BUh fU 9b[`UbX 7ca a]gg]cbYX FYdcfh B97F &)- Á

DfcXi Wj]cb cZ U XfU Zia cXY``YX a Uf]bY Vci bXUfmYI hYbg]cb Zcf h\ Y =g`Yg cZ GWj`m GD5 Á

NATURAL
ENGLAND

: cfYk cfXÁ

Þæˇ¦æ¼ÁÒ}*¦æ}åå&¦[{{ã•ã¦}Áå¼Áæ}*^Á¦-Á^][¦σ-Á¼[{¼Á¢σ^¦}æ¼&¦}¢æ&¢¦•Á¦[¼Á
]¦[çãˆ^Áçãˆ^}&¼Áæ}åÁæ¼çã¼^Á¦Áæ•ãˆ¦•Á¦åÁ^¦ã¼^¦ã¦*Á¼ˇ¦¦å¼ˇã•ÆV@¼ã¸•Á¦Á¼¢ãˆÁ
¦^][¦¦¼æˇ^Á¦@•^¦-Á¦@¼¼æˇç@¦¦•Á¦}åÁ¼[¦,[¦^&^••æã^Á^]¦•^¦}¼¢@¦•^¦-ÁÞæˇ¦æ¼Á
Ò}*¦æ}åÆÁ

6 UW[fci bXˇ

This report is being published to support the formal consultation of a proposed marine extensionto the Isles of Scilly Special Protection Area (SPA).

Þæˇ¦æ¼Á}*¦æ}åå&¦[{{ã•ã¦}^åÁPãˇ^-ÁŒ¦¼æ¼Á
Ùˇ¦çãˆ¦*Sã¦ã¦å¼&¦}å¼¼&ã¦¼^¢æ¼æ¼å¼æ}æ^•ã¼Á
[¼Á¦çˇ}æ¼Áæ¼æ¼[ˇ¦&¼¼^¦&¦ãã¦*Á@ø¼Á
Phalacrocorax aristotelis åã¼bã¦¼¢¦}¼¼[¦}åå&@¼Á
Q¦^¦-Á¼¦&å¦¼ˇã¦¼Á¼@¼¼^^ã¦*Á^æ¦[¦ÆV@¼Á
¸[¦¼Á¼æ¼¦[{{ã•ã¦}^å¦Á&¼^ç¦¦{ã¼^Áㆼæ¼¦[^Á
¦^^å㦼æ}åÁ¼æã¼ç^}æ}&¼å&^æ^æ¦¼Á[¦@¼Á^&ã•Æ
ã¦Áˇ[¦}¼Á¼æ^}¼çㆼ¦•¼¼ãæ¼æ^¦¦¦ãã¼•Á¼Á¼an
ÚÚÓÆ¦¦[][•æ¼¼æ¼ˇ¦å¦@¼¼¦æ}¼å•Ææ}å¼æÁ¼^-ã¦^Á
à[ˇ}å椕涼¼[}å¼@•^¦¦¦ãã¼•Æ

V@ãˇ¼^][¦¼¼@ˇˇ¼å¼å¼^&æ¦^å¼æ•¼Á
Production of a Draft Modelled
Boundary Extension for the Marine
Isles of Scilly SPA. Natural England
Commissioned Report NECR259 .Á

Production of a draft modelled boundary extension for the Marine Isles of Scilly SPA

Contents

Figures

Tables

Executive Summary

Natural England ("NE") commissioned HiDef Aerial Surveying Limited ("HiDef") to conduct a detailed analysis of potential data sources describing shag *Phalacrocorax aristotelis* distribution around the Isles of Scilly during the breeding season. This work was commissioned to determine suitable feeding and maintenance areas for the species, in turn to identify most suitable territories for a Special Protection Area ("SPA") proposal around the islands, and to define boundaries around these territories.

Guidelines have been produced by the Joint Nature Conservation Committee ("JNCC") for selection of most suitable territories for SPAs under the European Union Birds Directive [2009/147/EC]. These have been used to classify several sites in the UK in the marine environment. The approach taken for shag by the JNCC so far has been to identify important areas based on existing data from a variety of sources. Several methods have been described for determining the most suitable territories for seabirds based upon modern data modelling techniques applied to seabird distribution data derived from tracking studies and transect-based surveys. In the UK, a technique has been developed for delimiting boundaries from such modelled data known as maximum curvature, which describes the optimum trade-off between the number of birds a site contains and the size of the area.

Four potential sources of shag distribution data were identified for use in this analysis: non-targeted boat-based data from the European Seabirds at Sea ("ESAS") database, non-targeted visual aerial survey data, targeted digital aerial survey data and geographical positioning systems ("GPS") logger tracking data from two colonies in the Isles of Scilly.

Environmental covariate data were identified for use in data modelling: distance from land, distance from colony, bathymetry, EUNIS habitat maps, remotely sensed sea surface temperature. Digital Aerial Survey ("DAS") data for sitting shags (defined as all birds observed in contact with the water surface) collected from three surveys each in the breeding season of 2014 and 2015 were analysed using two methods: relatively simple, but robust kernel density estimation ("KDE") which uses information on the nearest neighbouring samples to predict the abundance at un-sampled locations, and density surface modelling ("DSM") which uses the Complex Region Spatial Smoother ("CReSS") method, with targeted smoothing capabilities using the Spatially Adaptive Local Smoothing Algorithm "SALSA"1D and SALSA2D methods. These methods are currently implemented in the Marine Renewables Strategic Environmental Assessment ("MRSea") package. This method describes the relationship between abundance, location and the habitat covariates to predict the abundance at un-sampled locations.

Tracking data from ten birds fitted with archival GPS tags in 2010, 2011 and 2012 were analysed based on presence and absence data. A random sampling method of these was used to equalise the number of presences and absences in the data to counteract problems in model fitting and model validation. The CReSS and SALSA methods were used to describe the relationship between presence and absence of shags and habitat covariates to predict the probability of presence at places for which there were no tracking data.

Non-targeted ESAS and visual aerial survey data were mapped using a DSM approach, but predictions were provided at the relatively coarse scale of 3km x 3km squares. The prediction grids from the density and probability models were summarised for 0.5km x 0.5km squares (KDE analysis), or 1km x 1km grid squares for DSM and tracking analysis, then sorted from highest to the lowest value. The cumulative abundance or probability and cumulative area was then calculated for each grid cell. The

point in the plotted curve in which there is the greatest change in the relationship is the point of maximum curvature and can be defined by a single exponential or a double exponential function. The method calculates the density at which this occurs and this density can then be used to delineate the boundary of the site.

KDE produced relatively simple prediction grids of shag distribution over a wide area in 2014 and 2015, whereas DSM analysis of the same data produced a tight prediction of shag distribution close to the islands. The DSM used depth and change in sea surface temperature ("SST") but rejected distance from colony, SST and EUNIS level 2 and 3 habitats as environmental covariates. The model provided mostly good fits of shag density compared to observed density and showed this species to be located mostly close inshore around the islands.

Model outputs from analysis of tracking data showed areas of high predicted presence also to occur close inshore around the Isles of Scilly, but extending further offshore than the model outputs from the DSM. The models found EUNIS level 3 habitat to provide a good fit for the model. However, the extrapolated prediction grids proved to be unrealistic and this habitat covariate was dropped from the model. Depth, SST and distance from colony did not provide a good fit for the models and were not used for the model predictions. This meant that it was not possible to use the model to predict the feeding areas used by birds at different nesting locations from the tagged birds within the Isles of Scilly archipelago.

