Potential Use of Joint Cetacean Protocol Data for Determining Changes in Species' Range and Abundance: Exploratory Analysis of Southern Irish Sea Data

Len Thomas

Centre for Research into Ecological and Environmental Modelling

University of St Andrews

June 2008; revised January 2009

Report to Joint Nature Conservation Committee; National Parks and Wildlife Service; and Countryside Council for Wales

This work was funded by the Countryside Council for Wales and the National Parks and Wildlife Service. The contract was managed by the Joint Nature Conservation Committee

Comhshaol, Oidhreacht agus Rialtas Áitiúil
Environment, Heritage and Local Government

Cyngor Cefn Gwlad Cymru
Countryside Council for Wales

JOINT NATURE CONSERVATION COMMITTEE

JNCC Contract No: F90-01-1208

Foreword

Since publication of the Atlas of cetacean distribution in north-west European waters (Reid *et al.* 2003), many new data relating to cetacean distribution have been collected. The European Seabirds at Sea and Sea Watch databases have been augmented by data from opportunistic, generic surveys as well as systematic, geographically targeted ones. The SCANS (Small Cetacean Abundance in the North Sea and adjacent waters) survey was repeated over a more extensive area in 2005, and survey coverage further extended by the 2007 CODA (Cetacean Offshore Distribution and Abundance) survey. In addition to these, other European data-sets now exist that were not available for the Atlas and would enhance any future update.

There are several national and international instruments that would be served by an updated Atlas; However, perhaps the most appropriate legal driver is the EC Habitats Directive (HD; EEC 1992). All species of cetacean are included in Annex 4 of the Directive, which places an obligation on all European Union Member States to accord them strict protection. This carries with it the requirement to report on species conservation status every six years. Favourable conservation status (FCS) must be assessed with regard to four parameters: natural range, population size, habitat (extent and condition) and future prospects. While some flexibility is accorded to national experts, EC advice recommends the use of time series data to detect change in FCS, possibly in comparison against favourable reference values.

In 2006 a working group was established to update this cetacean data resource and produce a new atlas. It aims to achieve this through establishment of a Joint Cetacean Protocol (JCP) rather than a static database. The JCP will comprise standards for the integration of cetacean abundance and distribution data collected from European waters using a variety of methods. Data will be shared under a common agreement, ideally through a web-based portal demanding little maintenance, which would, if necessary, restrict access to data not in the public domain. This project has received support from ASCOBANS and a growing number of European governmental and non-governmental organisations.

As part of the initial phase of the project, exploratory analyses were commissioned by UK and Ireland on a subset of data from the southern Irish Sea, which were considered representative of the eventual JCP data resource. The aim of this pilot study was to determine how disparate data types might be integrated and what power the final data resource may have to detect trends in range and abundance. The results of this work are presented here.

Further Work

- The JCP working group has considered the findings of this report and aim to proceed with the project as follows:
- Clean and collate data-sets to be used for final analysis of the JCP data resource;
- Develop methods required to allow integrated analysis of all JCP data;
- Use this analysis to consider how data collection methods can be improved;

- Produce summary data-sets using best methods derived from the above analysis;
- Create agreement for sharing data between partners of the JCP project;
- Share data via an internet data portal.

JCP working group 21st October 2008

Index

Executive summary

The purpose of this work was to make a preliminary assessment of how Joint Cetacean Protocol data might be used to detect changes in abundance or range of UK and Ireland cetacean species. We review the monitoring objectives that arise from Article 11 of the EC Habitats Directive, and consider what measures might feasibly be monitored. We show that targets such as having high power to detect a 1% annual decline in abundance or range over a 6 year reporting period are not remotely feasible, and suggest that a 15-30% annual decline may be detectable over that period. Analysis of JCP data is difficult because the data are sparse, are collected over a range of spatial and temporal scales and often lack direct information about detectability. We review potential analysis methods, suggest methods for data integration and conduct an exploratory analysis of JCP datasets from the southern Irish Sea. More research is required on optimal methods for analysis. JCP data would be of much greater utility if more effort were devoted to implementing field methods that gather direct information about detectability, such as rigorous line transect protocols, even on platforms of opportunity. Land based surveys, currently part of JCP, likely have limited utility for UK and Ireland-level cetacean monitoring, except perhaps for very coastal species.

Main recommendations:

- Continue current efforts to define monitoring metrics and realistic targets.
- Allocate significant effort to developing and testing statistical methods for spatio-temporal trend estimation from JCP data.
- Implement rigorous line transect protocols on all JCP surveys where feasible, including dedicated marine mammal observers and double-platform methods. This includes platforms of opportunity and volunteer-based surveys.

Introduction

EU Member States have a legal obligation under Article 11 of the Habitats Directive to undertake surveillance of all cetacean species occurring in their waters to determine their "conservation status", and to report on this every 6 years. A species is in "favourable" conservation status if: "population dynamics data indicate that the species is maintaining itself on a long-term basis as a viable component of its natural habitats, the natural range of the species is neither being reduced nor is likely to be reduced in the foreseeable future, and there is, and will probably continue to be, a sufficiently large habitat to maintain its populations on a long-term basis." The exact measures reported are open to interpretation by Member States (see European Commission 2006), but the above guidance leads naturally to a focus on (1) trends in species' abundance; (2) changes in species' range; (3) designation and monitoring of suitable habitat; and (4) Future Prospects.

In the UK and Ireland, one cost-effective method for potentially addressing the first two of the above measures is via the Joint Cetacean Database and its prospective successor, the Joint Cetacean Protocol (JCP). The JCP is essentially a collection of survey data and related meta-data that have been gathered by various governmental and non-governmental organizations. It is being assembled by a group of interested parties called the JCP working group. Because the data come from a variety of sources, and much of it was not collected with the primary aim of determining large-scale trends in range and abundance, obtaining robust estimates is potentially very problematic. The primary aim of the short research contract reported here was to make a preliminary determination of what outputs might be possible given a concerted effort. This was achieved in part by an exploratory analysis of a subset of the JCP data, from the southern Irish Sea.

More precisely, the research objectives were:

1. With members of the JCP working group use an effort-related cetacean data set, from the southern Irish Sea, to determine how to integrate data types collected using different methods and from a range of spatial and temporal scales.
2. Determine whether these data have the power to detect changes in range and abundance, making recommendations where appropriate for ways to improve, if possible, their explanatory power.
3. Identify any further analyses that should be considered for determining how to detect trends in range and abundance in the final JCP data set
4. Report on the findings of the above study to the JCP working group.

The test data were delivered to CREEM during the first part of April 2008, and a workshop was held in St Andrews on 21-22nd April 2008 where JCP working group members presented information about each dataset including the survey goals, field methods, data collection and storage protocols and outputs to date. A summary of the data sources is given in Table 1 (figures and tables are at the end of this document). Data collection methods included dedicated line transect surveys that employed a random design for positioning transects, transect surveys from platforms of opportunity (with and without distances to detected objects, and from both large and small platforms), and land-based watches. At the workshop there was also discussion and refinement of the research objectives. (Minutes of that meeting are available from Tim Dunn, JNCC.) CREEM members then pursued the above objectives.

The remainder of this report is structured as follows. Section 0 contains a discussion of which monitoring objectives and metrics are implied by the EC Habitats Directive, and what may be feasible. Section 0 contains a discussion of potential statistical methods for analysis of JCP data. Section 0 reports a brief exploratory data analysis. Lastly, we give some conclusions and recommendations for further work in Section 0.

Monitoring Objectives and Metrics

It is beyond the scope of this report to provide a comprehensive investigation of exactly what objectives may be appropriate and practical for compliance with the EC Habitats Directive, and what metrics may therefore be used for cetaceans. Indeed, a separate effort is underway by JNCC to produce a UK surveillance strategy scoping document (Sónia Mendes, JNCC, pers. comm.). Nevertheless, some discussion here will be useful in highlighting what outputs may be required from an analysis of JCP data, and possibly in aiding the scoping effort.

Population dynamics data, population viability and temporal trend analysis.

The first criterion for determination of favourable conservation status is to determine whether "population dynamics data indicate that the species is maintaining itself on a long-term basis as a viable component of its natural habitats". Ideally, data would be available on both demography (survival, fecundity, etc.) and population size, and these could be integrated within a population viability analysis (Bessinger and McCulloch 2002; here assumed to include the more up-to-date versions using a stochastic population dynamics model fit to the data e.g., Buckland *et al.* 2007b). Such analyses could be used to determine the long-term viability of the species, perhaps under different possible future management scenarios. While such analyses are possible for some species in the UK, for example some species of songbird (Besbeas and Freeman 2006, Baillie *et al.* in press) and seabird (Crespin *et al.* 2006), little demographic data exists for cetacean populations in the UK and Ireland (and indeed for almost all cetacean populations worldwide). Nevertheless, useful progress can be made in this direction (e.g., Thompson *et al.* 2000, Berggren and Wade 2001).

Such population assessment via mechanistic modelling of the population dynamics, based on a knowledge of the species' biology, is in practice difficult, and by far the more common approach is to engage in empirical modelling of abundance data with the goal of determining past trends in species' abundance and using this as an indicator of population viability. Mechanistic vs. empirical modelling approaches are compared in, e.g., Thomas *et al.* (2004) and Berggren *et al.* (2008a, b), with the latter having a particular emphasis on cetacean monitoring in the EU. In the remainder of this report, we concentrate exclusively on the empirical approach.

There are many potential methods for determining population trends, in part because the definition of trend, "a long-term change in mean level", is intrinsically subjective – what is "long-term"? There are also many issues inherent in trend estimation that can cause undesired results. For example, for relatively short time periods, the state of the population at the start of the monitoring period will have a strong influence on the trend – a population that is at the bottom of a decadal population cycle will appear to have a strongly increasing trend for the first 10 years of monitoring. Another example is that if the trend is defined as average change in abundance over the monitoring period (i.e., slope parameter from a linear or log-linear regression model) then sudden catastrophic population declines become harder to spot the longer the monitoring goes on for. These, and other issues with trend estimation, are discussed in detail by, e.g., Thomas *et. al.* (2004) and Berggren *et al.* (2008a, b).

