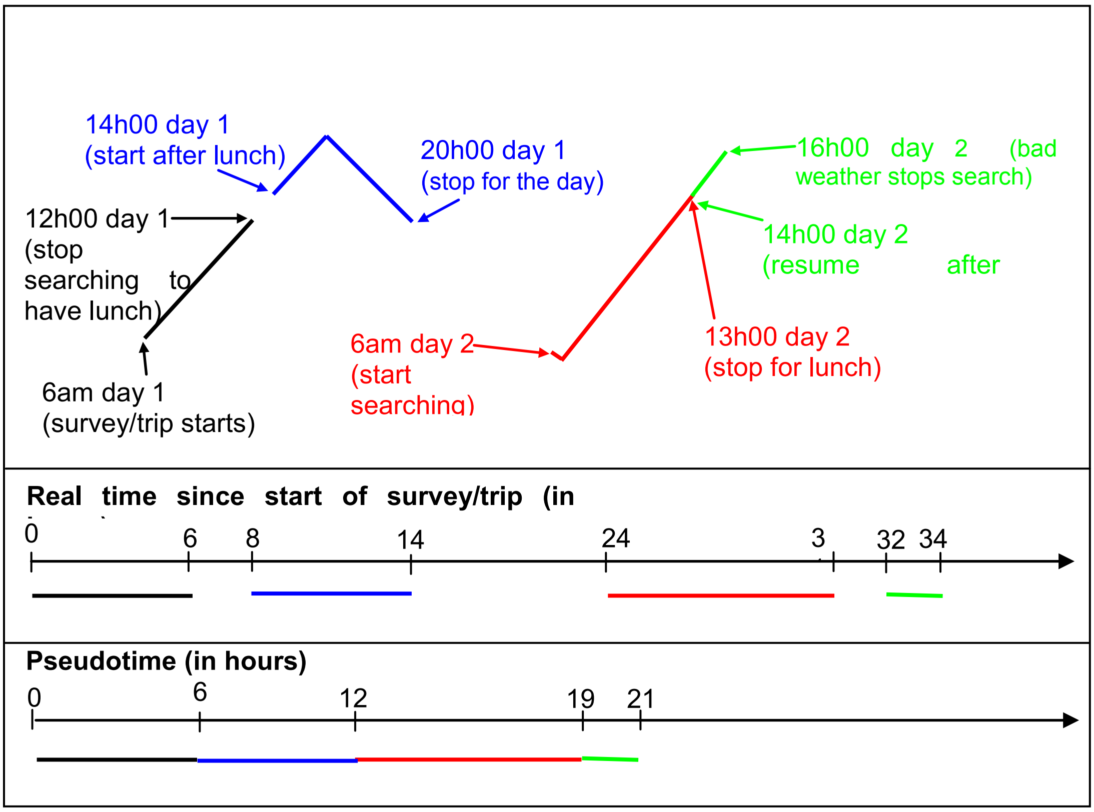

Figure 1. Example of how survey effort is tallied in terms of pseudotime. The top panel shows a section of survey transects; the middle panel shows the search effort in real time; the bottom panel shows the search effort converted to pseudotime.

Pseudotime is tallied separately for each trip. Mathematical details of the analysis methods are given in Appendices 1 and 2.

Sighting rates are modelled as a function of covariates, treating sightings as independent events (a Poisson process) in pseudotime, with the probability of a sighting in the next instant depending on the covariates only (independently of when the last sighting occurred). This general formulation needs to be adapted to the specific circumstances of these data, for example by assuming that density does not vary much within a grid square of predefined size, or over a certain season of the year. The strong independence assumption in the model is an oversimplification, but it allows the construction of a simple consistent estimator. Deviations from the independence assumption are handled by a bootstrapping approach to uncertainty (see Appendix 1).

Because the analysis is based on pseudotime, nonsensical time values cause critical errors in analysis.

2.2.2 Pod size bias

Size bias in sighting surveys arises when larger pods are more likely to be seen than smaller pods. In the context of relative abundance studies (such as looking for hotspots), size bias causes a potential problem if the underlying distribution of pod sizes varies from place to place. Unless the extent of size bias is known *a priori*, it is not obvious which area contains more animals if, say, more large pods are seen in area A but many more small pods were seen in area B. There is also a potential interaction between size bias and sighting conditions, in that it is not obvious whether relative visibility of different-sized pods will remain consistent across conditions, and average conditions may vary from place to place.

As an attempt to circumvent the problem of bias arising from unmodelled size bias, double-platform data from SCANS were used to obtain estimates of absolute sighting probability for pods of different sizes under different sighting conditions (sea state: 0, 1 or 2 used). Although the absolute probabilities will not apply generally to the JCD data because of different protocols and sighting platforms, it might be reasonable to expect that the relative probabilities follow the same pattern, regardless of platform. Further details of this analysis are given in Appendix 2. Because the analysis of non-SCANS data here rests on analysis of the effects of conditions from SCANS that operated only under sea state 2 or below, it has only been possible here to use non-SCANS sightings in sea states 2 or below. This removes about 35% of sightings. Further methodological developments (e.g. along the lines of Bravington 2002) would be required in order to utilize the high-sea-state observations.

Having established relative sighting probabilities by pod size and conditions, it is possible to estimate the frequency distribution of pod size (i.e. the probability that a porpoise pod will be of size 1 or of size 2 or of size n) as a function of local time and place, based on local observed frequencies of pod size corrected by the estimated sighting probabilities under local sighting conditions. This allows calculation of true mean pod size, and overall relative sighting probability of any pod, also as a function of time and place. A more elaborate version of this procedure is described in Bravington (2002).

In principle, the adjustments described here remove the effects of size bias, and allow comparable density estimates to be calculated across a wide range of times and places from sightings data, with the effects of sea state already allowed for. However, this rests on the assumption that the estimated adjustments from SCANS are more widely applicable. If sea state is included as a covariate in an analysis of sighting rates even after the SCANS adjustment, there is still a strongly significant improvement in Goodness of Fit (about 81 units of log-likelihood for only 2 degrees of freedom). Even after allowing for the directly-estimated effects on sightability from SCANS, sighting rate data suggest a further reduction in sightability of around 65% in sea state 2 compared to sea state 0. Either there are difficulties with the double-platform analysis (e.g. over the appropriateness of the linearity assumptions), or the effects of sea state on SCANS-double-platform-style observations are not applicable more widely. The implications of this for the size-bias corrections above, require further investigation.

2.2.3 Patchiness

Analysis using the methods described in Appendix 1 reveals significant clustering/patchiness of sightings. This apparent clustering potentially reflects several issues, including genuine short-term fluctuations in local abundance, so that surveying the same line twice in successive weeks can produce quite different results; unmodelled observer and platform effects; and systematic variations in local abundance on a scale too small to include in the model.

Analysis suggests that the effective sample size is only about 45% of the observed sample size as a result of patchiness. At an intuitive level, the effective sample size corresponds to the number of independent bits of information available in the sample. Patchiness decreases the effective sample size because the data from sightings within a patch are not independent. (Consider an extreme case in which the patch is concentrated at a point; in this case as soon as you observe one animal in the patch, you know the location of all the others in the patch, and the effective sample size from the patch is 1.) Because lower effective sample size increases uncertainty, CVs for these data are always likely to be about 50% higher than an analysis assuming independence would suggest.

One way to think about this, is to consider what sample sizes would be necessary to give a good chance of detecting a statistically significant difference between two areas, assuming for example that the real density is 50% higher in one area than the other. In that example, it turns out that, with about 50 sightings per area, a test would successfully detect the difference at the 5% level about half the time. For a more reliable test, many more sightings would be required. However, more extreme variations in density might be detected reliably with fewer sightings. These ideas set some practical limits on the size of region which can be tested for local effects, based on the number of sightings in that region.