Maximum curvature analysis of these prediction grids was based on single and double exponential functions; the use of more complex and potentially inappropriate additional functions was not found to be necessary. The grid squares selected from these analyses differed in extent and nature according to the type of analysis carried out. The KDE output resulted in the largest area being selected, whereas those selected from the DSM were all from close inshore within the Isles of Scilly archipelago. Modelling of tracking data, using the CReSS method, with targeted smoothing capabilities using the SALSA, also predicted the majority presence to occur close inshore around the islands, but extended further offshore than the DSM method.

Of the three methods, the DSM data from the digital aerial surveys were considered to be the most robust on the basis of the inclusion of two habitat covariates in the model fitting. The modelled distribution of tracking data needed to be interpreted with care, but still gave reasonably realistic predictions of abundance around the islands, if further offshore than the DSM. The KDE analysis of the digital aerial survey data resulted in grid cells being selected for inclusion in the marine SPA proposal quite a long distance offshore all around the islands which were considered to be the least realistic of the three methods employed.

Two potential boundary options were recommended for the marine SPA proposal for shags: a boundary around the DSM selected grid squares and the colony location using a double exponential maximum curvature model (option 1), and a precautionary boundary around the tracking data selected grid squares using a single exponential maximum curvature model (option 2). The use of the colony data ensured that the flight lines between the colonies and the predicted feeding areas selected from the models would also be included within a potential SPA. The option 1 boundary was recommended as being the most realistic representation of a most suitable territory for shags based on the observed distribution from digital aerial survey, tracking data and non-targeted ship-based and visual aerial survey

data. The option 2 boundary is not recommended by itself, but if merged with the option 1 boundary, could represent a more precautionary third option.

1 ˙ Introduction

1Á   Member states of the European Union ("EU"), including, at present, the United Kingdom ("UK"), are required by the Birds Directive [2009/147/EC] to find and protect the "most suitable territories" of species listed under Annex I of the Directive, or migratory species. These "most suitable territories" are required to be designated as "Special Protection Areas" ("SPA").

2Á   In the UK, one migratory species that requires most suitable territories to be identified and suitable SPAs to be designated for the protection of its populations is European shag *Phalacrocorax aristotelis*. European shags (hereafter "shag") are medium sized seabirds that dive from the surface of the water to forage on pelagic or benthic prey. They breed in coastal colonies, often with other seabirds. To date the only SPAs designated to protect shag populations are at these breeding colonies. Despite their exclusively marine foraging behaviour no SPAs entirely in marine habitats have been designated to protect birds foraging at sea from coastal colonies.

3Á   The approach taken for shag by the Joint Nature Conservation Committee ("JNCC") so far has been to identify important areas based on existing data from a variety of sources. Areas have been identified based on the seabird aggregations analysis (where surveys were undertaken sufficiently close to the shore), visual aerial survey where observers could identify the species, and tracking data collected by the Centre for Ecology and Hydrology ("CEH").

4Á   At present, there are 13 terrestrial SPAs in the UK that include shag as a qualifying feature, all at terrestrial breeding colony locations. Only one of these SPAs is located south of 55°N, in the Isles of Scilly. The Isles of Scilly SPA is designated for its qualifying populations of European storm-petrel *Hydrobates pelagicus*, lesser black-backed gull *Larus fuscus*, and its seabird assemblage of more than 20,000 individuals in the breeding season (including shag as a named qualifier).

5Á   Natural England ("NE") commissioned HiDef Aerial Surveying Limited ("HiDef") to conduct a detailed analysis of potential data sources describing shag distribution around the Isles of Scilly during the breeding season to identify most suitable territories for a SPA proposal around the islands and to define boundaries around these territories.

6Á   The selection of "most suitable territories" in the UK follows guidance from JNCC (JNCC, 1999). This is a two-stage process. Stage 1 identifies areas that are likely to qualifying as SPAs. Stage 2 then considers these areas further to determine whether these are the most suitable in the number and size of SPAs. JNCC provides four guidelines to determine whether a site may qualify under Stage 1:

1. An area is used regularly by 1% or more of the Great Britain (or in Northern Ireland, the all-Ireland) population of a species listed in Annex I of the Birds Directive in any season;
2. An area is used regularly by 1% or more of the biogeographical population of a regularly occurring migratory species (other than those listed in Annex I) in any season;
3. An area is used regularly by over 20,000 waterfowl (waterfowl as defined by the Ramsar Convention) or 20,000 seabirds in any season; and
4. An area which meets the requirements of one or more of the Stage 2 guidelines in any season, where the application of Stage 1 guidelines 1, 2 or 3 for a species does not identify an adequate suite of most suitable sites for the conservation of that species.

7Á Once a site has been determined likely to qualify as a SPA under Stage 1, a further seven categories are recommended to determine whether the site is a most suitable territory:

1. Population size and density – areas holding or supporting more birds than others and/or holding or supporting birds at higher concentrations are favoured for selection;
2. Species range – areas selected for a given species provide as wide a geographic coverage across the species range as possible;
3. Breeding success – areas of higher breeding success than others are favoured for selection;
4. History of occupancy – areas known to have a longer history of occupation or use by the relevant species are favoured for selection;
5. Multi-species areas – areas holding or supporting the larger number of qualifying species under Article 4 of the Directive are favoured for selection;
6. Naturalness – areas comprising natural or semi-natural habitats are favoured for selection over those which do not; and
7. Severe weather refuges – areas used at least once a decade by significant proportions of the biogeographical population of a species in periods of severe weather in any season, and which are vital to the survival of a viable population, are favoured for selection.

8Á JNCC note that the Stage 2 guidelines are particularly important for selecting and determining boundaries of SPA for "thinly dispersed and wide-ranging species".

1.1˙ Review of boundary setting methods

9Á Recent publications have reviewed much of the available information on boundary setting methods for Natura 2000 sites in the UK. The Nature Directives give little direct guidance on designating sites, only that "most suitable territories" must be designated by member states. The European Commission ("the Commission") provide good, high level, guidance on designating sites in the marine environment (European Commission, 2007). To collect suitable data for designating a feature such as breeding shags, the Commission recommends that a variety of survey types are suitable, including aerial and boat-based surveys, existing ESAS data and telemetry data. The Commission guidance urges care in the analysis of existing data, as data quality, extent and the original purpose of the data collection may not match the needs of the analysis required to identify sites. Where data are analysed it is also recommended that habitat and other covariates are used (where available) to help explain the variance in bird numbers and distribution. These could also be used to predict use outwith the areas that have been surveyed. The Commission suggests that there are two key features of site boundaries: site size and site shape. It is noted that site size should, "*provide the basis for adequate protection of the features of conservation interest*". Particularly pertinent to the designation of marine areas for shag in the Isles of Scilly is the guidance that, "*where the nature or scale of the species dispersion results in more loose aggregations, resulting in smaller concentrations disjunct from the core aggregation, the decision whether to include such satellite concentrations within a boundary should be made by reference to the overall size of the qualifying interest*". On site shape the Commission advocates the use of simple shapes using straight lines to delineate site boundaries to aid surveys and monitoring that may be necessary for management. Birdlife International (2010) recommends that greater weight is provided to areas with more layers of data (i.e. more than one data collection methodology), for its Important Bird Areas programme.