Issues arising from the use of relative abundance as a metric

An ideal metric for trend analysis is population abundance (or equivalently density) within the area of interest. However, many of the data collection methods in the JCP database do not lend themselves to estimation of absolute abundance or density (and even for those that do, there may be significant problems that mean such estimates are not reliable – see later). Therefore relative abundance may be considered for use instead.

The pitfalls of estimating trend based on relative, rather than absolute, abundance are well known (e.g., Anderson 2001, Pollock *et al.* 2002, Sauer and Link 2004, Nichols *et al.* in press). To summarize them requires some notation. Let N_s be the true abundance within the area of interest at time *t*, let R_t be the measure of relative abundance, and let β_t be the constant of proportionality linking the two (i.e., $R_t = N_t\beta_t$). Assume β_t is a random variable, with some mean μ_β and variance σ_β^2.

The first pitfall is that there may be a trend in the βs that is (positively or negatively) correlated with the trend in *N*s. If this is the case then the trend in relative abundances *R* will not accurately reflect the trend in abundance. It is easy to imagine how this can take place – for example, if animal sightability generally increases over time due to better observer training, then this will generate a misleading positive trend in the *R*s. This is just one example – there are many other reasons why animal sightability may increase or decrease over time. One partial solution is to standardize the data collection protocol as much as possible. However, despite rigorous standardization, trends in the constant of proportionality may still remain for unexplained reasons (e.g., Norvell *et al.* 2003). Also, the potential for standardization of survey methods across the datasets in the JCP may be limited. A second partial solution is to include covariates in the trend modelling that may explain variation in the constant of proportionality – e.g., observer, observer experience or observer training in the above example. The problem with this approach is that if the pattern in covariate values is correlated with the trend in absolute abundance, then including the covariate will mask the true trend. A significant limitation of this second solution is therefore that covariates should only be included if they are thought not to vary in the same way as the trend.

The second pitfall is that even when there is no trend in the constant of proportionality, its' value will vary from year to year (according to the magnitude of σ_β^2). This causes variation in the *R*s that is not related to variation in N, and therefore decreases the ability of the relative abundance measures to detect a trend. This issue is often ignored when determining the power of methods based on indices of abundance to detect trends, but is discussed in detail by Berggren *et al.* (2008b, especially in Adjunct 1).[1]

Despite these problems, it is often possible to make suitably qualified inferences about temporal patterns in relative abundance – examples include those from the North American Breeding Bird Survey (e.g., Sauer *et al.* 2007) and harbour porpoise sightings from the European Seabirds at Sea (ESAS) database (Winship 2008). The caveats should always be borne in mind, however, and trend estimation using absolute density or abundance measures (or as close to this as possible) should be preferred where possible. These issues are well rehearsed and debated in the literature – see, e.g., Anderson 2001

---

[1] It should be noted that the alternative of estimating absolute abundance is often no better in this respect, since the variance of estimates of absolute abundance given the true abundance at a time point (i.e., $\mathrm{var}(\hat{N}_t \mid N_t)$) may have nearly equivalent variance to the corresponding variance for the relative abundance index ($\mathrm{var}(\beta_t \hat{R}_t \mid N_t)$), because one component of the true abundance estimate is an estimation of sightability.

vs. Engerman 2003 vs. Anderson 2003 or Hutto and Young 2002 vs. Ellingson and Lukacs 2003 vs. Hutton and Young 2003.

Power to detect trends

Power analysis has become a standard tool in assessing the potential utility of planned monitoring programs (e.g., Berggren *et al.* 2008b): it allows us to address questions such as "What is the power to detect a 50% decline in (relative) abundance over 25 years", or "How much survey effort is required to have an 80% power to detect a 50% decline in (relative) abundance over 25 years". The concepts are discussed in many articles (e.g. Gerrodette 1987, Steidl and Thomas 2001, Thomas *et al.* 2004) and user-friendly software exists for performing simple power analysis (e.g., Gerrodette and Brandon 2003).

Here, for simplicity, we will assume that trend is to be estimated using log-linear least-squares regression on relative abundance indices, and that the criterion for assessing trend is a two-tailed significance test on the slope. We further assume that the coefficient of variation of relative abundance is independent of the estimate (see Gerrodette, 1987, for some alternatives). To perform an analysis of the power in this circumstance we are required to specify four out of five of the following: number of time points (assumed equally spaced, with an observation at each one), rate of population change (i.e., regression slope), residual coefficient of variation, alpha level for significance test and power. If any four are specified, the fifth can be calculated.

The most problematic quantity in the above is the residual CV – i.e., the variability in relative abundance indices not explained by the trend line. It has potentially three components: (1) sampling variation in the relative abundance indices, given some fixed true abundance and constant of proportionality; (2) variation in the constant of proportionality (we already referred to the corresponding variance as σ_{β}^{2}); and (3) residual lack of fit between the trend line and the true annual abundances. Assuming these sources of variation are independent, they can be combined using the delta method (CV of a set of independent random variables multiplied together is square root of the summed squared CVs); other assumptions are also possible (e.g., Berggren *et al.* 2008b). Many power analyses mistakenly only consider the first source of variation in making their calculations – e.g., Gibbs *et al.* 1998. Others (e.g., Berggren *et al.* 2008b) ignore the third source of variation, although this is valid if true abundance is treated as a fixed rather than random quantity – see Thomas *et al.* (2004, section 5.2.5) for a discussion of this point.

The general outcome of such analyses is that the monitoring goals initially specified when designing surveillance programs are unrealistic. To show an example of this, we need to obtain a realistic set of CVs to work with. Winship (2008) estimated relative indices of harbour porpoise abundance for seven subsets of ESAS data and obtained CVs for component (1) above (i.e., just sampling variation) in the range 0.3 to 1.2. Berggren *et al.* (2008b) estimated combined CVs for components (1) and (2) above using the SCANS II survey to be in the range 0.2-0.6 for an ESAS-like survey method and 0.7-0.9 for a towed acoustic method, depending on assumptions about independence between components (1) and (2) (using their approach 1 – see paper for details). In general, it is hard to imagine obtaining a residual CV (i.e., components (1), (2) and (3) above) of less than 0.2 from analysis of future JCP data. Conversely, it is easy to imagine that such a CV may be 1.0, or higher.

The Habitats Directive reporting interval is 6 years. We therefore calculated the power to detect a range of annual population declines (or increases) over 6 years, given CVs in the range 0.2-1.0, and assuming an alpha-level of 0.1 (which is generally considered liberal). The results are summarized in Figure 1. For example, imagine the target power of a monitoring program was 0.8. If the program achieved the best conceivable CV (0.2), then an annual rate of decline of 0.15 is required for that level of power. Note that a decline of this rate over 6 years corresponds to the loss of approximately 60% of the population[2]. If the CV was 0.4, an annual rate of decline of 0.33 would be required for a power of 0.8; this corresponds to the loss of approximately 90% of the population over 6 years.

It has been suggested (European Commission 2006, p. 43) that to comply with the EC Habitats Directive, one should be able to determine if each species has undergone a large decline in population size, defined as a loss of more than 1% (i.e., 0.01) per year. Given the above figures, this is clearly not feasible. Indeed, given annual monitoring, it would require 32 years to have a power of 0.8 to detect a decline of 1% per year given a residual CV of 0.2; the corresponding figure with a (more realistic, but still optimistic for JCP data) CV of 0.4 is 49 years. The total population change over this long time period may also not be one considered of biological significance: a decline of 1% per year for 32 years corresponds to the loss of approximately 30% of the population, and over 49 years to around 40%. Clearly, a re-think of the quantitative targets for monitoring is required.

Changes in natural range: spatio-temporal trend analysis

A second criterion for determination of favourable conservation status is that "the natural range of the species is neither being reduced nor is likely to be reduced in the foreseeable future." A species' range is a difficult thing to quantify precisely, especially for taxa like cetaceans that may be highly mobile. Unusual sightings of many species have occurred far outside what is considered to be their core range. Many cetacean species are also known to undergo seasonal migrations. However, without an unambiguous definition of range, it is not possible to quantify whether range is being reduced.

There are many potential range metrics. One is a definition in terms of probability of presence – for example, a species range may be defined as the polygon enclosing all points where the instantaneous probability of presence in a 1km^2 square centred on that point is greater than a given value. A second definition is in terms of density – range is a polygon enclosing all points where the average animal density (averaged over a defined time window) is greater than some given value. For species known to migrate, either of the above definitions could include a seasonal component.

[2] Calculated using $1-(1-0.15)^6 \times 100$

Given estimates of probability of presence or density at defined spatial locations and times, estimates of range and changes in range over time can be produced via spatio-temporal analyses – i.e., modelling the presence or density data with covariates that are spatially and temporally referenced. In the case of JCP data, however, we are not able to estimate absolute probability of presence or density due to limitations in the data collection methods for many of the composite datasets. We are therefore not able to estimate "absolute" range[3], and any spatio-temporal modelling to changes in range will be restricted to "relative" measures of range, such as changes in encounter rates per unit time and area. The issues here are similar to those that arose in assessing trends in relative abundance.

We note that the use of relative (rather than absolute) measures of range is nearly ubiquitous for other taxa, often in the form of atlas schemes based on standardized survey protocols (e.g., Asher *et al.* 2001, Gibbons *et al.* 1993). One difference between these schemes and the JCP is that the former use a single (or small number) of standardized protocols, while the JCP comes from many diverse surveys. Nevertheless, the monitoring of "relative" range appears to be quite acceptable for other taxa.