This very rough analysis is not definitive, for several reasons: it does not attempt to take into account individual observer/vessel effects; it does not allow for estimation of other parameters, such as platform or sea state effects; does not allow for confounding between particular platforms/conditions and particular times/places; it is dependent on the spatial scale of the original model from which autocorrelations were estimated; and it does not reflect local patterns in segments (e.g. whether or not several different platforms were used within one area, whether observations were made over one short period or split over several periods some time apart). In addition, it does not reflect uncertainties about pod size. Most of these factors will mean that real CVs are worse than one might expect, even after allowing for clustering. However, the clustering-based CV correction does take account of the main factor controlling uncertainty in local estimates of abundance/density, namely the absolute number of sightings.

3 Results

3.1 Breaking up the analysis by geographical zone and by season

Because of the geographical complexity of the entire study region, and the wide variations in amount of sampling, the analysis has been broken up into three main geographical areas: Irish Sea, North-West coast (Western and far Northern Scotland), and Eastern Britain. In general, these areas have been analysed separately. In theory it is possible to fit completely different geographical models within each area, while still fitting a shared term to describe, for example, the effects of sea state or platform speed across all regions, but this complicated procedure was not possible in the time available for the analysis.

An important point is how sparse the non-summer observations have been over the last 10 years, apart from a few set routes. This is shown in Figures 2 and 3:

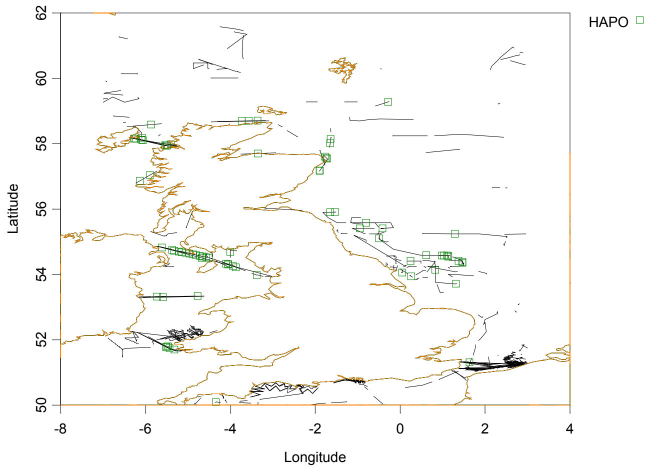

Figure 2. Winter tracklines (January to April) and sightings

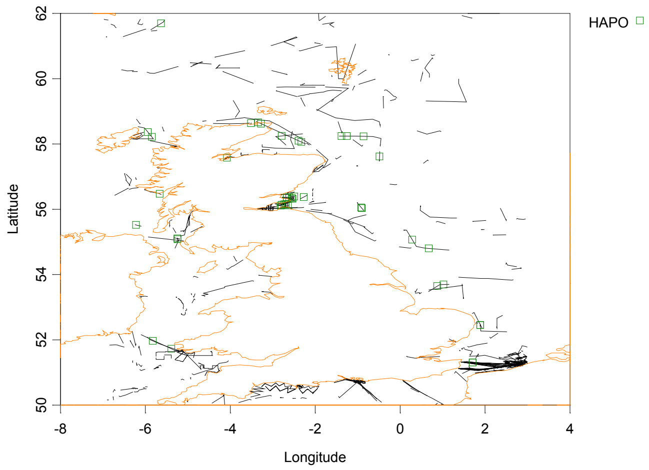

Figure 3. Autumn tracklines (September to December)

Apart from the usual paucity of sightings in the English Channel, it is clearly impossible to make any rigorous inferences about hotspots at those times of year since 1990, because effort and sightings are relatively sparse. However, some seasonal modelling is possible in some areas, assuming that seasonal patterns pre-1990 have been consistent.

3.2 Irish Sea

The data for the Irish Sea illustrates many of the practical and methodological issues common to all areas, and the remarks below are pertinent to analyses of the JNCC data as a whole.

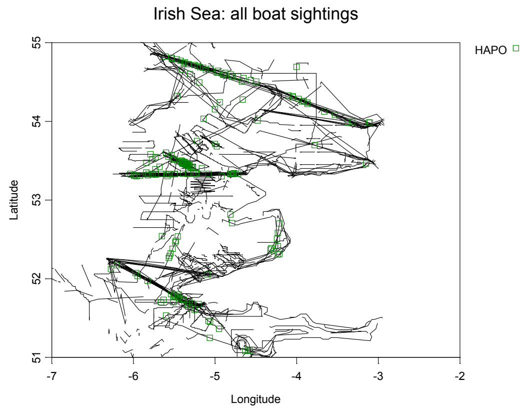

Figure 4. Tracklines and sightings in the Irish Sea region

Of the 280 boat-based sightings at sea state < 3 in the Irish Sea (Figure 4), almost all are concentrated along set routes. This again makes identification of hotspots difficult. Along the Welsh coast, there is a cluster of sightings around 52.5°N despite low effort, but there are only 7 sightings here, which is far too few to make any reliable statement (see below). These are aircraft sightings, which have been excluded from analyses here because no independent corrections for size-bias effects can be applied; there is no reason to assume that SCANS-based corrections will apply to aircraft.

The cluster of sightings around 5.5°W, 51.7°N is more interesting, although there are still only 33 sightings in total (see the discussion on Clustering above). A fit to sighting rate, adjusting for platform speed and sea state, suggests that about 80% of the abundance in the box bounded by [52.3°N, 51.5°N, 4.5°W, 6.8°W] lies within about 30nm of 5.5°W, 51.7°N, with a standard error of about 10%. This is based on the whole time period considered, but with very few sightings in winter.

The southern Irish Sea sightings are based on several boats and observers, so the apparent concentration is unlikely to be due to a boat or observer effect. However, to fully understand the contribution of “southern animals” to the abundance in the whole Irish Sea), it would be necessary to allow for the possibility of boat and/or observer effects, because different boats and observers are used in the south and in the north.

This raises a statistical problem, because of the limited number of samples available. Although it is simple to fit a ‘boat effect’ and an ‘observer effect’, the number of per-boat and per-observer observations is very small. About 16 Irish Sea observers failed to see any porpoises at all. As far as a naive statistical model is concerned, the most restrained explanation is that these observers had zero probability of detecting porpoises, but this is not very realistic.

The issue of unwanted ‘random’ covariates is particularly problematic for presence/absence data such as sightings, because there is a realistic possibility of getting no sightings for some observers/platforms, which will correspond to estimates of zero probability of detecting porpoises. Further, in platform of opportunity datasets such as in JNCC, coverage of some areas and times is completely correlated with who was doing the observing, so data from observers who make no sightings will convey no information at all about abundance in those areas.

To get around this problem, it is possible to fit 'random effects' models, which take account of the fact that all observers and boats are drawn from a common population. Among observers with plenty of accumulated sighting effort, it is possible to see how much sighting rates vary (after allowing for other factors such as time, place, and platform). If for these observers, there is only a 30% difference between the best and the worst, then a similar range is likely to hold for other observers without enough observation time to be able to develop an accurate estimate of personal sighting rates. Random-effects models include this variation; in this framework, it becomes more parsimonious to ascribe one observer's lack of sightings to a low but finite spotting ability, rather than to an infinite lack of talent. This allows some inferences to be made about animal density in areas covered only by that observer.

The difficulty with random effect models, is knowing how much variance to ascribe (say) to between-observer variability, versus to between-platform variability, versus to how much density can realistically fluctuate over a short space-time scale. In linear-model statistics, this is by now regarded as a standard problem, although the data requirement can still be quite high to get precise estimates. Methods do exist for objectively 'trading off' variances between the different components of the model, and the methods could in principle be translated to sighting-rate models of the sort used here, but this has not yet been done. There is a significant complication because of patchiness in the sightings.