10Á   Different methods for estimating the abundance of birds across an area of sea to designate an SPA have been used included spatial kriging (e.g. Webb *et al.* 2004) and kernel density estimation ("KDE") (O'Brien *et al.* 2012). Methods used to determine boundaries have included gradient analyses (Garthe *et al.* 2012), Getis-Ord Gi* statistics (Kober *et al.* 2010), and maximum curvature analysis (O'Brien *et al.* 2012). Across much of the rest of the European Union, apparently, many SPA boundaries were based on expert judgement, with no explanation of the process of determining these being published, so no methodology or audit trail of decision making is available.

11Á   Many countries in the EU have used the BirdLife International Important Bird Areas information to designate sites. Many of the criteria used by BirdLife to determine IBA locations (BirdLife International, 2001) are very similar to those recommended by JNCC in the UK (Stroud *et al.*, 2001). In addition, for application in the marine environment BirdLife International have advocated a simple model of foraging range information (Soanes *et al.* 2016), which can be modified with habitat and/or bathymetry data where this is available and is known to be a suitable predictor of species abundance.

12Á   The most recent draft SPAs in the UK have used different methods for establishing species (or multi-species) hotspots of abundance, but have all applied the maximum curvature approach to delineating boundaries. This method selects an area that represents the optimum trade-off between the number of individuals contained within the site and the size of area. The maximum curvature approach is detailed in O'Brien *et al.* (2012). In general terms, this selects the cells of a prediction grid with the highest bird density in plot of the relationship between the cumulative population size as each grid cell is added and compared to the cumulative area of the selected grid cells. The point in the plotted curve in which there is the greatest change in the relationship is the point of maximum curvature and can be defined by a single exponential or a double exponential function. The method calculates the density at which this occurs and this density can then be used to delineate the boundary of the site.

13Á   This method has been used to describe SPAs in the Outer Thames Estuary and Liverpool Bay for red-throated divers *Gavia stellata* and common scoters *Melanitta nigra* (O'Brien *et al.* 2012) and for inshore waterbirds around the coast of Scotland (Lawson *et al.* 2015) and England (Lawson *et al.* 2016). A variation on this method was employed for an extension to existing tern colonies around the UK, in which models, based on tracking of adult terns foraging flights around their colonies during the breeding season, were used to predict relative usage of the sea around these and un-surveyed colonies (Wilson *et al.* 2014). They used the measure of relative usage of the sea by the different tern species as a proxy for density or the number of birds, and were able to use maximum curvature of the relationship between cumulative relative usage and cumulative area. A similar method was used on tracking data from shag colonies in and around the Firth of Forth (Daunt *et al.* 2016).

1.2   Sources of shag spatial abundance data

14Á   There were three sources of bird data available for mapping the spatial abundance of shags from the Isles of Scilly SPA: digital aerial survey ("DAS") data; geographical positioning systems ("GPS") logger tracking data; and, European Seabirds at Sea ("ESAS") and visual aerial survey data processed in the Seabird Mapping and Sensitivity Tool ("SeaMaST").

15Á   The DAS data were collected by HiDef during the breeding seasons of 2014 and 2015 for NE. Two reports were provided to NE on the spatial abundance of shags around the Isles of Scilly, one for each

breeding season (Webb & Irwin 2016 & Irwin 2015). These data were the main data source for modelling the spatial abundance of shags in relation to the key environmental covariates.

16Á   Shags from three colonies on different islands in the Isles of Scilly were fitted with GPS archival tags during the nesting season (Evans *et al.* 2016). Results showed that birds were mostly foraging within the archipelago, and the foraging ranges of birds from different colonies overlapped considerably (Figure 1). These data are limited as a relatively small number (13) of birds were tagged and recovered. This is also a small proportion (0.5%) of the Isles of Scilly shag population (approximately 1296 pairs), and is fewer than the number of individuals recommended by Soanes *et al.* (2013) needed to represent 95% of the active use area (39 – 135 birds sampled), but is sufficient to represent 50% of the active use area (5 – 9 birds sampled). These data would therefore not be able to inform the boundary of the site on their own, however, they are likely to be useful additional information in checking whether the boundary using other methods is similar.

Figure 1   Tracking data from breeding shags on the Isles of Scilly (from Evans *et al.* 2016)
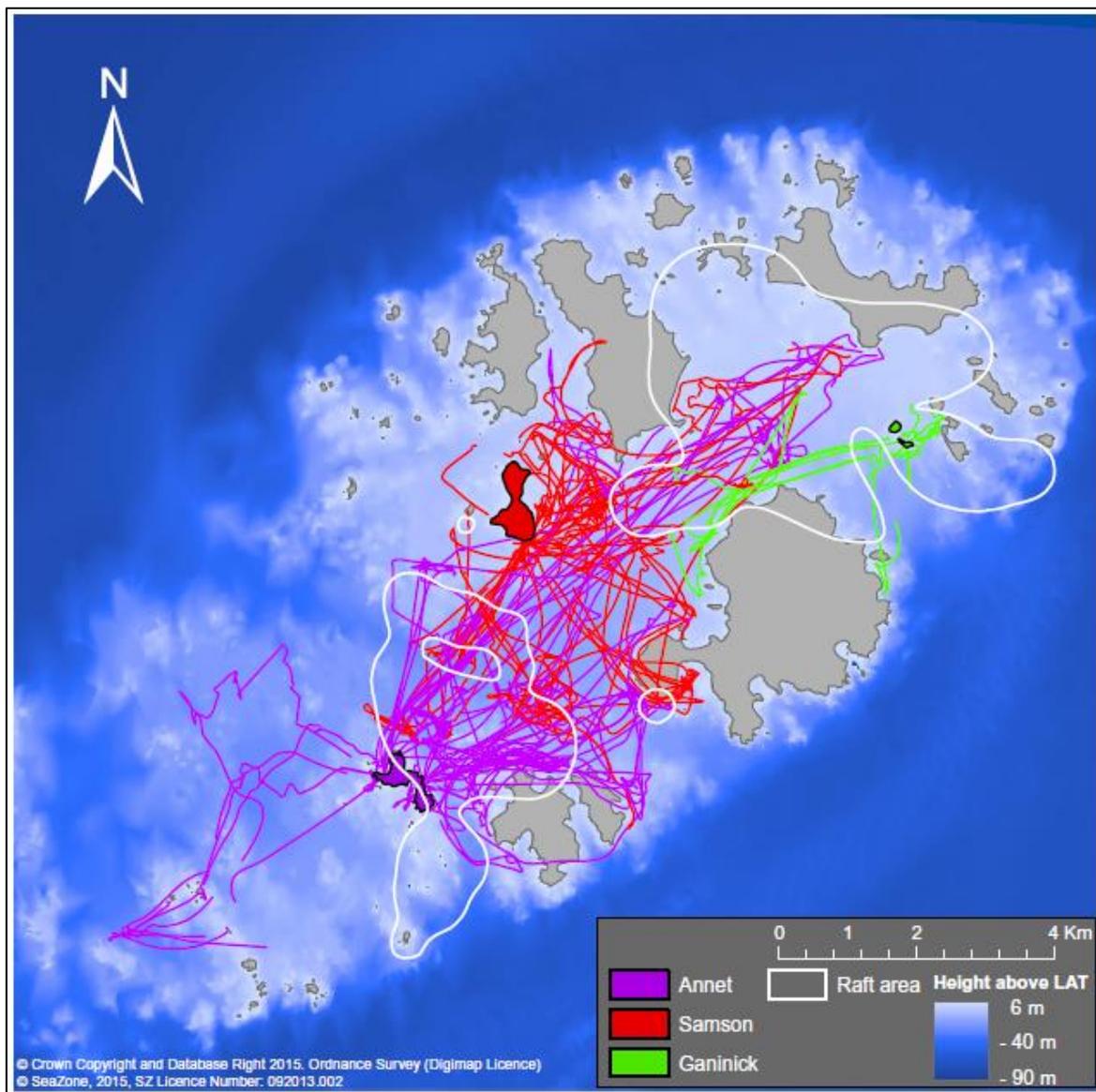
17Á   Data used in the SeaMaST project came from broad scale ESAS boat based and visual aerial surveys (Bradbury *et al.* 2014). These data were explored to determine their value in understanding the spatial abundance around the Isles of Scilly. Due to the coarse nature of the prediction grid of the SeaMaST data it was acknowledged early in this project that the data may not be useful for boundary setting, but they may also be useful as a verification tool.