Habitat assessment

The third criterion, for determination of favourable habitat status, is that "there is, and will probably continue to be, a sufficiently large habitat to maintain its populations on a long-term basis". Research into determination of what comprises "habitat" for free ranging species is an enormous topic, encompassing many branches of ecology (see, e.g., review in Aarts 2007). Determining habitat extent will be particularly difficult for most cetacean species if the definition of habitat is taken to include places where there are food stocks, since robust information about diet is lacking for most cetacean species, as is information about the presence of potential food such as fish stocks (see, e.g., Torres *et al.* in press). The importance of other factors that potentially affect habitat suitability (under some definitions of "habitat"), such as level of ocean noise, are also unknown.

A practical approach, at least for species that are not central-place foragers, is to define habitat empirically, in terms of measurable biologically-relevant and temporally-stable environmental variables that correlate with higher animal density. This reduces identification of habitat to a model selection exercise, given suitable data on animal density and potential covariates. Notwithstanding the caveats given in previous sections about the problems of using relative density and abundance measures in place of absolute measures, it may be possible to include identification of species' habitat preferences within the same modelling framework as the spatial or spatio-temporal models of the previous sections. Note, however, that the identification of potential candidate covariates will be different depending on the analysis goal: for spatial or spatio-temporal modelling to look at range and changes in range, it is quite acceptable to include explicit space covariates (such as latitude and longitude); for identification of suitable habitat, the

[3] "Absolute" does not imply here total range, since the total range of many species exceeds the survey boundaries of UK and Ireland waters. Instead, "absolute" is in terms of the previous definitions – e.g., absolute probability of presence at defined locations.

inclusion of such covariates would make little sense unless to highlight the absence of a key biological covariate determining animal density patterns.

A framework for Spatio-Temporal Trend Estimation from JCP Data

In this section we consider statistical models that may be appropriate for the analysis of JCP data, with metrics from the previous section in mind. We begin with a general formulation, focussed on the context of line transect sampling where the formulation was initially developed. We discuss various potential approaches for modelling line transect data, before considering how one might perform similar analysis on transect data without distances and land-based watch data. Lastly, we consider methods for joint analysis of multiple datasets. We assume throughout that each species will be treated separately, although joint analyses of multiple species is a conceivable extension to these methods. Due to time limitations, we largely ignore the issue of species that occur in compact groups, although extensions to these methods can be (and have been) used for this situation (e.g., Bravington *et al.* 2002).

A more comprehensive discussion of the various approaches recently developed for spatial and spatio-temporal analysis of line transect data, arising from a recent technical workshop on this subject held in St Andrews, is in preparation (Thomas and Scott-Hayward, in prep.).

General formulation – line transect context

One general approach for considering the random processes generating observations is that of a spatial point process (e.g., Schweder 1974, Hedley 2000). Assuming independence between animals, their locations in space and time is a realization of an inhomogeneous Poisson process with intensity $D(x, y, t)$, where x and y are spatial coordinates and t is time (the time index is intended to account for seasonal and annual changes in intensity, rather than temporal changes at the level of a survey boat moving along a transect; temporal models are not considered in the references cited in this section). Under this definition, the number of cetaceans in an area a is a Poisson random variable, with mean

$$\mu(a,t) = \int_{(x,y)\in a} D(x,y,t)\,dxdy \,. \tag{1}$$

This formulation can be extended in various ways. For example, to allow for cetacean species that occur in compact groups, one could use a marked spatial point process model where the mark is the group size (e.g., Hedley 2000). To relax the assumption of independence of locations of cetaceans (or cetacean groups) and allow for effects such as clumping or repulsion, one could consider more general models such as Neyman-Scott models (e.g., Waagepetersen and Schweder 2006) or Markov-modulated Poisson models (Skaug 2006, Bravington 2008).

In a line transect survey, transect lines are traversed by an observer (or observers), and the location of detected animals recorded. Not all animals within each survey segment are detected. Under the assumption that detections are independent random events, the number of animals recorded in a segment a can be thought of as the outcome of a thinning of the point process. A thinned point process is also a Poisson process, so the number of animals recorded in segment *a* is a Poisson random variable with mean

$$\mu_d(a,t) = \int_{(x,y)\in a} D(x,y,t)\, g(\theta)\, dxdy$$

where $g(\theta)$ is the probability of detecting an animal given covariates θ, which may include spatial location (perhaps relative to the transect line) and other factors of importance such as observer, sea state, etc. Thus, given parametric models for *D*() and *g*(), a likelihood can be constructed for the location of detected animals (e.g., Hedley 2000; related approaches are taken by Royle *et al.* 2004 and Buckland *et al.* 2007a).

Two-stage approaches to modelling density in line transect surveys

In practice, a two-stage approach is usually taken, where the detection function *g*() is fit first, and then the density function *D*() is fit conditional on the estimated detection function parameters. This is because probability of detection is sensibly thought of (and parameterized) as being a function of local effects such as distance of the animal from the transect line, while variation in density occurs at much larger spatial scales. Hence there is little information shared among the two components, and therefore little loss in separating them. By dividing the problem into two pieces it becomes conceptually and practically easier. This is discussed further by Hedley (2000), and a good example of the small loss in information caused by implementing a two-stage approach is given in Royle *et al.* (2004), where the detection probability estimates are nearly identical, within nearly identical precision, in a one-stage and two-stage approach.

The two-stage approaches can be broadly divided into two sub-groups, depending on the response variable modelled in the density function estimation stage (Hedley 2000, Hedley and Buckland 2004, Hedley *et al.* 2004). The first approaches ("count methods") are based on modelling the counts of numbers of animals (or groups) detected in segments of transect line. The second ("waiting times/distances/areas methods") are based on modelling the times, along-trackline distances or areas surveyed between successive detections. We briefly discuss these below.

In count methods, the transect lines are divided into segments (or legs) where covariate values affecting density and detectability (apart from animal-level covariates such as group size) are judged to be constant. The response variable, the number of animals (or groups) detected per segment, is then modelled as a function of covariates judged to affect density. There are several ways to take detectability into account; a common method is to include the area effectively searched (i.e., transect length times width of the transect strip times average detection probability) as an offset term in the model. Such methods are discussed in detail in Hedley (2000), Hedley and Buckland (2004), Hedley *et al.* (2004) and are available in a test version of the software Distance (Thomas *et al.* 2008). An advantage of count methods are that they are generally relatively easy to apply, since the model can be formulated as a standard generalized linear or additive model (GLM/GAM, with the counts assumed to have a Poisson distribution, for example), for which robust general fitting software is available. A significant disadvantage is that dividing up transect lines into a large number of small segments (so that covariate values can be assumed constant) usually results in a large number of segments with zero counts, so that standard GLM/GAM models provide a very poor fit. Relatedly, there is often residual autocorrelation after modelling, so that the standard assumption of independence between data values does not hold, and model selection and variance estimation are compromised. Again relatedly, the determination of segment size is arbitrary and can affect the results. Various solutions have been proposed, including zero-inflated models (for dealing with the excess zeros), mixed effects models (for dealing with the residual autocorrelation), Markov-modulated models (for dealing with the same issue via the model rather than error term), parametric bootstrap (for better variance estimation); these are reviewed by Thomas and Scott-Hayward (in prep.).

Waiting time/distance/area methods are related to time-to-event methods in other areas of statistics, such as survival analysis in industrial and medical statistics. Here the response variable is the time (e.g., Bravington *et al.* 2002) or the along-trackline distance (e.g., Skaug 2006) to next detection, which is modelled as a function of the covariates thought to affect density. Detectability can be incorporated in various ways; for example Hedley (2000) modelled the area effectively searched before next detection as the response variable. Variation in density between detections can also be accounted for (see previous 3 references for examples). The major advantage of these methods is that there is no need to divide the trackline into arbitrary segments, and so it is much easier to construct models that provide an adequate fit to data. However, unexplained local clustering in sightings is still a common issue, leading to implementation of methods such as Markov-modulated models (Skaug 2006) to deal with it. A disadvantage is that analysis is not easily performed using standard software, once variation in density between detections is accounted for, making them less accessible at present to practitioners. An example of the application of these methods to potential JCP data is Henrys (2005), who analyzed line transect data collected from two ferry routes by the Biscay Dolphin Research Programme.

Various other issues have been identified in the spatial and spatio-temporal density modelling of line transect data, and solutions are an active area of current research (see review by Thomas and Scott-Hayward in prep). The appropriateness of biologically-motivated covariates (such as habitat covariates) versus covariates of convenience such as latitude and longitude is an issue that has already been mentioned. Semi-parametric methods based on smoothing are often used in modelling covariates; there are many such methods and their relative merits is a matter of debate. Data-based methods for selection of the amount of smoothing is an issue, as is the tendency of standard smoothing methods to extrapolate wildly into un-surveyed regions and to cope poorly with complex topography (one solution to these is proposed by Wood *et al.* 2008). For the newer methods that allow small-scale local variation as well as large-scale smooth patterns (e.g., Skaug 2006), an issue is how to separate these two scales, and when it is appropriate to exclude the small-scale variability from variance estimation when making inferences. Lastly, appropriate methods for modelling spatial (and spatio-temporal) variation in group size, and for cascading uncertainty through the various stages of modelling to produce valid overall variance estimates, are issues.

Other two-stage approaches exist that lie outside the dichotomy outlined above. For example, Kaimi (in prep.) has developed a Bayesian approach to spatial density surface modelling from line transect data, based on tessellations.

Relevance to transect data without distances

Most of the methodology discussed in the previous section is applicable to the situation where distance data are not available, or are not of sufficient reliability for probability of detection for each observation to be estimated using distance sampling methods. Clearly, however, explicit corrections for detection probability cannot be made – for example through the use of area effectively surveyed as an offset in the count method. Instead, segment length must be used (or area surveyed if a fixed truncation width was used), and the potential confounding between density changes and changes in detectability can be addressed, as much as possible, through the use of covariates. As noted earlier, care must be taken when trying to remove effects of variation in detectability, not to use covariates that might be correlated with density changes – otherwise changes in density will be lost in the modelling.