However, by making different assumptions about the variance trade-offs, it is at least possible to fit random-effect models to sighting rate data, using 'shrinkage' (see Appendix 1); it is not yet possible, though, to determine which model is most appropriate. A number of such models were explored for the Irish Sea, with random effects ascribed to observers, platforms, both, or neither. It turns out that the *qualitative* pattern of abundance in the Irish Sea, is similar whatever model is fitted (see Figure 5 below, where scales have been omitted because the point is to show the pattern). There are three (sometimes two, with the northernmost absent) peaks in abundance, with reasonable separation in distance. The estimated contrast between high (yellow) and low (red) densities is substantial (four-fold or more), except for models where almost unrestricted observer effects *and* vessel effects are included. With both observer and vessel effects, it is possible to explain almost all the variation in densities without invoking much spatial change in density, although the small estimated variations are qualitatively similar to those shown above.

It is of concern, however, that these peaks occur only in places where there has been quite heavy sampling. Although not every heavily-sampled area is suggestive of high densities, the question must arise of what 'hotspots' might be unobserved in areas where sampling has been less intense.

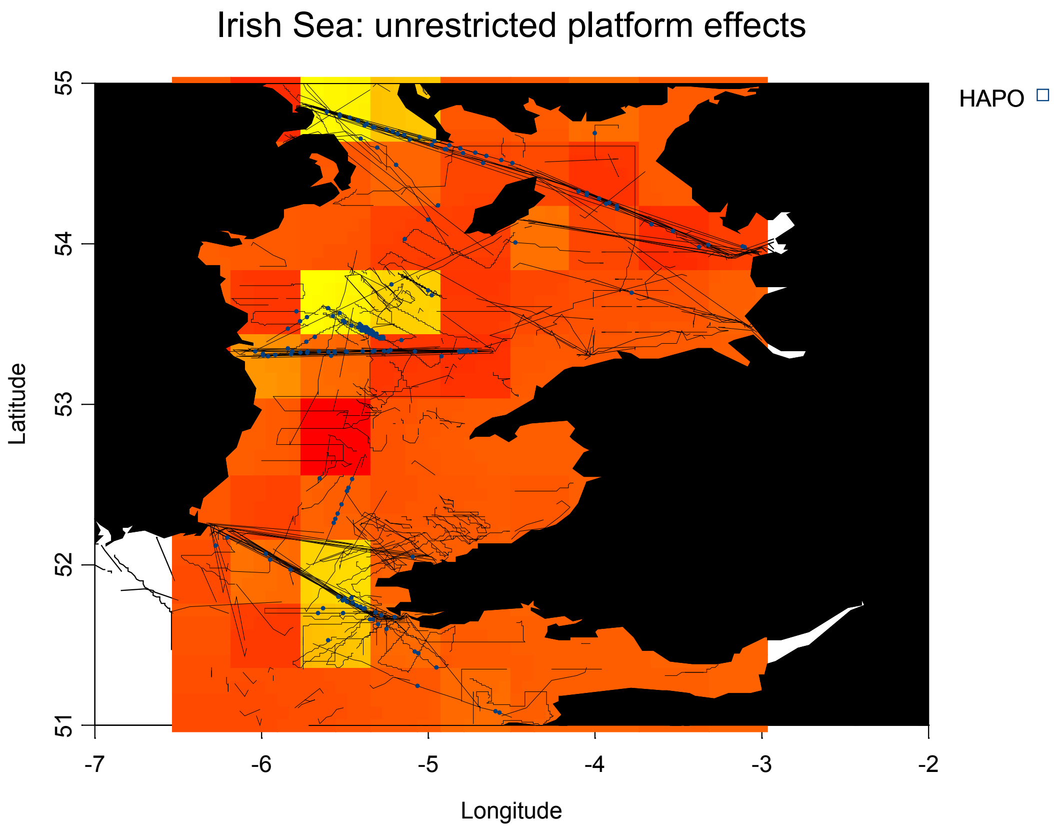

Figure 5. Estimated relative distribution of harbour porpoise in the Irish Sea

Further, almost all of the sightings in the southern Irish Sea are pre-1995, and almost all the sightings in the north are post-1995. This makes north-south comparisons particularly difficult.

At present, all that can safely be said is that there is a strong suggestion of a local concentration within the southern Irish Sea, around 5.5°W, 51.7°N, but it is too early to assess how significant this is for Irish Sea porpoises as a whole. Possible movement of porpoises in and out of this region should be investigated before strong conclusions about a hotspot are drawn.

There is also a hint of a high density patch around 5.5°W, 51.7°N, but effort is somewhat sparser in this region and the patch may be a sampling artefact.

3.3. Eastern Britain

There are no obvious gaps in distribution, in either summer or winter.

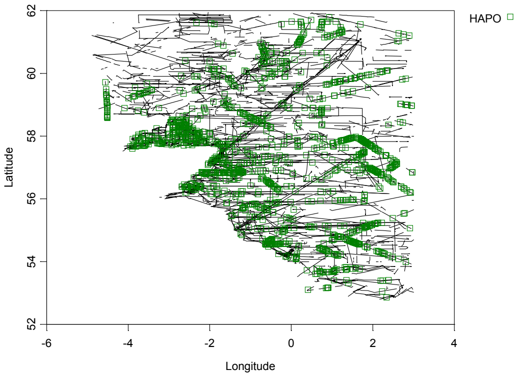

Figure 6. Tracklines and sightings off Eastern Britain. Summer

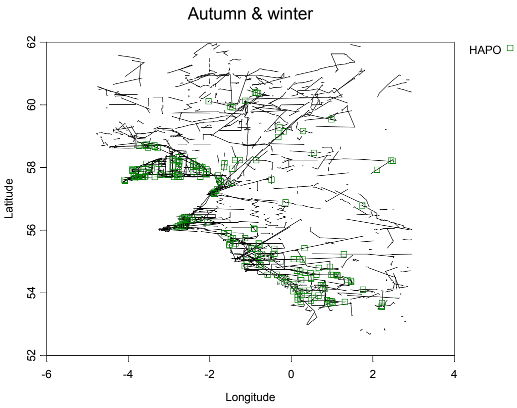

Figure 7. Tracklines and sightings off Eastern Britain. Autumn and Winter

Mean pod size and size bias effects were investigated as described under Section 2.2.2 for this subset of the data. Various model structures were explored, and the use of Akaike's Information Criterion (Akaike, 1973; an objective way of trading off increased complexity against apparent goodness-of-fit) suggested that a complex model with seasonal effects, distance from shore, and distance along shore effects within season, was most appropriate. However, no particularly consistent pattern emerged from subsequent analysis, and this issue therefore requires further investigation. In autumn and winter, estimated mean pod sizes were largest offshore and to the North, but the opposite was evident in summer.

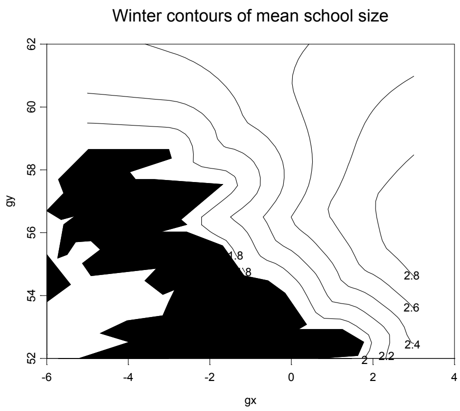

Figure 8. Mean pod size contours in Winter

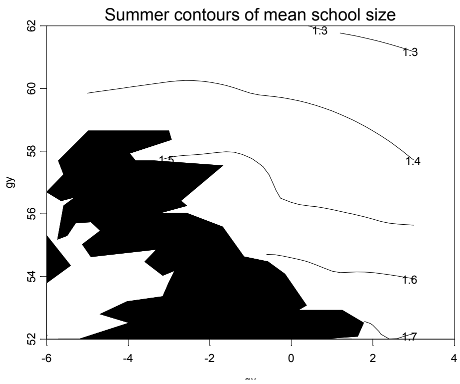

Figure 9. Mean pod size in Summer

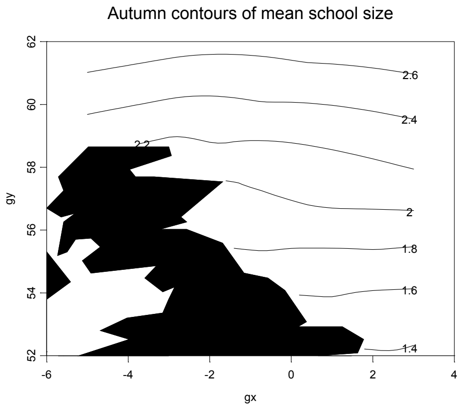

Figure 10. Mean pod size contours in Autumn

At least with current techniques, smooth fits of sighting density have a tendency to place illusory hotspots in low-data areas. For hot-spot purposes, it seems safer to divide up the area of interest in grid-fashion, with density in each grid cell being estimated independently (and simultaneously estimating other shared effects, such as sea state). Even when using grids[2], strange estimates (very high abundances) can result when there are very few observations; in the results below, obviously strange results have been removed. For the British East coast, a 'grid' was used based on distance offshore and distance alongshore (moving towards the northwest), the outlines of which are apparent in Figures 8 to 10.