1.3˙   Sources of covariate data

18Á   A variety of covariate data were available for analysis. Those thought most likely to successfully parameterise spatial models to be used in determining the appropriate site boundary are discussed in turn below.

19Á   **Distance from land** – shags are a very coastal species, seldom being recorded far from coastal waters. It is therefore likely that the distance from land will be an important explanatory parameter in the models. The distance from the nearest permanently dry land (i.e. above mean high water spring) was calculated using ArcGIS.

20Á   **Distance from colony** – since the aim of this project was to determine the boundaries of a marine SPA for breeding shags, the distance from the breeding colony may be a key factor in limiting the distribution of birds. Shags have relatively short foraging ranges, with a mean of only 5.9 ± 4.7 km and a mean of the maximum foraging range of 14.5 ± 3.5 km (Thaxter *et al.* 2012). The geodesic distance was calculated using ArcGIS, as coastal breeding shags rarely fly over land to reach marine foraging areas.

21Á   **Bathymetry** – while shags are a diving bird their maximum diving depth is relatively shallow at only 10 – 40m (Daunt *et al.* 2015). Since they are predominantly a benthic feeding species, this could also limit their ranging behaviour in the breeding season. As such the bathymetry within the study area was also included as a covariate in the spatial models. Bathymetry could also be used to limit the boundary of the SPA at a maximum depth contour (e.g. 80m), as suggested by BirdLife for marine Important Bird Areas (BirdLife International 2010).

22Á   **EUNIS habitat maps** – NE provided shapefiles of habitats around the Isles of Scilly based on the EUNIS habitat categories. Two levels of data were available: level 2 and level 3. Further information is available in section 2.2.

23Á   **Remotely sensed sea surface temperature** – a suitable data source of sea surface temperature ("SST") is from the European Organization for the Exploitation of Meteorological Satellites ("EUMETSAT"). Data is based on retrievals from the Visible Infrared Imaging Radiometer Suite ("VIIRS"). It has a 0.02 x 0.02 degree spatial resolution (about 2.2km x 1.4km at 49.9 degrees North.) and a 12 hour temporal resolution. This data set was explored to determine whether SST varies sufficiently in both spatial and temporal scales to provide useful explanatory power to the models.

2 Methods

2.1 Digital aerial survey data

24Á A series of strip transects were flown monthly between May and July 2014 and 2015. Transects were spaced at 5 kilometre (km) intervals for a wider zone and at 2.5km in a higher interest area over the Isles of Scilly. The broader transects provide contextual information for assessment of seabird abundance. The closer spacing of transects in the high interest area results in greater survey effort and therefore more encounters of birds which is therefore more likely to give greater precision to abundance estimates.

25Á Surveys were undertaken using four (4) HiDef Gen II cameras with sensors set to a resolution of 2 centimetres (cm) Ground Sample Distance ("GSD"). Each camera sampled a strip of 125m width, separated from the next camera by ~25m, thus providing a combined sampled width of 500m within a 575m overall strip.

26Á The surveys were flown using a Diamond DA42 aircraft flying at a height of approximately 550m above sea level ("ASL") (~1800'). Flying at this height ensures that there is no risk of flushing those species which have been proven to be easily disturbed by aircraft noise.

27Á Position data for the aircraft was captured from a Garmin GPSMap 296 receiver with differential GPS enabled to give 1m for the positions, and recording updates in location at 1 second intervals for later matching to bird and marine mammal observations.

28Á Data was viewed by trained reviewers; who marked any objects in the footage as requiring further analysis. For Quality Assurance ("QA"), an additional "blind" review of 20% of the raw data was carried out and the results compared with those of the original review. In the QA process, if 90% agreement is not attained then corrective action is initiated. This includes reviewing the remaining data set and if appropriate, discarding the reviewer's data. If necessary, the data is re-reviewed and additional training provided to the reviewer to improve performance.

29Á Images marked as requiring further analysis were reviewed by specialist ornithologists for identification to the lowest taxonomic level possible. This includes an assessment of approximate age of each animal, as well as any behavioural traits visible from the imagery. At least 20% of all objects are subjected to an external QA process and if less than 90% agreement is attained, the reviewer's data is discarded and the data re-reviewed. Disputed identifications are then passed to a third expert ornithologist for adjudication.

30Á Species identifications were assigned a confidence rating of possible, probable or definite. Ornithologists also noted, where possible, their flying direction, age, plumage and sex.

31Á All data were geo-referenced, taking into account the offset from the transect line of the cameras, and compiled into a single output; Geographical Information System ("GIS") files for the Observation and Track data are issued in ArcGIS shapefile format, using UTM30N projection, WGS84 datum.

2.2 Kernel Density Estimation

32Á Density maps for shags recorded as sitting on the water surface (this included birds diving from the water surface, but excluded all birds in flight) have been derived using a Watson-Nadaraya type KDE

technique (Simonoff, 1996). In KDE, a small 'window' function (the kernel) is used to calculate a local density at each point in the study area. To evaluate the density at a given point, the kernel is centred on that point and all the observations within the window are summed to obtain a local count. The total area of the transect(s) intersecting the window is then summed to obtain a local measure of effort. By dividing the local count by the local effort, a local density estimate is obtained. To build a density map, the study area is covered with a fine mesh of study points and the density is calculated at each point in the mesh in turn.

33Á   Kernel techniques are robust and not as complex as other density estimation techniques because they have fewer parameters; as a result, they are arguably the easiest density surface technique to reproduce independently. The only variables are the size and shape of the kernel or window function. Here, we have used a Gaussian window function, which has the advantage of being smooth, rotationally symmetric and easy to compute. The shape of the Gaussian function is determined by a single width parameter; the selection of this parameter is the only variable in the computation of the density maps.

34Á   Rather than set the width parameter arbitrarily a leave-one-out cross validation method was used. Cross validation estimates the predictive power of a model by removing some of the data from the data set and using the remainder of the data and the model to predict the values for the data that was removed. The closer the predicted values represent the removed data, the better the model performance and the width parameter used in the model.

35Á   To apply cross validation to the survey area, each transect was subdivided into 1 km long segments. To evaluate a particular choice of kernel width, each segment was removed in turn, and the remaining data was used to predict the density of the missing segment and subtract the known value from the prediction to obtain an error score. This process was repeated for every segment and the error scores for all segments were squared and summed to give a total performance score for that particular choice of kernel width. The kernel width was then varied and the process repeated; if the new score was lower than the old, the new kernel width was a better choice than the previous value. An exhaustive search over all kernel widths was then used to identify the best global choice. The result of the process was a smooth density estimate which had been derived without any manual parameter selection.