An example of a waiting times-based method applied to relative abundance data (ESAS data for harbour porpoise) is Bravington *et al.* (2002), although their goal was identification of spatial hotspots rather than long-term temporal trends. A correction for detectability differences induced by school size and sea state was used, based on a separate analysis of SCANS data. In addition, sea state (again) and vessel speed were used as detection related covariates. Random effects-based methods for dealing with potentially important covariates such as observer and vessel that have many factor levels were suggested.

An example of a count-based method applied to relative abundance data is given by Winship (2008), who modelled spatio-temporal patterns in harbour porpoise encounter rates using 7 subsets of ESAS data, chosen to be as internally homogeneous as possible with respect to factors affecting detectability, and to limit potential confounding between such factors and variation in density. Factors shown to impact on encounter rates, and unlikely to be related to variation in density included sea state, visibility, time of day, observer and vessel. A number of strategies were used to cope with various issues in the data: the response variable was assumed negative binomial (to deal with over-dispersion); some nuisance covariates were modelled as random effects (as they were present in the data at many levels); diagnostics included checking for autocorrelation in the residuals.

Relevance to land-based watches

Similar methods can be used to analyze temporal patterns in encounter rates in land-based watches. However, even if positional data are collected on sightings, it is not straightforward to use this to estimate detection probability using distance sampling methods, as a fundamental assumption is violated: that the distribution of true animal density with respect to distance is known[4]. Protocols that require two independent

[4] This distribution is known in standard distance sampling applications because a large number of point or line transects are placed in the study area using a random survey design – see Buckland *et al.* (2001).

observers to operate, or supplemental data on animal distribution from a separate survey, can be used to circumvent this problem, but such advanced methods are rarely used in practice in land-based watches. A more feasible approach may be to use data truncation to restrict the survey area to one where it is assumed no animals are missed. An analytically complex alternative that may be feasible for the Cardigan Bay data is estimate the distribution of animal density with respect to distance from the watch points using data from the independent line transect surveys – see Marques (2007) for an example of this type of analysis.

Multiple land-based watches could potentially be used to provide spatio-temporal information, but there are rarely enough spatial locations to enable spatial modelling methods of the types described in previous sections to be used.

One potential issue when comparing absolute or relative density estimates from land-based watches with those from transects is that of time. Distance sampling methods, including those that take place at points, are conceived as taking a "snapshot" of the density at an instant in time. When the observer is stationary at a point, clearly the longer the watch period, the greater the number of animals that will be observed, due to animal movement. Hence to make fixed point surveys comparable with those from transects, either a snapshot approach must be implemented (where the data are the location of animals known to be in the monitored area at a set of instants in time), or other approaches based on instantaneous cues (such as surfacings) used (see Buckland 2006 for a discussion of this issue in the context of point surveys for birds). If naïve encounter rate measures are used, such as number of detections per 15 minute interval, then land-based methods may be used to supplement temporal trend information from transect surveys, but their relative encounter rates will always be different and this will need to be accounted for in any joint analysis (see section 0)[5]. Possible ways to implement snapshot-type methods for land-based watches include so-called "scan sampling", where regular scans are made and the locations of animals in view are recorded, as well as methods where accurate records are kept of each sighting event; there are many caveats with each approach and these are discussed (in the bird context) by Buckland 2006.

Methods for integrating data from multiple sources

JCP data come from a variety of survey types: designed surveys (where survey effort is located with an element of randomization) vs. platforms of opportunity; transect surveys with and without a rigorous distance sampling field protocol, and land-based watches. There are (at least) seven approaches for integrated spatial or spatio-temporal modelling of these data:

1. Use only designed line-transect data sets, and perform a standard design-based distance sampling analysis treating each dataset as a stratum. This is clearly not tenable given current JCP data, as it will involve discarding the majority of the surveys and leave too little to make reliable inferences about spatial or temporal

[5] Note that animal movement can be a problem for transect-based surveys too, if vessel speed is slow relative to animal speed – see Buckland *et al.* (2001) for a discussion of the issues and DiTraglia (2007) for some initial steps towards methods for addressing it. The scale of the potential problem will vary from species to species, depending on their speed and reaction to vessels.

patterns. Also, it would be important to assess the extent to which each line-transect data set is capable of producing an unbiased estimate of density, or, more generally, how potential biases may vary among surveys. For example, variation in vessel speed and noise between surveys mean that there is a varying effect of animal movement. Double-platform survey methods have the potential to deal with responsive movement issues (in trial mode, see Laake and Borchers 2004), but rely on at least one platform being high enough to detect animals before movement. Even non-responsive movement will bias estimates when vessels move relatively slowly compared to the rate of animal movement (discussed in Buckland *et al.* 2001).

2. Use all line-transect data (designed and platform of opportunity) in a distance sampling model-based analysis of spatial and spatio-temporal trends in absolute density or abundance, using one (or more) of the methods described in sections 0 and 0. This is also unlikely to be tenable due to loss of data, for historical surveys at least, although if future data collection protocols are modified it may be possible in the future (see section 0, Conclusions, below).
3. Use all line transect data, plus "comparable" transect data that does not have distances, in a two-stage analysis. Firstly, a distance sampling analysis of the line transect data is performed to estimate detection probability and covariates affecting detectability (such as sea state, vessel height and speed, group size, etc.). This analysis is then used to "correct" both the line transect and the transect data that does not have distances (e.g., by calculating effective strip width given covariate values in the non distance surveys, etc.). A model-based analysis as described in Section 0 is then performed. Determining which parts of the data are "comparable" can potentially be problematic – for example this type of analysis was performed on ESAS data by Bravington *et al.* (2002), using SCANS data to correct ESAS data for detectability – although sea state was still required as a further correction for the ESAS data, and the SCANS data were not used at the density surface modelling stage. This approach deserves further investigation.
4. A more integrated, single stage, approach to the previous method could be envisaged, in which the detection function is fit to the line transect data at the same time as the spatial or spatio-temporal density surface model is fit. There is potentially information in the change in encounter rate with different covariate conditions in the non-distance data that could be used to help fit the detection function. Again, this kind of approach deserves further attention.
5. Use all transect data, whether distances were collected or not, and ignore the distance data from the line transect portion, instead modelling relative density or abundance. Instead, include covariates in the encounter rate modelling that are common to all datasets to attempt to deal with detection issues. This is a multi-dataset extension of the approach described in section 0, as implemented, e.g., by Winship (2008). It does not make use of the distance sampling data to estimate detectability, but avoids the need to discard non-comparable data as was required by the previous method. This is another approach that requires further investigation.
6. Use the results from separate analyses of all possible datasets in a meta-analysis of absolute or relative density or abundance. The methodology is discussed

further in section 0, below. The conclusion in that section is that it does not make best use of the JCP dataset.

7. Include all available data, including that from land-based watches, in a joint analysis. This is discussed further in section 0, below.

Meta-analysis of survey results

In this approach, comparable estimates that are the outputs from different surveys are jointly modelled to produce a composite estimate of spatial and/or temporal trend. For example, imagine there are multiple estimates of a species' density, made by various survey methods over a range of time periods and subsets of the area of interest. The density estimates could be modelled in a GLM/GAM framework (likely as gamma-distributed random variables), with covariates including (smooths of) time and space (e.g., centre of area studied by each survey). The number and spatio-temporal coverage of the survey estimates will limit the amount of flexibility possible for the covariates. If there were thought to be biases in the density estimates, then additional covariates thought to affect bias, such as survey method, survey agency, etc, could be included. Such an approach is also feasible with relative density (e.g., encounter rate) as a response for surveys that cannot estimate density.

It would be sensible in such an analysis to weight the individual survey estimates by a measure of precision, so that more precise estimates get a higher weighting in the analysis, and by area studied, so that estimates covering a larger part of the overall area get a higher weighting. The latter could also be achieved by modelling estimated abundance and using studied survey area as an offset.

Using the centre of each studied survey area as a covariate will clearly result in a loss of information about the shape of the area in each survey. A possible extension, would be into a hierarchical modelling framework where, for example, expected abundance for each survey was modelled using a formulation like that of equation (1) and estimated abundance is related to true abundance via a second stochastic process. For example, if N_s denotes the true abundance in the region studied by survey s with region area a_s, and $\hat{N}_s$ denotes the survey estimate, then a potential model is

$$
\begin{aligned}
N_s &\sim Poisson\left(\int_{(x,y)\in a_s} D(x,y,t_s)\,dxdy \right) \\
\hat{N}_s &\sim gamma(N_s, \sigma_s^2)
\end{aligned}
$$

(lognormal, rather than gamma, would be another potential distribution for the second level). Such models could not be fitted with standard GLM/GAM tools. These ideas may warrant further investigation; nevertheless, the spatial aggregation required to produce per-survey estimates (even if results are stratified into regions) will still inevitably result in a loss of information relative to working with the raw sightings data. Since raw data are available within the JCP we do not pursue this method further here.

Analysis of raw sightings data from all data sources

As discussed already, although all component surveys of the JCP contain data that are spatially referenced, different components naturally produce different output measures.

Surveys that implement a line transect protocol are potentially capable of producing absolute density estimates, at least along the transect lines surveyed (in the case of platforms of opportunity). Transect surveys that do not rigorously implement line transect protocol, or do not collect distance information at all, can produce encounter rate per unit transect length statistics. As currently implemented, the land-based watches provide encounter rate per unit time (15 minute intervals for the Cardigan Bay watch data).