The graphs below show the estimated relative porpoise abundance (shading, with darkest red corresponding to zero) and the corresponding percentage of total abundance (the numbers) within each 'grid square', within the time period of the title.

[2] Decisions on grid size were made on the basis of rough calculations that suggested about 50 sightings per grid rectangle might be required to detect a twofold increase in density between a hotspot in the rectangle and that surrounding it. There is always some subjectivity in decisions on grid sizes and locations.

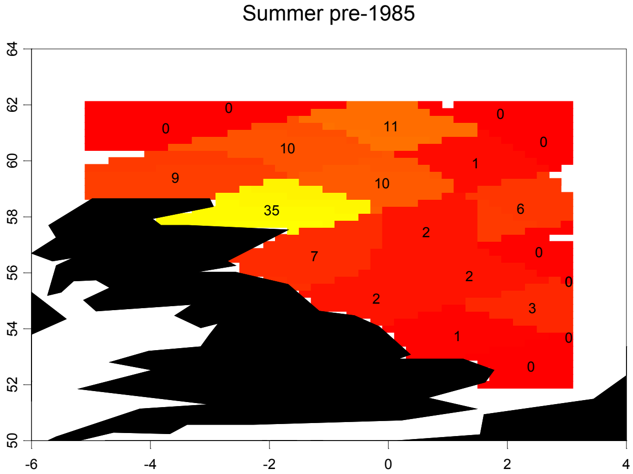

Figure 11. Estimated relative distribution off Eastern Britain: Summer pre-1985

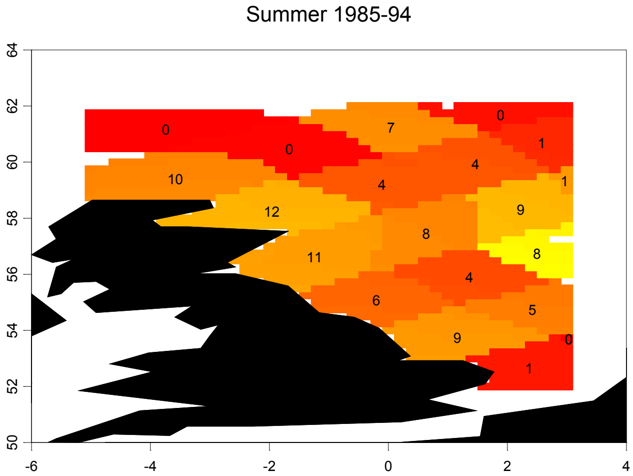

Figure 12. Estimated relative distribution off Eastern Britain: Summer 1985-94

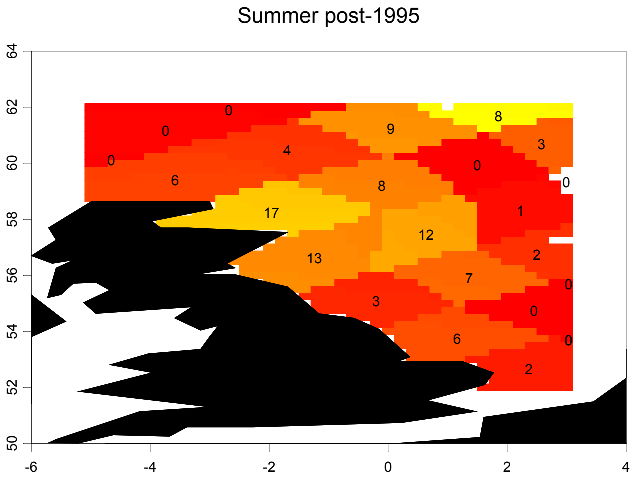

Figure 13. Estimated relative distribution off Eastern Britain: Summer post-1985

The three plots consistently show higher abundances of animals in the area east of Scotland and up towards the Shetlands (which are not shown), in particular near the Moray Firth. This is particularly apparent in the pre-1985 results; however, as shown below, there was little effort outside the Moray Firth in those years, so the pre-1985 patterns are particularly uncertain. Post-1985, there are several hundred sightings during summer in each of the two blocks east of the Moray Firth and Firth of Forth, and several hundred sightings outside these blocks, so these results are unlikely to be artefacts of sample error. Of course, other phenomena such as unmodelled platform effects in particular areas, may still have affected the results. Note that it has not been possible to explore random effects models for observers or platforms in the North Sea. Out of 100 boats used in this area, 45 recorded no sightings (at least within this reduced dataset), and over 85% of all sightings were made by just 10 boats.

Even though abundances are generally higher in this region, it is not obvious that the region is a ‘hotspot'' in any useful sense. The distribution covers a large area of sea, makes up no more than 50% of total abundance within the map area, and there are no obvious gaps in distribution (see the effort data). There may still be more local hotspots, but there has not been time to examine this during the current study. In any case, any such smaller local hotspots cannot contain a high proportion of North Sea porpoise population, since the whole east-of-Scotland region seems to contain less than 50% of the total abundance shown on these maps.

The median number of summer sightings in each grid squares above is 43, and the mean is about 67, so inferences *across the whole period* on a per-grid square basis are likely to be fairly reliable. Inferences based on shorter periods (e.g. the decadal intervals shown above) are based on fewer sightings and are thus likely to be less reliable. Nevertheless, the maps show some suggestion of spatial patterning at a scale larger than that of the individual grid square, which lends credence to the results.

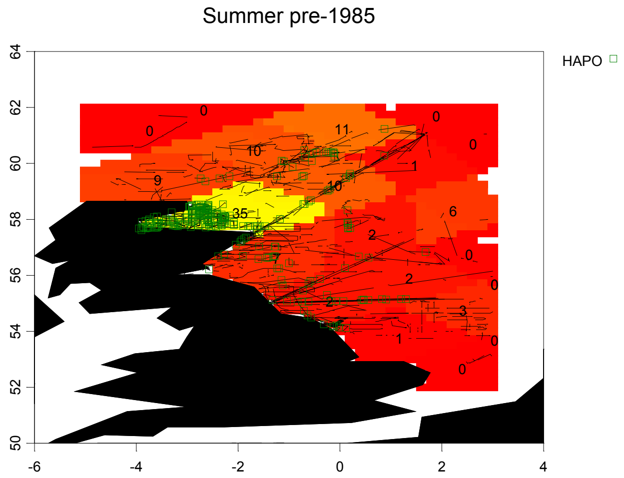

Figure 14. Relative distribution, location of effort and sightings: Summer pre-1985

For non-summer distribution, lack of data makes it difficult to extract separate decadal effects, so a model without a long-term time effect was used (incorporating summer data for purposes of estimating vessel effects). Only two of the grid squares have more than 40 non-summer sightings from 1978 onwards. Clearly, this means that it is very difficult to draw small-scale inferences.