36Á   All data for the 2014 and 2015 seasons were combined into a single analysis for each year because of the large differences in overall abundance between the two years. Thus, each analysis combined the transects (survey effort) and observations from three flights of similar transect length.

2.3˙ Density Surface Modelling

37Á   The density surface modelling was undertaken using the Complex Region Spatial Smoother ("CReSS") method, with targeted smoothing capabilities using the Spatially Adaptive Local Smoothing Algorithm "SALSA"1D and SALSA2D methods. These methods are currently implemented in the Marine Renewables Strategic Environmental Assessment ("MRSea") package in R (Mackenzie *et al.* 2013). This has recently been updated as part of a Scottish Government contract developing a power analysis tool for data of this kind. These methods are spatially adaptive and are designed to describe surfaces with both local surface features (e.g. patchy surfaces with locally acting hotspots) and/or global surface features (e.g. at surfaces or far-reaching trends). The CReSS smooth implemented employed Euclidean distances alongside robust standard errors for inference to cope with residual correlation within

transects. Quasi-Poisson based models were trialled alongside an offset term for the area associated with each count observation.

2.3.1ˑ Model Selection

38Á The choice of surface flexibility (knot number and their locations) was carried out using the Bayesian Information Criterion ("BIC") score (which balances model fit with model complexity for over-dispersed data) while decisions about whether to retain covariates in a model were assessed using cross-validation since this measure explicitly quantifies fitness of the model to data unseen by the model. Model-based p-values (corrected for residual correlation) were also recorded for comparison. An empirical runs test was used to determine the level of evidence for spatio-temporal residual correlation (if present) and both the within transect residual correlation and over-dispersion were accounted for as part of the model(s) fitted.

2.3.2ˑ Model covariates

39Á Year and Month were included as a factor in the model to permit a flexible time-based element. The "YearMonth" variable is a combination of year and month, so to permit changes in average numbers across years and months. Distance from colony and distance from coast relationships were not used here, as they assume that any patterns are the same for a given distance from the colony or coast regardless of the spatial position of the locations. The spatial term used here allows more flexibility than assuming a one-dimensional relationship with distance from coast/colony, which itself assumes the same relationship in all directions from the coast/colony.

40Á Bathymetry data was available for the whole study area and derived from the GEBCO Database (IOC, IHO, and BODC 2003).

41Á Sea surface temperature was accessed from the Group for High Resolution Sea Surface Temperature ("GHRSST") Level 3 collated North Atlantic Regional Subskin Sea Surface Temperature from the Visible Infrared Imaging Radiometer Suite ("VIIRS") on the Suomi NPP satellite (GDS version 2) from the Ocean and Sea Ice Satellite Application Facility ("OSISAF") of the European Organization for the Exploitation of Meteorological Satellites ("EUMETSAT") (EUMETSAT/OSI SAF, 2015). SST is retrieved from the AVHRR and VIIRS infrared channels using a multispectral algorithm. This product is delivered as four six hourly collated files per day on a regular 2km grid. Two possible covariates were calculated from these data: mean SST across the 6 days prior to each survey being undertaken, and the standard deviation ("SD") of the SST across the same 6 days. This has previously been shown to be an explanatory variable in the spatial distribution of shags, albeit at a larger spatial scale (Virgili, 2014).

42Á Habitat variables were available using the EUNIS ("European Nature Information System") marine habitat codes. Data at two hierarchical levels were available. Level 2 categories are:

- Á A1: Littoral rock and other hard substrata;
- Á A2: Littoral sediment;
- Á A3: Infralittoral rock and other hard substrata;
- Á A4: Circalittoral rock and other hard substrata;
- Á A5: Sublittoral sediment;
- Á A6: Deep-sea bed;
- Á A7: Pelagic water column; and
- Á A8: Ice-associated marine habitats.

43Á   Level three categories are more detailed at the next level of habitat type. There are a total of 56 EUNIS level three habitat codes for the marine environment. These are not provided here but can be found here: http://eunis.eea.europa.eu/habitats-code-browser.jsp?expand=#level_A

2.4ˑ GPS logger tracking

44Á   GPS logger data from shags tagged on the Isles of Scilly were obtained from the Royal Society for the Protection of Birds ("RSPB"), as part of the Future of the Atlantic Marine Environment ("FAME") project, and from the University of Exeter.

45Á   Shags were captured, colour-ringed, and equipped with a GPS logger (iGOTu GT-120, Mobile Technology), under licence from the British Trust for Ornithology ("BTO"), from two colonies on the Isles of Scilly. The tags used were manufactured by removing the original casing and sealing them in heat-shrinking plastic, to improve water resistance and reduce weight. Tags were then attached to the back feathers of birds using Tesa tape. The total tag weight was 17 g (<1 % of body weight). Once activated the birds' position was logged every 100 seconds, resulting in a total recording time of about 3 days. Birds were then recaptured to recover and download the data archived on the tag. Data was retrieved from three tags in 2010, seven in 2011 and three in 2012.

46Á   Tracking data was not manipulated and used as is. Every fix was merged with the available covariates. It was considered important that flights of tracked birds were included in the analysis, as the space between the breeding colonies and any important feeding and maintenance areas are included for a highly coastal species, like shag. Thus, any SPA boundary recommended using the tracking data will include the connection between the breeding colonies and the at sea areas used by those birds.

2.4.1ˑ Model specification

47Á   The tracking data were analysed using a case-control approach to provide the 'absences' for the presences supplied by the tracking data. This involved randomly sampling locations from all candidate locations (recorded at some point) along with their corresponding covariate data. These were sampled randomly to achieve an equal number of presences and absences. This followed the recommendations of Barbet-Massin *et al.* (2012). Absences were generated by lifting points off the prediction grid at random. The covariate values associated with those locations were used in the modelling as part of the case control approach. This case control approach renders all but the intercept parameter to be reliable even though we are retrospectively sampling the absences. Since the birds were tracked their location is known at each point in time for that bird, therefore it is also known where birds were absent. A fine grid could have been used and each tracked bird assigned a 1 at the noted location for each time point and allocated zeros to every other point on that grid at that time point (to recognise their absence in those locations). However, that approach would flood the data and model with zeros, which is why a case-control approach was used instead. Note, this approach only works for models which employ a logit link function which was also implemented here.

48Á   To ensure this approach was reasonable, an examination of how the ratio of zeros affected model results was carried out. This focused on the coefficient of variation about each parameter (standard error/coefficient) and in particular any evidence that this stabilises with some proportion of zeros sampled. This metric was chosen since all other measures are no longer comparable when the response

data is changed and for example, cross-validation and pseudo-R^2 scores vary substantially with the response variable modelled.

49Á   Predictions across models with different numbers of zeros sampled can also not be compared due to the successive lowering of the fitted probabilities with the addition of zeros. For this same reason the average and, in particular, the level of predicted probability of presence in each grid cell are also somewhat artificial. What is more meaningful is the probability of presence in each location relative to the probability in other locations.

50Á   After sampling, the presence/absence data were modelled using the CReSS method with targeted smoothing capabilities using SALSA. This method is spatially adaptive and is designed to describe surfaces with both local surface features (e.g. patchy surfaces with locally acting hotspots) and/or global surface features (e.g. at surfaces or far-reaching trends). This approach also incorporates any temporal autocorrelation and returns adjusted standard errors if residual correlation is observed within panels (deemed to be measurements collected on the same day).