Jointly maximizing over separate sub-models

One option for an integrated analysis is to model each data source using whatever measure and detectability-related covariates that are most natural for it, but to include in all models the same spatio-temporal covariates, and then jointly maximize the likelihood over all models. For example, line transect data could be modelled by assuming that counts follow a Poisson distribution and using effective area as an offset; non-distance transect data could be modelled by again assuming counts follow a Poisson distribution but using line segment length as an offset and including covariates such as observer, sea state and platform in the sub-model for those data; land-based watch data model count per 15-minutes as a Poisson variable with sea state and observer. In addition, all three types of models would include the same spatio-temporal covariates: for example assuming a linear trend through time (log-linear if a log link function is used) and a low-dimension spatial smoother for spatial pattern. A likelihood could be derived for each sub-model, and this would be jointly maximized. This could not be done using standard GLM/GAM tools (because there are multiple models), making the best approach a statistical research problem.

Integrated analyses where multiple data types are jointly fit are common in other branches of statistical ecology, and an issue that can arise is the need to scale different components of the likelihood if the input measures are very heterogeneous (e.g., Maunder and Langley 2004). This is something that would need to be investigated, but could particularly be a problem if a different measure were used for some survey types – for example if waiting time between detections (see Section 0) were used for the line transect surveys but counts per 15 minute watch period were used for the land-based data. Without scaling, some surveys could receive undue weight in the joint maximization just due to the nature of the input measure.

Joint modelling using a common measure

An alternative approach for integrated analysis is to search for a common measure that could be used for all datasets. The advantages of this approach is that it may be possible still to use standard statistical tools such as GLM/GAM; a disadvantage is that some compromises or inefficiencies will inevitably arise, as will become clear.

One potential common measure is the waiting time to next detection. However, this is not recorded in a few surveys, and is recorded in discrete time in others (e.g., the land-based watches). It is also hard to know how to reconcile land-based watch waiting times with those from shipboard surveys without invoking some model of animal movement, which would considerably add to the complication (although some accounting for movement may be needed in any case; see earlier note).

A more natural common measure may be the count per unit space and time. Modelling would then be on the counts, with the amount of space and time that each count record was collected in used as an offset. However, transect surveys are usually regarded as “snapshots” in time, which means that animal movement does not have to be considered (although it is sometimes a problem in practice, leading to unreliable density estimates in line transect surveys – see earlier note). Rather than get into issues of modelling movement (if it can be avoided), it may be simpler to use a measure of count per unit space, and to consider the 15-minute watch periods for the land-based surveys as a snapshot. The offset would then be area effectively searched for line transect data, distance along transect line for non-distance transect data, and possibly area searched for watch data (if some data outside this area could be truncated). Note that these offsets are different for the different data sources, which means that, at the least, the model would have to include a factor covariate for “survey type” that reflects the different meanings of the counts for each survey type. Also, nuisance covariates (such as sea state, etc) would need to be included for the non-distance sample surveys, but the factor levels could be set to be all the same for the distance sampling data, as such things will have already been accounted for in the effective area offset.

It is not clear at this stage whether such an approach has merit and is workable in practice. This is a suggested area of further research.

In all of the above cases, inclusion of land-based watch data at 15 minute intervals is likely to cause significant temporal autocorrelation in residuals, unless a model that allows small-scale (perhaps hourly, daily or weekly) temporal clustering in sightings is included. This will be another topic for further research.

Exploratory Analyses of Irish Sea Data

Table 1 gives a summary of the data sources discussed at the April workshops. Data were not available for two data sources, and time limitations prevented us from requesting the missing data. These are source C (Sea Watch *ad libitum* line transect surveys in Cardigan Bay) and E (ESAS data) – for the latter sightings information but no effort data were provided. The absence of these data did not affect our conclusions as time for exploratory analysis was extremely limited. We also note that other datasets are potentially available for the study area; nevertheless, we make some conclusions based on the data available for this analysis.

Figure 2 shows the approximate spatial extent of the datasets (including C and E). A few things are immediately apparent:

- While some areas (e.g., Cardigan Bay SAC) are very well covered by survey effort, the effort in other areas is very sparse, or non-existent.
    - There is no survey effort in the current dataset in the south-east part of the study area, between Pembroke and Lundy Island (around 2,300km^2; for comparison, the whole study area is approximately 28,750km^2). Similarly,

there is almost no effort in the north-east part, east of Anglesey (around 2,000 km^2). Any estimates on animal populations into these areas would necessarily be extrapolations.
    - There is also no effort in the middle portion of the study area (from Holyhead down past Bardsey Island and east, around 5,000km^2). Estimates of animal populations here would necessarily be interpolation.
    - In general, effort off the Irish coast between Dublin and Rosslare is very sparse.
    - We note, however, that other data may be available to fill in these gaps.
- The area covered by the land-based watches is insignificant in comparison with the size of the study area, and this area is also covered by other data sources (although clearly not with the same level of temporal resolution).

Survey effort for each of the available datasets is shown by year in Table 2. The table indicates that there is some overlap in time between all the surveys, raising hope that it may be possible to disentangle true changes in density from survey-specific variation in sighting rates. However, there is, for example, almost no spatio-temporal overlap between the St George's channel (D) and ferry (F) data, and the Cardigan Bay (A and B) and Sulaire (G) surveys: while there is some Sulaire survey effort outside of Cardigan Bay, very little of it occurs in 2003-2007 which is when the St George's channel survey and the bulk of the Ferry data were collected. This type of lack of overlap is a common feature of these type of data (e.g., Bravington *et al.* 2002, Winship 2008), and can severely limit inference. A full account of all available data for this area would be needed before the potential level of confounding could be assessed in this area.

Nevertheless, for illustrative purposes, we attempted a very simple modelling exercise on encounter rate, using just the data for which we could conveniently obtain spatially referenced effort segments, segment length, and counts of cetaceans detected (all species combined) per segment. These datasets were the Sea Watch dedicated surveys in Cardigan Bay (B) and St George's Channel (D), the IWDG ferry surveys (F), and the Sulaire surveys (G). The years used for each survey are indicated in bold face in Table 2; others could potentially have been included as well, as could additional data sources, but lack of time precluded this. Our expectation is that the annual and large-scale spatial estimates would have been almost identical with inclusion of the land-based watch data, but may have changed somewhat with the inclusion of ferry and ESAS data.

Initial data screening rejected effort legs that were clearly incorrectly positioned (e.g., on land), or had unfeasibly long or short lengths, or that could not be matched with sighting data (for example because time stamps were missing) – these only accounted for a few percent of the data.

The models fitted were generalized additive models (GAMs), using the mgcv package within R version 2.8.1. A Poisson response was assumed, and a log link function used. The response variable was cetacean sightings (all species pooled) per segment and segment length was used as an offset. A tensor product smooth of latitude and longitude (both at the mid-point of each segment) was included in all models. Survey dataset (B, D, F or G) was included as a factor covariate. Additionally, models with a linear year term, a smooth year term (thin plate spline), and no year term were tried. Degrees of freedom for all smooths were estimated by cross validation.

All models fit the data quite poorly, with very little of the deviance explained (<10%), and a poor residual distribution due, largely to the many zeros in the data. The survey covariate appeared to have a significant effect, with the St George's Channel and Sulaire surveys having a slightly higher encounter rate than the Cardigan Bay survey, and the ferry surveys having a significantly lower encounter rate. Note that this could either reflect spatio-temporal variation in density, or a true effect of the surveys. The model that included a smooth year term had the lowest UBRE score (0.186), closely followed by that with the linear smooth term (0.197), and the model with no year term last (0.222). The estimated year smooth term is shown in Figure 3, and the spatial smooth in Figure 4. These results are shown merely for illustration – the model is too naïve to warrant serious interpretation. The bounds for predictions from the spatial smooth in Figure 4 are given by a minimum bounding rectangle on the survey points, intersected with the study area boundary. Nevertheless, the predictions are extrapolations in both southern corners of the rectangle, and based on very few data along the northern edge. Predicted encounter rate in the south-west is quite high, and this is almost certainly spurious, as is the hotspot in the north-west. Similarly, the much higher encounter rate predicted for the south-east is likely due to leakage from the Cardigan Bay area. Better smoothers that address both extrapolation and leakage problems have been developed by Wood *et al.* (2008). Overall, the only high encounter rate area that seems well supported by abundant data is that in the St George's Channel area, to the south of the study area.

Despite the clear problems, we have shown that the datasets can be combined and analyzed together. Obvious next steps in the short term are to include the other datasets, split the encounter rates up by species, use better smoothing methods and investigate the potential for inclusion of covariates such as sea state, vessel speed, platform height and observer (perhaps as a random effect) that are known to affect sightability and are collected on all surveys, as well as covariates such as season that are known to affect density.

Conclusions

Monitoring objectives and metrics

- In the absence of any demographic data, count data may be able to yield appropriate metrics for compliance with the EC Habitats Directive. There is a need for Member States to continue their consideration of which metrics are to be used for monitoring.
- It is premature to discuss the exact size of trends that may be detectable, since the metrics are not yet defined and the required variance components unknown. Nevertheless, it is quite clear that very small population trends, such as 1% per year, are not detectable in any reasonable time span. Trends in the order of 15-30% per year may be detectable over the 6 year time-span imposed by the EC Habitats Directive. Smaller per-year trends will require a longer time span to detect.

Statistical methods:

- There are many outstanding issues in spatial and spatio-temporal modelling even of high quality line transect data (Thomas and Scott-Hayward in prep.). These issues are even more prominent when diverse sources of data are combined, some without the information to directly assess detectability.
- There are many potential methods that can be applied to JCP data, to inform the monitoring objectives. We have reviewed many of them here, but are unable to reach a conclusion as to which is currently best. Further work is needed.