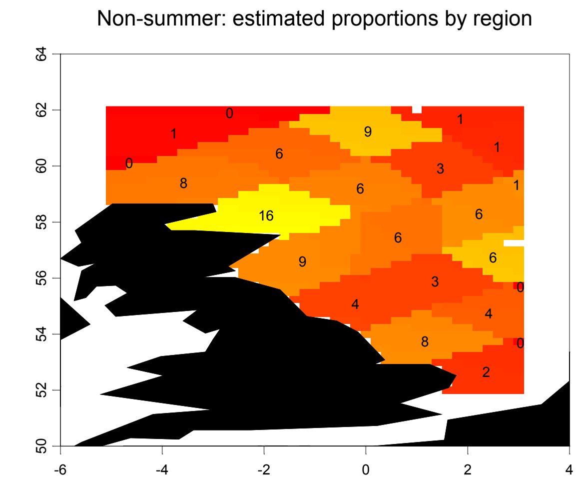

Figure 15. Estimated relative distribution off Eastern Britain: Non-summer

3.4 The North-West Coast

Sightings here are fewer in number; there are only 270 boat-based sightings in sea states up to 2 over the whole period (see Figure 16). The most obvious feature of the data, is the concentration of sightings between the Outer Hebrides and the mainland. However, this is also where effort is highest.

Analysis does suggest a significant ‘North Minch effect’, with densities about 60% higher within a 45nm circle of 58°N,5°45’W than outside. However, most of this reflects an inshore/offshore effect. Restricting attention to effort within 70nm of the coast, actually suggests that coastal densities are lower within the North Minch. There is therefore no real evidence of a high density region here.

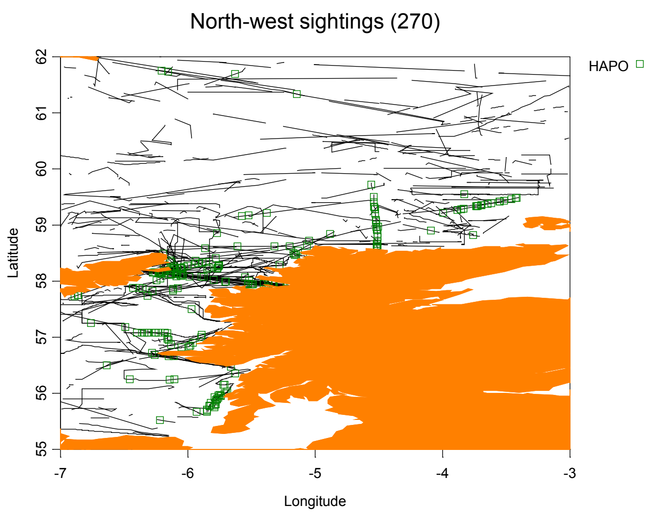

Figure 16. Location of effort and sightings off North-west Britain

4 Results

In addition to perpendicular distance, *x*, the set of covariates considered was Beaufort (*bf*), vessel (*ves*), platform height (*hgt*), pod size (*ss*), depth of sea bed (*depth*), distance to the nearest coast (*dist*), latitude (*lat*) and longitude (*lon*). The covariate vessel, was entered as a factor with five levels using the vessel groupings determined from the original SCANS analysis (Hammond *et al.* 1995: 1 = Abel-J; 2 = Corvette, Dana, Gorm; 3 = Henny; 4 = Holland, Tridens, Isis; 5 = Gunnar Thorsen) and Beaufort was fitted as a factor with 2 levels; Beaufort 0 and Beaufort 1-2. Stepwise model selection was used.

Starting with all the covariates listed above, the model selection procedure chose the covariates *x*, *bf*, *ss*, *depth*, *dist*, *lat* and *lon*. This is referred to as model 1. The fitted detection function is shown in Figure 17. The estimates of the regression parameters are given in Table A2.1.

Excluding latitude and longitude from the set of covariates available for inclusion, results in a model (model 2) which contains the covariates *x*, *bf*, *ss*, *depth* and *ves*. Thus, the spatial coordinates (*lat*, *lon*, *cdist*) have been replaced with the vessel factor. Since one vessel surveyed one block, apart from one block, these covariates are essentially confounded. The estimates of the regression parameters for model 2 are in Table A2.1 and the detection function is plotted in Figure 18.

A further 4 models were fitted, as follows:

Model 3: *seen ~ x + bf + ss + ves*
Model 4: *seen ~ x + bf + ss + ves + depth*
Model 5: *seen ~ x + bf + ss + ves + dist*
Model 6: *seen ~ x + bf + ves*

Results are shown in Table A2.2. Although it is not the best model for the SCANS data, Model 3 has been used in the analysis of JNCC data because other models contain location-specific covariates (*lat*, *lon* and/or *depth*) and the effect of these outside the SCANS survey area might be quite different to their effect within it.

Table A2.1. Estimates of the regression parameters for the selected covariates and the residual sums of squares (RSS) for models 1 and 2.

Covariates	Units	Estimates of regression parameters	
		Model 1	Model 2
Intercept		-11.1090	-0.0425
x	km	-3.0190	-2.9927
bf (1-2)		-0.5937	-0.5871
ss		0.4705	0.4703
depth	metres relative to sea level	0.01002	0.0049
dist	km	-0.0034	
lat		0.1978	
lon		-0.0681	
ves (2)			-1.2477
“ (3)			-0.3180
“ (4)			-0.9110
“ (5)			-1.0459
RSS		519.47 on 611df	521.63 on 610df

Table A2.2. Estimates of the regression parameters for the selected covariates and the residual sums of squares (RSS) for models 1 and 2.

Covariates	Model 3	Model 4	Model 5	Model 6 ss 1	Model 6 ss 2	Model 6 ss 2+
Intercept	-0.5244	0.0493	-0.4830	-0.6031	2.03402	0.8594
x	-2.9520	-3.0361	-2.9341	-2.7247	-3.4978	-2.7776
bf (1)	-0.5025	-0.5256	-0.4987	-0.1893	-0.7978	-0.8192
bf (2)	-0.6569	-0.6733	-0.6512	-0.5413	-0.7225	-0.5666
ss	0.4667	0.4647	0.4671			
ves (2)	-1.1841	-1.2365	-1.1183	-1.0490	-2.4800	-0.5930
ves (3)	-0.0051	-0.2939	-0.0319	0.1987	-0.9652	-0.4731
ves (4)	-0.5275	-0.8843	-0.4803	-0.1447	-1.7101	-0.2094
ves (5)	-0.7521	-1.0428	-0.7720	-0.4439	-1.8867	-0.6904
depth	-	0.00486	-			
dist	-	-	-0.00078			
RSS	525.5 on 610df	521.4 on 609df	525.4 on 609df			

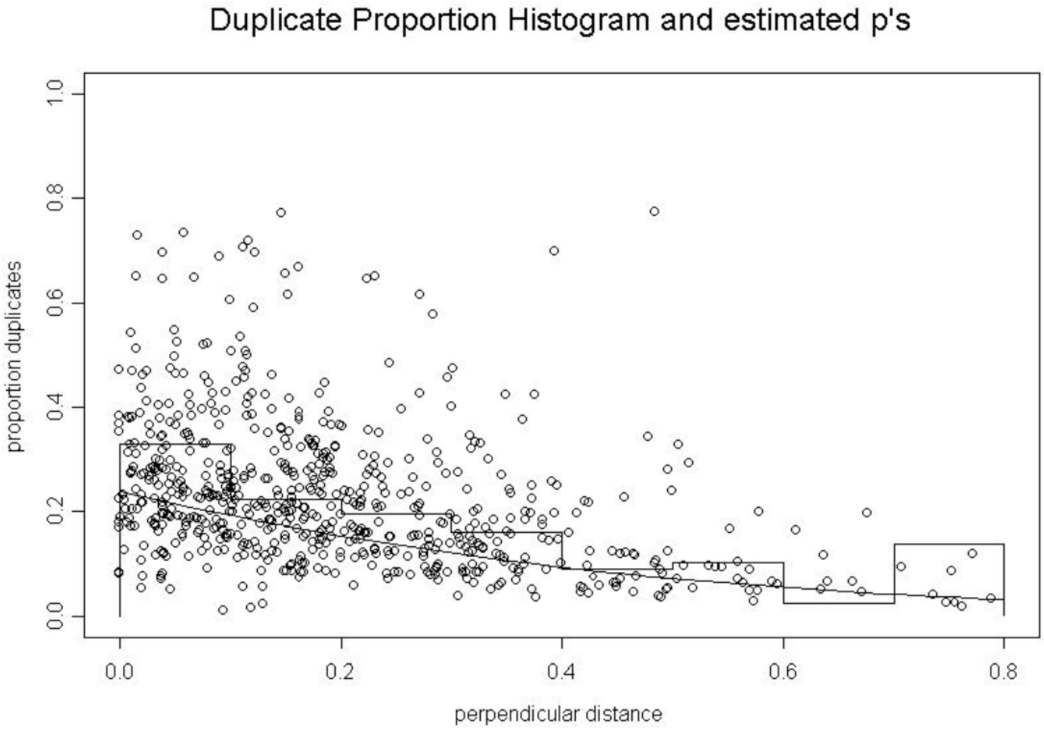

Figure 17. The smooth curve shows the detection function for model 1. The dots are the estimated detection probabilities for individual detections and the stepped curve shows the duplicate proportions.