51Á   For the one dimensional covariates, we permitted nonlinearities on the link scale, alongside a flexible two-dimensional smoother based function for the spatial element. In this case a Binomial-based model was fitted with good success.

2.4.2˙ Model selection

52Á   Model selection (including the choice of surface flexibility and the covariates retained in the model) was carried out using the cross-validation scores, since this explicitly measures fitness of the model to data unseen by the model. The Mean Square error (the average squared difference between the fitted probability and the binary outcome) was used as the comparative measure. Spatio-temporal residual correlation was assessed using an ACF plot and should this be seen then robust standard errors were employed.

53Á   The probability of presence surface given a set of covariates was returned alongside confidence intervals for these predicted probabilities.

2.4.3˙ Model covariates

54Á   In addition to the model covariates used in the density surface modelling, distance to coast and distance to colony were also used in the predictive modelling of the tracking data. The tracking data also provided information on the breeding status of all the birds tracked. Only two birds, both tracked in 2011, were not breeding adults. One bird was a known immature bird and the other was of unknown breeding status. Since data from these individuals provided 3,903 fixes, it was considered important to retain them but, with breeding status as a categorical variable in the model.

2.5˙ Non-targeted visual boat-based and aerial survey data

55Á   Data from SeaMaST (Bradbury *et al.* 2014) was mapped for comparison with the other three data sources above. These data were analysed from existing ESAS and visual aerial survey data. Data were Distance corrected and mapped using a Density Surface Mapping ("DSM") approach. Further details are available in WWT Consulting (2013). Due to the availability of data the mapping was at a much coarser resolution (3 km x 3 km) than the maximum prediction grid used for the other analyses here (1 km x 1 km).

2.6˙ Boundary determination

56Á KDE outputs using only non-flying shags were prepared for the years of 2014 and 2015 and these were considered to represent the best equivalent to modelled DAS data and modelled tracking data for shags which in turn were focussed on predicting the foraging distribution of shags around the islands. The KDE analysis generated a grid of cells of variable size and shape, but generally around 0.2km – 0.3km in dimension. The mean density of all cells from both years combined in which the centre occurred in square grid cells of standard 0.5km x 0.5km dimension was calculated and this was used for a single maximum curvature analysis representing both years.

57Á The DSM outputs generated predictions of density and the lower and upper 95% confidence intervals ("CIs") for a range of covariate parameters (year, month, depth, standard deviation of SST) in grid cells at a scale of 1km x 1km. We calculated the average density of sitting shags for each grid cell as well as the lower and upper CIs, although only the average density was used for maximum curvature analysis.

58Á The outputs from the tracking data models also generated grids of 1km x 1km dimension containing predictions of probability of presence with 95% CIs of the probability for covariates used in the modelling (month, year and breeding status). We calculated the mean predicted probability and the mean lower and upper CIs of the probability in each grid cell and did this separately for breeding tagged birds only and for all tagged birds separately. Only the mean probability was used for the maximum curvature analysis. We did not use any model outputs which included the EUNIS level 3 habitat covariate because of the inherent unreliability of the outputs (see Section 3.2.1).

59Á After sorting the grid from highest mean density or probability to lowest, the cumulative number or probability of shags was calculated as the density x area of grid cell for each grid cell. The cumulative area was also calculated from the highest density grid cell to the lowest. The cumulative number was expressed as a proportion of the total number and the cumulative area as a proportion of the total area for the workings of the maximum curvature analysis.

2.6.1˙ Maximum curvature

60Á The maximum curvature of the relationship between the cumulative number or probability of presence was plotted against the cumulative size of area. The point of maximum curvature was calculated using the single exponential and the double exponential formulae given by O'Brien et al. (2012). The density or probability of the cell at which the point of maximum curvature occurred was used as the minimum threshold for selecting cells that should be contained within a marine SPA proposal. Other, more complex formulae were considered and discarded primarily on grounds of sufficiency of the existing formula options used, but because the alternatives were over-complex and in most cases, inappropriate.

2.6.2˙ Determining options for a marine SPA proposal

61Á The data analysis provided a range of options for a marine SPA proposal for feeding and maintenance areas used by shags during the breeding season. While this focus on feeding areas was necessary to use habitat covariates to predict all of the areas used by shags, it does not address all of the protective requirements of shags nesting on the islands. The final step in the process therefore was to provide a connection between the breeding sites and the feeding areas by using a convex hull analysis around all the breeding sites and all of the selected feeding areas for each option. A convex hull is a polygonal area that is of smallest length and so that any pair of points within the area have the line segment

between them contained entirely inside the area. This was performed using ET GeoWizards v11.3 in ArcGIS 10.4. A shapefile of points, consisting of all colony locations (using data supplied by NE from 2015 and 2016) and the centroid of all selected grid cells, then a buffer set around these of 707m (the distance from the centroid of each 1km square to each corner) which was considered sufficient to capture variation in the flight lines of shags between their breeding sites and feeding areas. This approach was used because it was found that there was a considerable amount of overlap in usage of feeding areas used by shags tracked from different breeding sites (Evans *et al.* 2016).

3.   Results

3.1   Digital aerial survey data

62Á   The total number of all shags observed from the DAS data are shown in Table 1. The raw observations of shags recorded interacting with the water surface (sitting, diving or taking off/landing) from digital aerial surveys are shown in Figure 2. These data suggest a highly coastal distribution with more observations in the south-western waters of the archipelago than in other areas. These data were the basis for the KDE and DSM analysis.

Table 1 Raw counts of all shags recorded in each month of survey effort in 2014 and 2015

Year	May		June		July	
	Sitting	**Total**	**Sitting**	**Total**	**Sitting**	**Total**
2014	332	449	82	134	60	145
2015	37	47	105	143	8	176