Exploratory analysis:

- Very superficial analyses were undertaken on a subset of the data that could be made available.
- There appears to be partial spatio-temporal confounding between time and survey method in this dataset, but this may be addressed by looking at all available data.
- Nevertheless, it appears possible to at least fit linear and additive models to the data, although whether they produce sensible results is an open question still.
- The analyses performed here relied heavily on large amounts of high quality data provided by the Sea Watch Foundation. Such data are not available elsewhere. We also used ferry data from the Irish Whale and Dolphin Group. On the other hand, there was not time to make proper use of the other large-ship platforms of opportunity data, nor were the SCANS datasets used. In an analysis of ESAS data, Winship (2008) found major confounding between factors that potentially explain density and those related to sightability. Hence whether approaches that try to separate the two are feasible in general is still not fully known.

Data sets:

- Land-based watches can cover only a very small proportion of the nation's waters, and so are of limited utility in a nationwide population monitoring scheme unless one is prepared to make the strong assumption that temporal patterns seen inshore will apply to some larger area. Examination of the extent to which temporal patterns seen on land-based surveys match those on off-shore surveys that also cover areas surveyed from land may prove useful. Land-based surveys may also have utility for very coastal species.
- The lack of direct information on detectability (e.g., via the implementation of rigorous line transect survey protocols) significantly limits the options open to the analyst, and therefore the reliability of the results. Either data without direct detectability information must be assumed to be comparable with those that do have such information, or covariates thought to influence detectability but that are not correlated with patterns of density (a very limiting restriction) must be included in the spatio-temporal density modelling. The collection of high quality detectability information should receive higher priority.
- Other survey modalities, such as acoustic monitoring, may be useful; a comprehensive review of alternative modalities is in Berggren *et al.* (2008a, b).

Recommendations

- The current effort to define monitoring metrics and realistic targets for precision or power should be continued and completed.
- A significant amount of research time should be devoted to developing and testing statistical methods for monitoring spatial and temporal trends from JCP data. It is clear that the topic is complex, with many issues and potential solutions. There are (at least) three alternative ways forward:
    - Sponsor a relatively short piece of work (in the order of a few months) to make significant progress on a spatio-temporal analysis of relative abundance in the southern Irish Sea data. This should be extended temporally to include SCANS I and II data, and the datasets that were not analyzed for this report, as well as any other available data, should be considered for inclusion. Analysis should be split by species, use better smoothing methods (such as those of Wood *et al.* 2008), and investigate the potential for inclusion of covariates affecting sightability that are collected on all surveys (potentially in some cases as random effects) as well as covariates such as season that are known to affect density. A second test dataset, as different as possible in character, could also be used. Time for data checking and obtaining cleaned data from data providers would need to be included in the work specifications. As well as producing initial spatio-temporal estimates for target species, a primary outcome would be recommendations on whether and what additional work is likely to lead to significant useful outcomes.
    - Commit at this stage to the significant work required to develop the required methods to perform an integrated analysis of all JCP data, and implement such an analysis. One aspect that was missing from the current study, but that should be investigated in a longer-term research topic, is the integration of other survey modalities, such as acoustic data (see Berggren *et al.* 2008a, b). It is hard to judge accurately how much work this would require, without performing the previous step, but a rough estimate is that it would require approximately 1 year of time by a well-qualified post-doctoral level statistician.
    - Fund a PhD student to focus on this topic for the next 3-4 years. Robust and efficient analysis of JCP data would seem like an ideal PhD student topic.
- Serious consideration should be given to whether a rigorous line transect field protocol could be implemented on all JCP transect surveys, so that direct information on detectability is available directly from all datasets. In many cases, there is no reason why platform of opportunity data should be any less rich than dedicated survey data, apart from the transect layout. For ESAS data, consideration should be given to adding dedicated marine mammal observers, using SCANS-like protocols. Training officers could be employed, together with an equipment budget, to help volunteer-based surveys, such as those from Ferries, to implement the same protocols. As well as allowing detectability to be determined independently from density variation, this should also help to

standardize results from the different surveys, potentially resulting in lower CVs of the spatio-temporal trend estimates.

- At this stage, it is not possible to say exactly what common set of covariates should be included in the JCP protocol, without further analyses and a discussion of the previous recommendation. Nevertheless, it is clear that it is essential to record information on covariates that affect detectability (such as sea state, platform height, vessel speed, vessel, observer, survey protocol), as well as the usual information about effort and sightings. For effort, it would be very useful to have easy access to the start and end times and locations of each leg of effort. Sufficient resources should be allocated to rigorous data checking.
- Land-based watches appear to have limited value for this purpose. However, if they are to be used, consideration should be given to implementing a snapshot-like protocol, rather than the current 15-minute window approach.

Literature Cited

Aarts, G. 2007. Modelling space-use and habitat preference from wildlife telemetry data. PhD thesis, University of St Andrews. Available at http://hdl.handle.net/10023/327.

Anderson, D.R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29: 1294-1297.

Anderson, D.R. 2003. Response to Engeman: index values rarely constitute reliable information. Wildlife Society Bulletin 31: 288-291.

Asher, J., M.S. Warren, R. Fox, P.T. Harding, G. Jeffcoate, and S. Jeffcoate. 2001. The Millennium Atlas of Butterflies in Great Britain and Ireland. Oxford University Press.

Baillie, S.R., S.P. Brooks, R. King and L. Thomas. In press. Using a state-space model of the UK Song Thrush Turdus philomelos population to diagnose the causes of a population decline. Environmental and Ecological Statistics

Berggren, P., J. Teilmann, D.L. Borchers, L.M. Burt, D. Gillespie, J. Gordon, K. Macleod; R. Leaper, M. Scheidat, R. Swift, M.L. Tasker, A.J. Winship and P.S. Hammond. 2008a. Review of methods previously used in monitoring temporal and spatial trends in distribution and abundance of cetaceans. Appendix D2.1 of Small Cetaceans in the European Atlantic and North Sea (SCANS-II) Final Report. EU Life project LIFE04NAT/GB/000245.

Berggren, P., J. Teilmann, D.L. Borchers, L.M. Burt, D. Gillespie, J. Gordon, K. Macleod; R. Leaper, M. Scheidat, R. Swift, M.L. Tasker, A.J. Winship and P.S. Hammond. 2008b. Quantitative comparison of monitoring methods and recommendations for best practice. Appendix D2.4 of Small Cetaceans in the European Atlantic and North Sea (SCANS-II) Final Report. EU Life project LIFE04NAT/GB/000245.

Berggren, P. & Wade, P. 2001. Bycatch threatens Baltic harbour porpoise with extinction does anyone care? Talk presented at the 14th Biennial Conference on the Biology of Marine Mammals, October 2001, Vancouver, Canada.

Besbeas, P., and S.N. Freeman. 2006. Methods for joint inference from panel survey and demographic data. Ecology 87:1138-1145.

Beissinger, S.R. and D.R. McCullough, Eds. 2002. Population Viability Analysis. University of Chicago Press.

Bravington, M., D. Borchers and S. Northridge. 2002. Analysis of harbour porpoise sightings data in relation to area-based conservation. Final report to JNCC Project XGLJ02.
Bravington, M.V.B. 2008. Saddlepoint approximations for Poisson & Binomial data with overdispersion induced by spatial random fields. Appendix 4 of Thomas, L., D.L. Borchers, E. Rexstad, S.T. Buckland, M.L. Bravington, S.N. Wood, and S.L. Hedley. 2008. Development of New Methods and Software for Distance Sampling Surveys of Cetacean Populations. Final technical report, Office of Naval Research grant N00014-04-1-0661.
Buckland, S.T. 2006. Point transect surveys for songbirds: robust methodologies. The Auk 123, 345-357.
Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas. 2001. Introduction to Distance Sampling. Oxford University Press.
Buckland, S.T., Borchers, D.L., Johnston, A., Henrys, P.A. and Marques, T.A. 2007a. Line transect methods for plant surveys. Biometrics 63, 989-998.
Buckland, S.T., K.B. Newman, C. Fernández, L. Thomas and J. Harwood. 2007b. Embedding population dynamics models in inference. Statistical Science 22: 44-58
Crespin, L., M.P. Harris, J.-D. Lebreton, M. Frederiksen, and S. Wanless. 2006. Recruitment to a seabird population depends on environmental factors and on population size. Journal of Animal Ecology 75: 228-238.
DiTraglia, F.J. 2007. Models of Random Wildlife Movement with an Application to Distance Sampling. MMaths Statistics Thesis, St. Andrews University. Available at http://eprints.st-andrews.ac.uk/archive/00000451/
Ellingson, A.R. and P.M. Lukacs. 2003. Improving Methods for Regional Landbird Monitoring: A Reply to Hutto and Young.Wildlife Society Bulletin 31: 896-902.
Engerman, R.M. 2003. More on the need to get the basics right: population indices. Wildlife Society Bulletin 31: 286-287.
European Commission. 2006. Assessment, monitoring and reporting under Article 17 of the Habitats Directive: Explanatory Notes & Guidelines Draft 3. Dated April 2006.
Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68: 1364-72
Gerrodette, T & Brandon, J. 2003. TRENDS Version 3.0. Available at http://swfsc.noaa.gov/textblock.aspx?Division=PRD&ParentMenuId=228&id=4740
Gibbs, J.P., S. Droege and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience 48: 935-940.
Gibbons, D.W., J.B. Reid and R.A. Chapman, R.A. 1993. The new atlas of breeding birds in Britain and Ireland: 1988-1991. T. & A.D. Poyser
Hedley, S.L. 2000. Modelling heterogeneity in cetacean surveys. PhD thesis, University of St Andrews.
Hedley, S.L. and S.T Buckland. 2004. Spatial models for line transect sampling. Journal of Agricultural, Biological and Environmental Statistics 9: 181-199
Hedley, S.L., S.T. Buckland and D.L. Borchers. 2004. Spatial distance sampling models. Chapter 4 in Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas (editors). Advanced Distance Sampling. Oxford University Press.