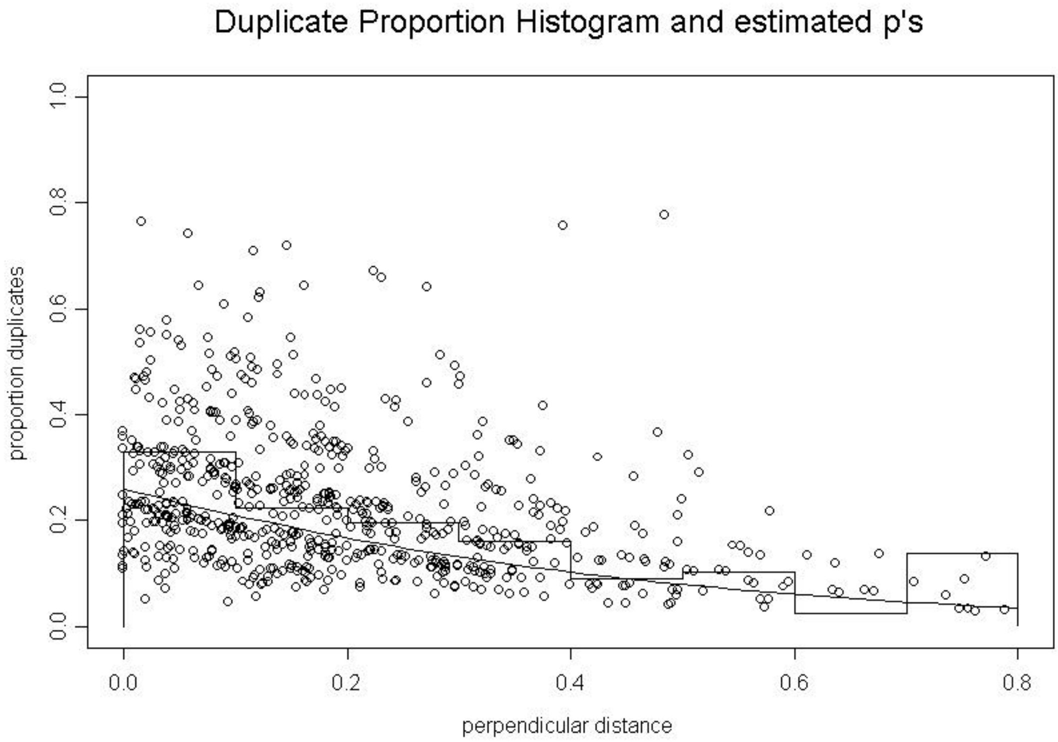

Figure 18. The smooth curve shows the detection function for model 2. The dots are the estimated detection probabilities for individual detections and the stepped curve shows the duplicate proportions.

5 Conclusions

This study combines an investigation of a complex dataset, with many different factors involved in a partly-aliased design (i.e. some variations in sighting rates could be explained equally by spatial effects or by platform effects), with a specific and fairly difficult question: identification of hotspots without pre-defined spatial scales, against a background of incomplete sampling of the full range of possible hotspot locales.

Evidence of clear hotspots is lacking. The available evidence is as follows:

- Irish Sea: There is a suggestion of a hotspot in the south Irish Sea. However, there are significant gaps in effort along the coastline to the north, which make it hard to assess the ‘separateness’ of the apparent southern Irish Sea hotspot. Aeroplane data may be helpful here, but some methodological work would be needed (see below) to work out how to incorporate those data. The suggestion of a hotspot in the central Irish Sea should be viewed with caution at this stage, because it is entirely due to concentrated effort from a single platform.
- North-West Britain: There is not enough data to the North-West of Britain to say much about hotspots, except to eliminate the north Minch as a hotspot.
- Eastern Britain: There may be high density regions just east of the Moray Firth and Firth of Forth. The distribution off eastern Britain merits further (finer-scale) investigation in a future study.

This is a rich and complex dataset, and given the constrained time available, it has been possible only carry out limited data analyses. Particular caveats attach to the following:

- Imperfect proxy for vessel effects: It was necessary to use a very imperfect proxy for vessel effects, here based on mean vessel speed on the assumption that it will be correlated with variables affecting sightability such as platform height. However, two vessels with the same mean speed, could differ widely in their utility as platforms-of-opportunity because of bridge layout, height, number of observers (e.g. SCANS vs. ESAS), vessel noise, etc. Since certain areas tend only to be surveyed by an individual vessel, there is serious potential for confounding between vessel effects and small-scale area effects.
- Possible individual observer effects. Previous experiments with JNCC data (Bravington *et al*. 2001) did not show much success in incorporating 'observer experience' as a covariate. Nevertheless, individual observers are known in other studies to vary widely in their porpoise-spotting prowess; this is likely to be particularly true when the main focus of observation is sometimes on other species such as seabirds. Again, some fairly sophisticated random-effects models might be useful.
- Time trends in sightability: This is a particular problem when there are too few data in an area to allow estimation of a time trend. Previous experiments with ESAS data have suggested a surge in sighting rates in the Moray Firth to an extent that seemed unlikely to be accountable purely in terms of porpoise abundance (Bravington *et al.* 2001). It may be that observers in some regions become more or less attuned to making porpoise sightings over time. When observations are so sparse that long time periods have to be lumped, there is a risk that data collected at different times in different areas will not be comparable. Investigation of this issue really requires detailed and time-consuming analysis of local data.

This analysis has necessarily been partly exploratory in its nature. While the results give some indication of the possible location of what might be high density regions, stronger conclusions regarding high density regions cannot be drawn without substantial further work. In particular, substantial methodological development will be required to address the issues raised above. This will require a substantial commitment of analysis time, probably of the order of 6 months of a postdoctoral researcher's time. As with any methodological innovation, there it is difficult to predict the extent to which the methodological development will answer the questions posed. What is clear, however, is that the data themselves cannot be changed and methodological development that addresses the methodological difficulties encountered in analysing this dataset will result in better use of these data and more reliable conclusions.

In order of importance, the main methodological issues that require further attention are:

(1) A full development of random-effect sighting-rate models in the context of patchy data.
(2) A better development of spatial smoothers, to cope properly with coastlines, varying scales of smoothing, patchiness; in particular, to avoid the symptom of wild results around unsampled edges of the region of interest.
(3) Consideration of better approaches to linking long-term and seasonal trends, to spatial models.

(4) Extension of pod-size-bias approaches, to sighting conditions not covered in SCANS.

At the time of the original MAFF-JNCC study (Bravington *et al.* 2001), where the remit was primarily to consider overall abundance, the issue of greatest concern was pod size bias. In the current study, this has been resolved at least partly by using SCANS data and new statistical methods. Of the remaining issues, although the random-effects and spatial-smoother issues were certainly in the air at that time, both are more important in the context of hotspots than they were for overall abundance. The reason that random-effects are particularly important, is that specific areas are quite likely to be covered only by one or few observers/boats, so that getting the observer effect right is very important; for overall abundance, on the other hand, one can legitimately expect that observer effects will to some extent be averaged out. As far as spatial smoothing is concerned, the tendency of existing GAM-based smoothers to put 'spikes' of abundance in unsampled areas, which is certainly disconcerting when considering total abundance, is potentially catastrophic when considering comparative abundance within an area. It is for that reason that grid-based density estimates have been used in this study as the least unreliable method available.