Figure 2   Observation of sitting shags from digital aerial surveys in 2014 and 2015 of the Isles of Scilly. Sitting shags includes birds diving and taking off or landing on the water surface. Note that this is only locational information and does not indicate abundance.
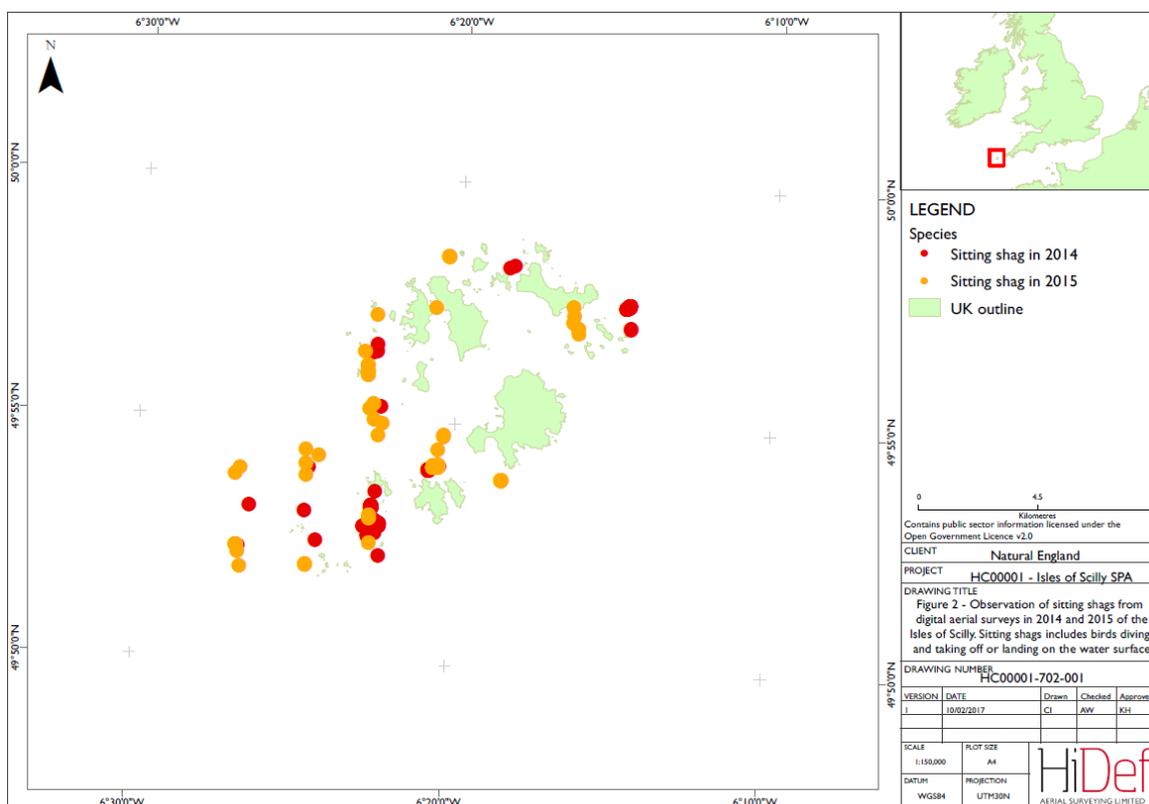
3.2 Kernel Density Estimation

63Á   KDE maps of predicted shag distribution were plotted separately for 2014 (Figure 3) and 2015 (Figure 4). These showed a difference in shag densities between the years, and different spatial distributions around the islands. In 2014 abundances were higher and proportionally more birds were recorded in the north-east of the archipelago than in 2015. As well as predicting more shags within the survey area in 2014 the KDEs also predicted birds being more coastal in 2014 than in 2015. This is likely a result of a fewer shags being recorded in 2015 than a reflection of true shag spatial abundance, as the raw data did not record any sitting birds in the areas beyond the coastal limits of the islands in either year. However, boundaries were determined from the combined data set, taking account of the year of survey.

Figure 3   KDE results for sitting shags in 2014
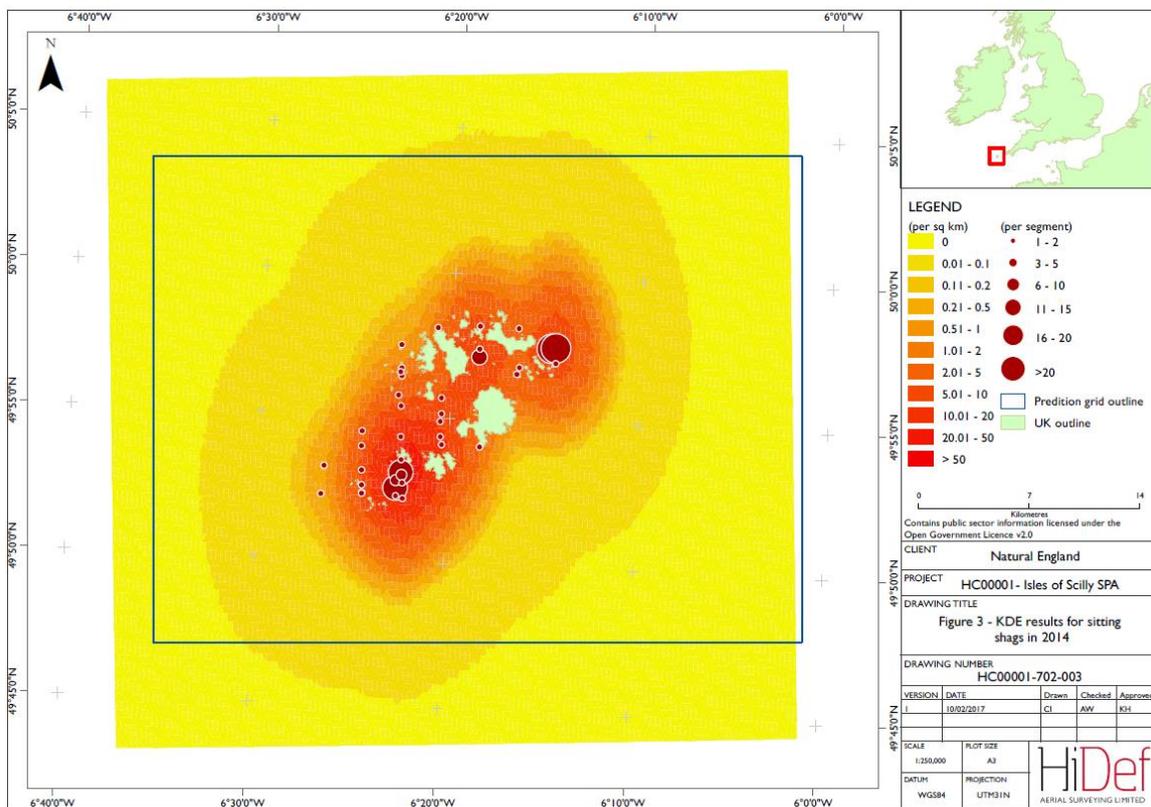
Figure 4   KDE results for sitting shags in 2015
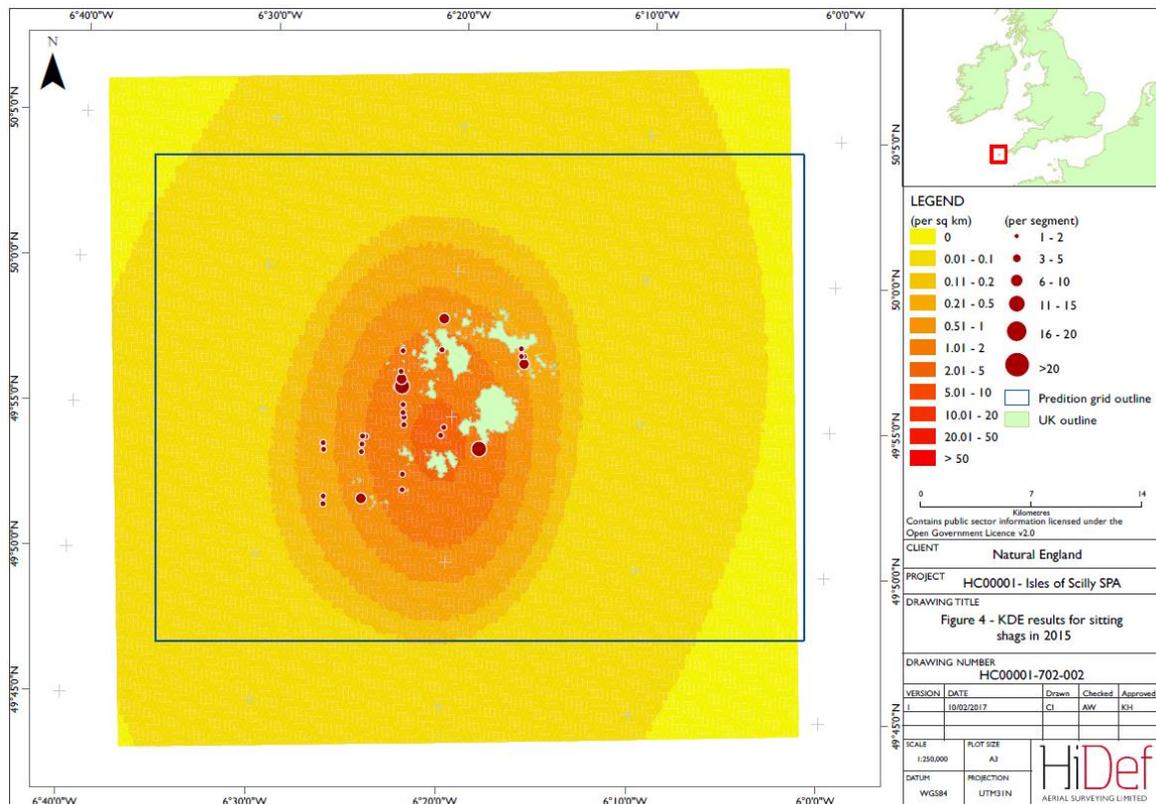
3.3˙   Density Surface Modelling