Henrys, P. 2005. Spatial Distance Sampling Modeling of Cetaceans Observed from Platforms of Opportunity. MMath dissertation, University of St Andrews. Available from http://www.creem.st-and.ac.uk/tiago/webpages/pdfs/Henrys2005.pdf.

Hutto, R.L. and J.S. Young. 2002. Regional Landbird Monitoring: Perspectives from the Northern Rocky Mountains. Wildlife Society Bulletin 30: 738-750.

Hutto, R.L. and J.S. Young. 2003. On the Design of Monitoring Programs and the Use of Population Indices: A Reply to Ellingson and Lukacs. Wildlife Society Bulletin 30: 738-750.

Kaimi, I. In prep. PhD Thesis, University of Lancaster.

Laake J.L. and D.L. Borchers. 2004. Methods for incomplete detection at distance zero. In Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) Advanced distance sampling. Estimating abundance of biological populations. Oxford University Press, Oxford

Marques, T.A. 2007. Incorporating measurement error and density gradients in distance sampling surveys. PhD thesis, University of St Andrews.

Maunder, M.N. and A.D. Langley. 2004 Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fisheries Research 70: 389-395

Nichols, J.D., L. Thomas and P.B. Conn. In press. Inferences about landbird abundance from count data: recent advances and future directions. Environmental and Ecological Statistics.

Norvell, R.E., F.P. Howe and J.R. Parrish. 2003. A seven-year comparison of relative-abundance and distance-sampling methods. The Auk: 120: 1013-1028

Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., and Sauer, J. R. (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13, 105–119.

Royle, J.A., D.K. Dawson and S. Bates. 2004. Modelling abundance effects in distance sampling. Ecology 85: 1591-1597

Sauer, J. R., J. E. Hines, and J. Fallon. 2007. The North American Breeding Bird Survey, Results and Analysis 1966 - 2006. Version 10.13.2007. USGS Patuxent Wildlife Research Center, Laurel, MD. Available at http://www.mbr-pwrc.usgs.gov/bbs/.

Sauer, J.R. and W.A. Link. 2004. Some consequences of using counts of birds banded as indices to populations. Pages 168-172 in Ralph, C.J. and E.H. Dunn. (Eds.) Monitoring Bird Populations Using Mist Nets. Studies in Avian Biology 29.

Schweder, T. 1974. Transformation of point processes: Application to animal sighting and catch problems, with special emphasis on whales. PhD thesis, University of California, Berkeley.

Skaug, H.J. 2006. Markov modulated Poisson processes for clustered line transect data. Environmental and Ecological Statistics 13: 199-211.

Steidl, R.J. and L. Thomas. 2001. Power analysis and experimental design. Pp 14-36 in Design and Analysis of Ecological Experiments, 2nd edition. S. Scheiner and J. Gurevitch, editors. Oxford University Press

Thomas, L., K.P. Burnham & S.T Buckland. 2004. Temporal inferences from distance sampling surveys. Chapter 5 in Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas (editors). Advanced Distance Sampling. Oxford University Press.

Thomas, L., Laake, J.L., Rexstad, E., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Burt, M.L., Hedley, S.L., Pollard, J.H., Bishop, J.R.B. and Marques, T.A. 2008. Distance 6.0. Beta 5. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/

Thomas, L. and L. Scott-Hayward. In prep. Report on technical workshop: Recent developments and future directions in spatial and habitat modelling from cetacean line transect and related data. St Andrews, Scotland, 6-8 May 2008.

Thompson, P.M., B. Wilson, K. Grellier and P.S. Hammond. 2000. Combining power analyses and population viability analysis to compare traditional and precautionary approaches to conservation of coastal cetaceans. Conservation Biology 14: 1253–1263.

Torres, L.G., A.J. Read and P. Halpin. In press. Fine-scale habitat modelling of a top marine predator: Does prey data improve predictive capacity? Ecological Applications.

Waagepetersen, R. and Schweder, T. 2006. Likelihood-based analysis of clustered line transect data. Journal of Agricultural, Biological, and Environmental Statistics 11: 264-279.

Winship, A.J. 2008. Analysis of harbour porpoise sighting rates from the European Seabirds at Sea database. Appendix D2.4 of Small Cetaceans in the European Atlantic and North Sea (SCANS-II) Final Report. EU Life project LIFE04NAT/GB/000245.

Wood, S.N., M.V. Bravington and S.L. Hedley. 2008 Soap film smoothing. Journal of the Royal Statistical Society, Series B 70: 931-955.

Acknowledgements

The author thanks Charles Paxton, Monique MacKenzie and Arliss Winship for their comments and ideas on this work. Thanks also to those attending the start-up workshop at St. Andrews: Andy Webb, Peter Evans, Giovanna Pesante, Simon Northridge, Tim Dunn, Sonia Mendes, Tom Stringell, Charles Paxton, Monique MacKenzie, Arliss Winship, Chris Pierpoint and Dave Wall.

Tables and Figures

Table 1. Summary of datasets discussed in this report.

Dataset	Organization	Contact	Survey method	Spatial range	Temporal range	Taxonomic range
A. Cardigan Bay land-based watches	Ceredigion County Council and others	Chris Pierpoint	Standardized land-based watches	6 sites in southern Cardigan Bay	1994 (3 sites) / 2004 (3 sites) - 2007 (ongoing); year round	All cetacean species, but mostly bottle-nose dolphin
B. Sea Watch dedicated line transect surveys in Cardigan Bay SAC	Sea Watch	Giovanna Pesante	Double-platform line transect survey from small boats	Cardigan Bay SAC	2001 and 2003-7; Apr-Sept.	All cetacean species but mostly bottle-nose dolphin and Harbour porpoise
C. Sea Watch *ad libitum* line transect surveys in Cardigan Bay (data not supplied)	Sea Watch	Giovanna Pesante	Photo-ID survey from small boats	Cardigan Bay, focussed on Cardigan Bay SAC and Pen Llyan A'r Sarnau SAC	2001-2007; Apr-Oct	Bottle-nose dolphin
D. Sea Watch dedicated line transect surveys in St George's channel	Sea Watch	Peter Evans	Double-platform line transect survey from moderate-sized boats	St George's channel (Celtic Deep)	2004-6; May-Nov	Mostly common dolphin; 7 other species

Table 1. (contd.)

Dataset	Organization	Contact	Survey method	Spatial range	Temporal range	Taxonomic range
E. ESAS surveys (including those from ferries)	JNCC and others	Andy Webb	Single-platform mostly ship-based transects, much with distances	Patchy; most from Dublin-Holyhead and Rosslare-Pembroke ferry routes.	1997-2001 (in dataset given for analysis); year round	All cetacean species
F. IWDG Irish Sea Ferry Surveys	IWDG	Dave Wall	Single-platform transects (no distances)	Principally Dublin-Holyhead and Rosslare-Pembroke ferry routes; also some Dublin-Cherbourg and Dublin-Liverpool	2001-2008; year-round	All cetacean species
G. Sulaire *ad libitum* surveys	Steve Hartley	Peter Evans	Single-platform line transect	Largely Cardigan Bay SAC	1995-2007 (more available; earlier data does not have effort however); May-Sept	All cetacean species; principally bottle-nose dolphin

Table 2. Annual effort from each survey for which there was effort data. For the land watch data (A), effort is observer hours. For the Sea Watch datasets (B and D), IWDG ferry surveys (F) and Sulaire data G, effort is line length in km and all available effort (designed line transect and otherwise, including casual watching) are included. Surveys and years used in the exploratory modelling exercise are indicated as bold cells in the table.

Dataset	**Effort units**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
A. Cardigan Bay land-based watches	Observation hours									739	755	611	655
B. Sea Watch dedicated line transect surveys in Cardigan Bay SAC	Transect length (km)						**2779**		567	**1321**	8269	2405	**2013**
D. Sea Watch dedicated line transect surveys in St George's channel	Transect length (km)									**1003**	**1635**	**2761**	
F. IWDG Irish Sea Ferry Surveys	Transect length (km)						**1030**	**1045**	**3889**	**1532**		**2823**	**1379**
G. Sulaire *ad libitum* surveys	Transect length (km)	**2551**	**2951**	**895**	**2774**	**2134**	**1761**	**4225**	**832**		**4505**	**6893**	**1311**

Figure 1. Statistical power to detect a log-linear population trend after 6 years of annual monitoring, over a range of annual rates of change and residual coefficients of variation (CVs), assuming a constant CV, an alpha-level of 0.1 and a two-tailed *t*-test for trend.