6 Recommendations

- Investigate further the possible hotspots in the Irish Sea and off Eastern Britain.
- Develop and apply methods that address the deficiencies highlighted in this analysis, to make optimal use of the data to detect hotspots.

7 References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory*, 2nd edn (eds. B.N. Petran and F. Csaaki), Akadeemiai Kiadi, Budapest, Hungary, pp 267-281.

Borchers, D.L., Zucchini, W. & Fewster, R.M. 1998. Mark-recapture models for line transect surveys. *Biometrics* **55**:11-24.

Borchers, D.L. & Burt, M.L. 2001. *MLRT 1.0* Research Unit for Wildlife Population Assessment, University of St Andrews, St Andrews, Fife, UK.

Bravington, M.V. 2000a. Covariate models for continuous-time sightings data. Paper IWC/SC52/RMP14 to IWC Scientific Committee.

Bravington, M.V. 2000b. Coping with clustered sightings. Paper IWC/SC52/RMP15 to IWC Scientific Committee.

Bravington, M.V., Northridge, S.P. & Webb, A. 2001. An exploratory analysis of cetaceans sightings data collected from platforms of opportunity. Report to MAFF.

Bravington, M.V. 2002. Spatial analyses of southern hemisphere minke whale data, allowing for size bias and sightability. Paper IWC/SC54/IA21 to IWC Scientific Committee.

Buckland, S.T., Breiwick, J.M., Cattanach, K.L. & Laake, J.L. 1993. Estimated population size of the California gray whale. *Marine Mammal Science* **9**: 235-249.

Buckland, S.T. & Turnock, B.J. 1992. A robust line transect method. *Biometrics* **48**:901-909.

Hammond, P.S., Benke, H., Berggren, P., Borchers, D.L., Buckland, S.T., Collet, A., Heide-Jorgensen, M.P., Heimlich-Boran, S., Hiby, A.R., Leopold, M.F. & Øien, N. 1995. Distribution and abundance of the harbour porpoise and other small cetaceans in the North Sea and adjacent waters. LIFE 92-2/UK/027.

Appendix 1

A1.1 Spatial modelling methods

Mathematically, the formulation for instantaneous sighting rate is

$$\mathbf{P}[\text{sighting in } (\tau, \tau+\delta\tau)] = e^{\lambda(\tau)}\delta\tau + O(\delta\tau)^2$$
$$= \exp\left\{\sum_{s=1}^{S} h_s\left(x_0(\tau), x_1(\tau), x_2(\tau), \cdots, x_R(\tau)\right)\right\}\delta\tau + O(\delta\tau)^2 \quad [1]$$

where $\delta\tau$ is a short time step. The real-valued functions $\{h_s : s \in 1 \cdots S\}$ are to be estimated, although their total number and individual forms are specified in advance: each h_s can be a linear, discrete, or general smooth function of some or all covariates. Typically, each h_s will depend on only one or two of the x_r, and each x_r will usually appear in at most one h_s. The exponential operator in [1] ensures that the instantaneous sighting rate $e^{\lambda(\tau)}$ is non-negative. Note that λ is sometimes used to indicate the rate of a process; here, though, the rate is e^{λ}.

Based on equation [1], it is easy to show that the log-likelihood of all the sighting events is

$$\Lambda(h) = -\int_0^T \exp\left\{\sum h_s\left(x(\tau)\right)\right\} d\tau + \sum_{i=1}^{N} h_s\left(x(\tau_i)\right) \quad [2]$$

Setting up the model entails deciding what form each of the functions h_s should take. For linear and discrete terms, we can just follow the usual procedure for setting up generalized linear models. In other words, if h_s is to be a linear function of x_r, then we write

$$h_s(x) = \alpha_s x_r$$

where α_s is a parameter to be estimated. If h_s is a discrete function of the 'factor' x_r, then h_s is replaced by a set of dummy functions with the q^{th} dummy taking the value α_q or 0 according as x_r takes its q^{th} possible value. The coefficients α_q are to be estimated. The usual rules apply for factor coding, contrasts, and extensions to interactions of two or more factors. This extension of h_s from one to several functions does not disrupt the notational convention of [2]; we can simply define a new extended set of h_s that includes the new dummies but not the original discrete function.

A1.2 Smooth terms

A smooth (but not necessarily linear or quadratic) form for h_s. cannot be obtained directly by using a smoother like S-PLUS's 's' or 'lo' functions, because our data are recorded in continuous time rather than at discrete points. However, we can mimic the effect of such a smoother by using pseudosplines (Hastie, 1996). A pseudospline replaces the single term $h_s(x(\tau))$ with a set of terms $h_{1s}(x(\tau)), h_{2s}(x(\tau)), \ldots, h_{ps}(x(\tau))$ [3] where the h_{is} are called 'eigenfunctions'. The first eigenfunction h_{1s} is a linear function of its argument, h_{2s} is roughly quadratic, and higher-order eigenfunctions are progressively more wiggly. To get round the problem of overfitting, the coefficients associated with higher-order eigenfunctions receive a quadratic penalty in the likelihood, which becomes

$$\Lambda(h) = -\int_0^T \exp\left\{\sum_s \beta_s \tilde{x}_s(\tau)\right\} d\tau + \sum_{i=1}^{N} \sum_s \beta_s \tilde{x}_s(\tau_i) - \sum \lambda_s \beta_s^2 \qquad [4]$$

where the penalty λ_s is zero if the s^{th} term is associated with a linear or discrete term, and positive if the s^{th} term is associated with a high-order pseudospline eigenfunction.

In a generalized additive model, it is possible to mimic the effect of a smoother like 'lo' by choosing (i) appropriate shapes for higher-order eigenfunctions and (ii) an appropriate set of matching penalties. The more terms are included in the pseudospline (the higher *p* is in equation [3]), the more accurate the approximation. In practice purposes, *p*=7 is often fine for one-dimensional smoothers. By setting up artificial datasets with large numbers of equally-spaced observations, and noting how smoothers behave when applied to these datasets, we can derive appropriate eigenfunctions and penalties for the continuous time case.

Since there is no really compelling reason to use one type of smoother over another (e.g. 's' vs. 'lo'), it doesn't matter too much exactly what 'eigenfunctions' one uses, as long as they satisfy the properties of smoothness, roughly constant height peaks and troughs, and increasing wiggliness with higher order. As the order increases, the eigenvalue should fall from one towards zero along some kind of sigmoidal curve. More-or-less any 'pseudosmoother' constructed along these lines will have reasonable properties in terms of visual fit and predictive power. In fact, one of the appealing features of pseudosplines is the ability to design one's own smoother to give whatever behaviour is desired. For example, it is reasonable to assume *a priori* that different cetacean species will each have a preferred range of temperatures, and that density will drop off at both higher and lower temperatures. If we want to use a smooth term for water temperature in modelling sighting rate, we would be quite prepared to believe a dome-shaped response. It therefore makes sense not to penalize the second-order quadratic-looking eigenfunction at all, and only to penalize third and higher-order eigenfunctions.

A particularly useful type of pseudosmoother is the *cyclic smoother*. Platform-of-opportunity data including the JCD are compiled on a year-round basis, and there can be marked seasonal changes in sighting rate. It makes sense to include a smooth term for time-of-year, but standard smoothers do not enforce continuity across 1st January. An inelegant solution is to use two smoothers, one for cos*t* and one for sin*t*, but this causes difficulties in subsequent interpretation. A better solution is simply to build a cyclic smoother, in which the eigenfunctions are 1, cos*t*, sin*t*, cos2*t*, sin2*t*, etc., with eigenvalues something like $\{1,1,1,(1+0.9\lambda)^{-1},(1+0.9\lambda)^{-1},(1+0.7\lambda)^{-1},(1+0.7\lambda)^{-1},\cdots\}$. Simulated data can be used to see if this is giving sensible behaviour.