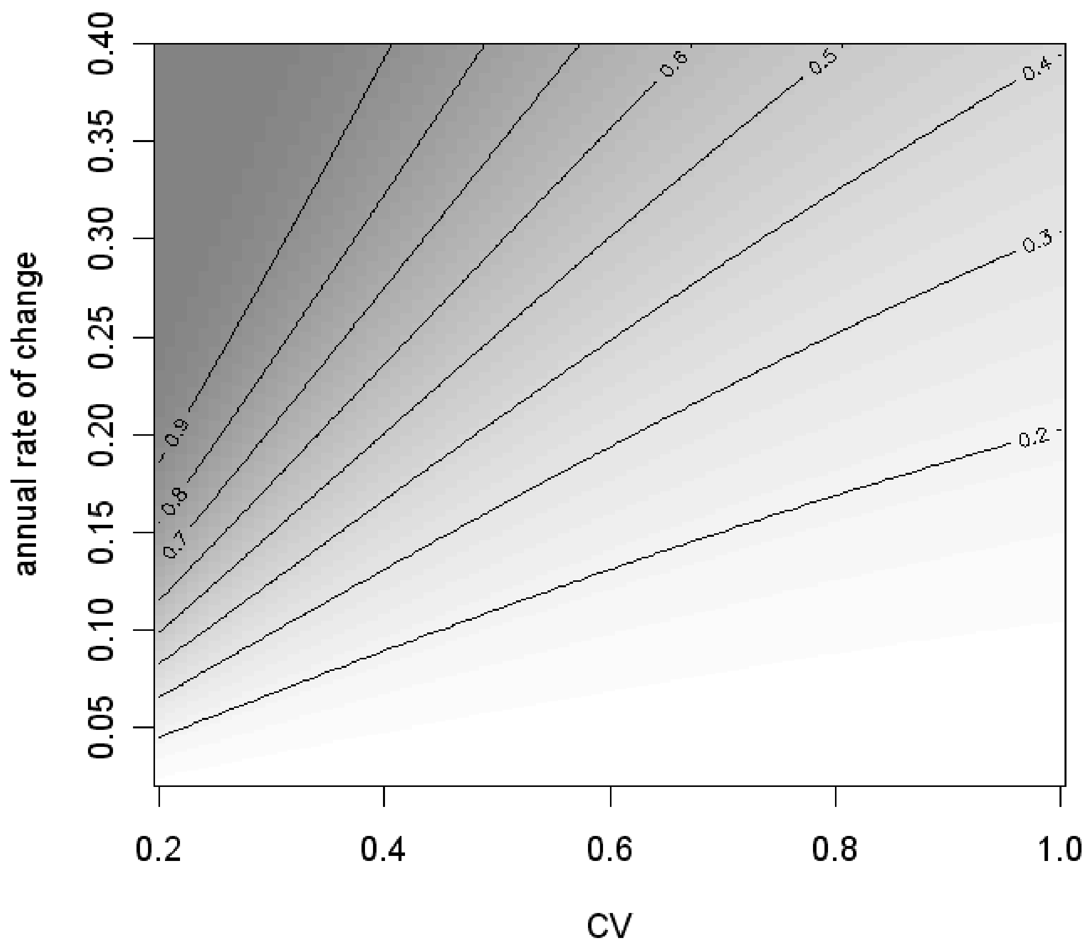

Figure 2. Sketch map showing the approximate spatial extent of each data source.

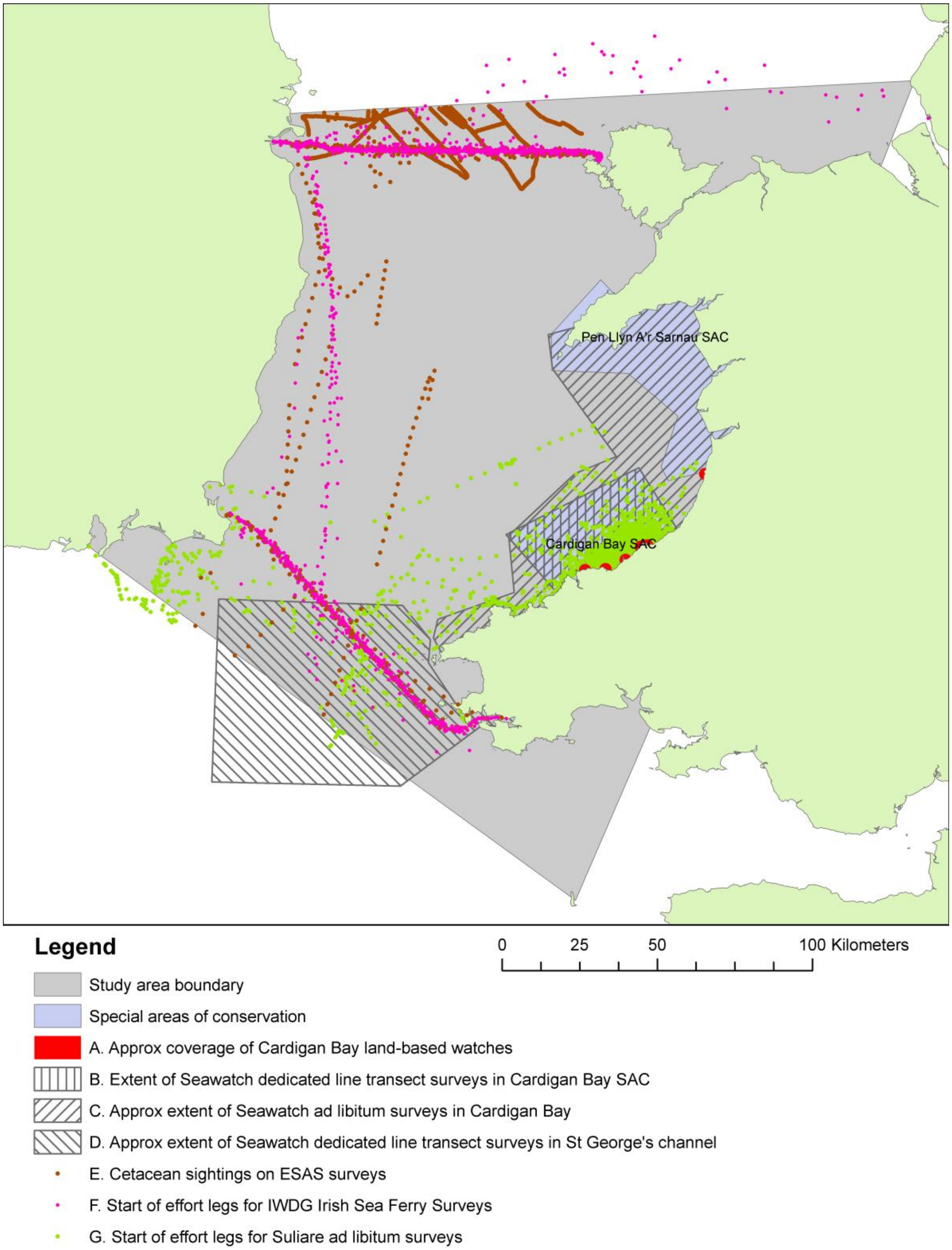

Figure 3. Estimated effect of year on sighting rate of cetaceans from exploratory spatio-temporal analysis of 3 datasets (B, D, F and G, Table 2). Year 0 corresponds to 1996. The smooth is shown on the scale of the linear predictor; the model was fit using a log link function.

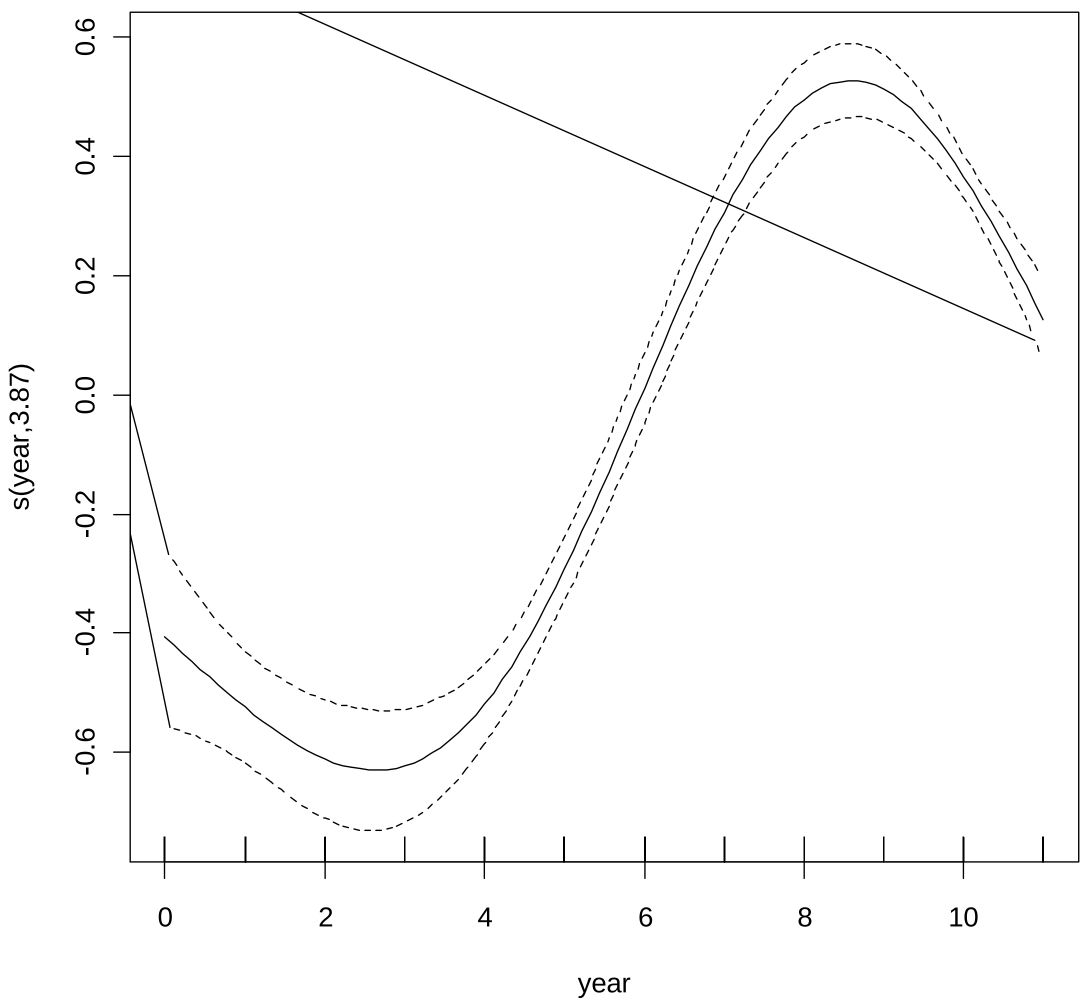

Figure 4. Heatmap of predicted relative abundance of cetaceans from exploratory spatio-temporal analysis of 3 datasets (B, D, F and G, Table 2). Predicted relative abundance is high in dark brown areas and low in yellow areas. Predictions are at a fixed level of the survey and year covariates. The spatial extent of predictions is given by a minimum bounding rectangle on the survey data used in the analysis (see Figure 1 for a map showing the spatial extent of each dataset).