A1.3 Shrinkage

The penalized-likelihood framework of [4] makes for very easy incorporation of *shrunken factors*. This is useful when dealing with discrete covariates that have a large number of levels, with relatively few observations at each value: for example, the observer's name. Different observers can have very different sighting rates, and it is important to allow for this if possible; but if we allow a separate independently-estimated sighting rate for each observer, we will grossly overfit the entire dataset, and there may be serious confounding with other covariate effects. Particular problems occur when there are combinations of levels with no sightings at all.

A simple and effective solution is to include the discrete covariate as a set of dummy 0-1 variables, one for each level, but to add a penalty term $\sum \lambda' \beta^2$ summed over the coefficients β of the levels. This tends to 'pull in' all the estimated effects towards their common mean. Statistically, it corresponds to placing a (fixed) prior distribution on the variable 'observer effect'.

A1.4 Fitting the model

Fitting the model amounts to maximizing equation [4] over the vector β. The form of [4] makes it simple to calculate first and second derivatives:

$$\frac{\partial \Lambda}{\partial \beta_r} = -\int_0^T \tilde{x}_r(\tau) \exp\left\{\sum_s \beta_s \tilde{x}_s(\tau)\right\} d\tau + \sum_{i=1}^{N}{}_s \tilde{x}_r(\tau_i) - 2\lambda_r \beta_r$$
$$\frac{\partial^2 \Lambda}{\partial \beta_r \beta_s} = -\int_0^T \tilde{x}_r(\tau) \tilde{x}_s(\tau) \exp\left\{\sum_s \beta_s \tilde{x}_s(\tau)\right\} d\tau - 2\lambda_r \delta_{r=s\,r} \qquad [5]$$

Note that the second derivative is independent of the sightings.

These equations form the basis for a Newton-Raphson or quasi-Newton iteration. For starting values, we can simply take $\beta_1 = \ln(N/T)$, $\beta_i = 0 \; \forall \, i > 1$ where $\tilde{x}_1(\tau) \equiv 1$ (the 'grand mean'). The main issue is how to tackle the integration over pseudotime. This basically has to be done numerically, though there is no need to call in a full-blown numerical integration routine to get adequate accuracy. Simple schemes based on weighted sums are quite satisfactory and much quicker.

Once coefficients have been fitted, it is a simple matter to summarize covariate effects or make predictions of sighting rates under different conditions. Because the eigenfunctions of a covariate to be smoothed are defined throughout the continuous range of that covariate, there is no technical difficulty in interpolating at parts of the range where there happen to be no observations. The usual warnings apply, however, to extrapolation or to interpolation across large gaps in the range.

A1.5 Assessing uncertainty: the problem of patchiness

If sightings really did follow a Poisson process with the functional forms suggested, then uncertainty could be assessed via the parameter variance/covariance matrix from fitting [4]. In practice, though, sightings tend to be more patchy than the Poisson form implies, even after fitting covariates. There are several reasons why this is only to be expected; the statistical upshot, though, is that we cannot trust likelihood-based approximations for variance and interval estimation.

Fortunately, though, the point estimates can be shown to be reasonable even under conditions of patchiness. We can therefore try to develop a bootstrap to provide more robust assessment of uncertainty. The key is to ensure that the resamples mimic the pattern of patchiness seen in the real data. It can be shown that, in the Poisson process formulation, the *rate gap* between sightings

$$G_i = \int_{\tau_i}^{\tau_{i+1}} e^{\lambda(\tau)} \mathrm{d}\tau$$

is exponentially distributed with mean 1. We can construct a 'residual rate gap' by taking

$$\hat{G}_i = \int_{\tau_i}^{\tau_{i+1}} e^{\hat{\lambda}(\tau)} \mathrm{d}\tau$$

If sightings are patchy, the mean of $\hat{G}_i$ will still be 1 (by construction), but there will be more short gaps and more long gaps than expected for an exponential random variable. This suggests the form of the bootstrap:

1) Calculate the empirical distribution of rate gaps.

2) To generate a new sample, pick a rate gap $G^*_{(1)}$ at random, and set the first observation at the pseudotime $\tau^*_{(1)}$ that satisfies

$$\int_0^{\tau_{(1)}} e^{\hat{\lambda}(\tau)} \mathrm{d}\tau = G^*_{(1)}$$

3) Pick another rate gap $G^*_{(1)}$, and set the next observation at $\tau^*_{(1)}$ s.t.

$$\int_{\tau^*_{(1)}}^{\tau^*_{(2)}} e^{\hat{\lambda}(\tau)} \mathrm{d}\tau = G^*_{(2)}$$

4) Repeat until the chosen rate gap 'overflows' the total elapsed rate in the fitted model, which will be equal to *N* by construction.

This guarantees that the resamples will have at least the right 'first-order' patchiness.

It is also possible that there is 'higher-order patchiness', that is, that a shorter-than-expected gap tends to be followed by another shorter-than-expected gap. The methods described in Bravington (2000b) allow such autocorrelations in the gap sequence to be studied. We applied this approach to the JNCC data, and it revealed a considerable impact of patchiness on uncertainty in estimated parameters.

Although the technical description above provides details of the main part of the analysis, there are other issues that we also found needed to be considered in the application of this analytical approach to these platform of opportunity data.

Appendix 2

A2.1 Double platform estimation

When the detection of animals on the transect line is not certain, one of the most successful ways of estimating detection probability, and hence abundance, is to use two teams of observers simultaneously surveying the same region independently of one another. Borchers *et al* (1998) call this type of survey a 'mark-recapture line transect' (MLRT) survey. It can be viewed as an experiment in which each animal in the survey area corresponded to a trial with four possible outcomes; detection by observer 1, detection by observer 2, detection by both observers and detection by neither observer. The last outcome is unobserved. A set of covariates, one of which would typically be the perpendicular distance of the animal from the transect, is associated with each trial.

Let $p_i(x,\underline{q})$ be the probability that an animal is detected by observer *i* (*i* =1, 2), at $(x,\underline{q})$ where *x* is the perpendicular distance and $\underline{q}$ are the other covariates and let $p_.(x,\underline{q})$ be the probability that an animal is detected by at least one of the observers. When two observers are searching independently of one another and the probability of detecting an animal depends only on the $(x,\underline{q})$ associated with the animal, then

$$p_.(x,\underline{q}) = p_1(x,\underline{q}) + p_2(x,\underline{q}) - p_1(x,\underline{q}).p_2(x,\underline{q})$$

Let β' = ($\underline{\beta}_1, \underline{\beta}_2$) represent the parameters of the detection function, where $\underline{\beta}_i'$ = (β_{i0}, .., β_{iR_i}) represent the R_i parameters for observer *i*.

The MLRT survey method conducted in SCANS involved a form of double platform survey called 'BT mode' survey (after Buckland and Turnock 1992). In this survey mode, the role of observer 2 is to generate detections of animals before they have responded to the observers and in SCANS observer 2 (known as the tracker) was searching with binoculars, farther ahead of the ship than observer 1 (primary), who was searching with the naked eye. The estimation of the detection function for observer 1 is conditioned on these detections, which serve as a set of binary trials in which success corresponds to detection by observer 1, from which $\underline{\beta}_1$ is estimated. The $p_i(x,\underline{q})$ can be modelled as a logistic function of the covariates as shown below.

$$p_1(x,\underline{q}) = \operatorname{log}it^{-1}\left(\beta_{10} + \beta_{11}x + \sum_{r=1}^{R_1}\beta_{1(r+1)}q_r\right)$$

The perpendicular distances are those recorded by observer 2 since it is assumed that the animals will not have reacted to the ships presence at this stage. The set of S+ functions implemented in the package MRLT (Borchers and Burt 2001) was used to estimate $\underline{\beta}_1$.