3.3.1˙ Covariates

64Á   The DSM data were segmented into 500m x 500m grid cells for a given time point (0.25 km^2) and so they constitute the observations. Depth information was present for all observations (n = 11,432) however the data set was reduced to 7,375 rows when either mean SST or the SD of SST was considered due to missing values. While including SST information in the model significantly reduced the number of rows available there was no reduction in the total number of birds recorded since bird numbers were universally zero when SST was recorded as 'not available'. By considering covariates in these models that have missing values, many zero observations are "lost". However, unless there is some reason to assume that the zeros for the SST data will unduly change the relationships for the other covariates (e.g. depth) then we don't believe there is any bias in the results based on models which include SST information. The range of the depth values was slightly smaller when SST information was either included or excluded. The deepest waters were 97.79m in the larger data set and 90.25 in the reduced size data set but the plot for the depth relationship in both cases were indistinguishable from each other. Additionally, the values at this upper end of the depth relationship were universally zero, so there were no undue effects from this 'loss' of information. It was not practical to fit both mean SST (s(meanSST)) and the standard deviation of SST in a model simultaneously (s(sdSST)) since collinearity was evidenced by a high variance inflation factor (VIF = 7.730). For this reason, model selection for these covariates proceeded on a parallel basis and the results compared to evaluate which, if any, form the SST information should appear in the model.

65Á   Habitat variables from both EUNIS level 2 and level 3 information were tested. For level 2 data missing values resulted in a smaller modelling data set. Additionally, bird numbers were universally zero in some categories of this habitat variable. Specifically, 4,178 observations had no entries for level 2 habitat data (i.e. shags were not recorded at the location, even though there was Level 2 habitat data at those locations) which meant that these rows were necessarily deleted before any models, which include this covariate, were fitted to the data. While this resulted in the loss of 4,178 rows from the data set, a total of only 9 birds were lost when these rows were deleted. Due to universally zero numbers of birds seen for some categories this necessitated an additional data reduction. Specifically, 3 of the 6 categories contained universally zero values (A1, A2 and B3) and so to include these covariates in the model a further 20 rows of data were removed. For this reason, inclusion of this covariate was considered after the selection of yearMonth, s(depth), the spatial term (s(x,y)) and either mean SST (s(meanSST)) or the standard deviation of SST (s(sdSST)) – if either SST information was retained in the model.

66Á   EUNIS level 3 data resulted in a drastically reduced modelling data set due to missing values and more importantly universally zero values in some categories of this habitat variable. Specifically, 4,178 observations had no entries which meant that these rows were necessarily deleted before any models were fitted to include this covariate. While this resulted in the loss of 4,178 rows from the data set, a total of only 9 birds were lost when these rows were deleted. In addition to this, 9 of the 17 categories contained universally zero values and so fitting this covariate in the model meant a total of 4,086 rows were further removed. Models were fitted with and without this variable and the CV scores compared for objective comparison. The CV score did improve when the EUNIS level 3 variable was included. However, there were combinations of covariates in the prediction grid (which were unseen in the modelling data) which returned unacceptably high predictions in those areas and for this reason the model including this variable was not considered further. While this did not change the total number of birds seen in this data set, removal of these points resulted in a model with much reduced (or weakened) spatial support since the coverage of the data when these rows were deleted were notable. For this reason, inclusion of this covariate was considered after the selection of depth, the spatial term and either mean SST or the standard deviation of SST (if either were retained in a model).

3.3.2ˑ Final Model

67Á   The final model retained YearMonth, a smooth term for the spatial coordinates (with 8 parameters) and smooth functions for both depth and sdSST (Table 2). These were also statistically significant at the 6% level (note the p-value for sdSST is 0.05005). The p-value of 0.05005 is extremely close to the 5% significance threshold and while the spatial extent was reduced in this case by using the SST information the model would be expected to report universally zero values in the excluded locations in any case (or very close to zero since a value of exactly zero is never reported for models of this sort). Despite having few parameters, this model captured the main spatial patterns in the data and the diagnostics reveal no concerns about the adequacy of the covariate relationship (Figure 13 & Figure 14) or the assumed mean-variance relationship (Figure 10). The residuals were not deemed to be significantly correlated within transects (Figure 9 & Figure 8); while the runs test statistic was -35.351 the associated p-value was p<0.00001. This notion of independence was also confirmed visually by the ACF plot (Figure 11).

68Á   The final model selected had the lowest CV score (Table 2). Amongst the other candidate models, the model which returned the lowest CV score included the EUNIS level 3 covariate (Table 3). However, this model returned extreme predictions on the prediction grid for a particular combination of depth

values (~ 20m) and a category of this covariate (the predictions were as high as 411 birds per km^2, while the fitted values were only ever as high as 166 birds per km^2 and were only 1.81 birds per km^2 on average). There was also unrealistically large uncertainty attached to the coefficients for the EUNIS (level 3) habitat variable (Figure 5). This covariate was a promising candidate but as a result had several categories with universally zero values (regarding bird counts). This makes estimation of coefficients for these levels difficult (since they have zero mean) and may have contributed to these unstable predictions to new data. We think that this covariate is worthy of further investigation but is outside the scope of this project.

Table 2   Final model covariates, p-values and CV score

Covariates	p-value
YearMonth	$p = 0.02991$
s(x,y, df=4)	$p < 0.00001$
s(depth, df=3)	$p < 0.00001$
s(sdSST, df=3)	$p = 0.05005$
CV Score= 14.88476	

Table 3   Other candidate model covariates, *p*-values and CV scores

Model	Covariates	p-value	CV Score
1	YearMonth	$p = 0.0593$	17.90672
2	YearMonth	$p = 0.003203$	17.90672
	s(x,y)	$p < 0.00001$	
3	YearMonth	$p = 0.02465$	17.34077
	s(x,y)	$p = 0.00002$	
	s(depth)	$p < 0.00001$	
	s(sdSST)	$p = 0.02066$	
	EUNIS (level 3)	$p < 0.00001$	

Figure 5 Coefficients associated with the **EUNIS (level 3)** habitat covariate and the prohibitively large uncertainties associated with each level in this case. Note, this graphic illustrates the 'best-case' scenario for this covariate, since these standard errors about the coefficients for some levels of this covariate were extremely high, even for a model which only included 'YearMonth' and EUNIS (level 3) (and the partial relationship for the latter is shown here)
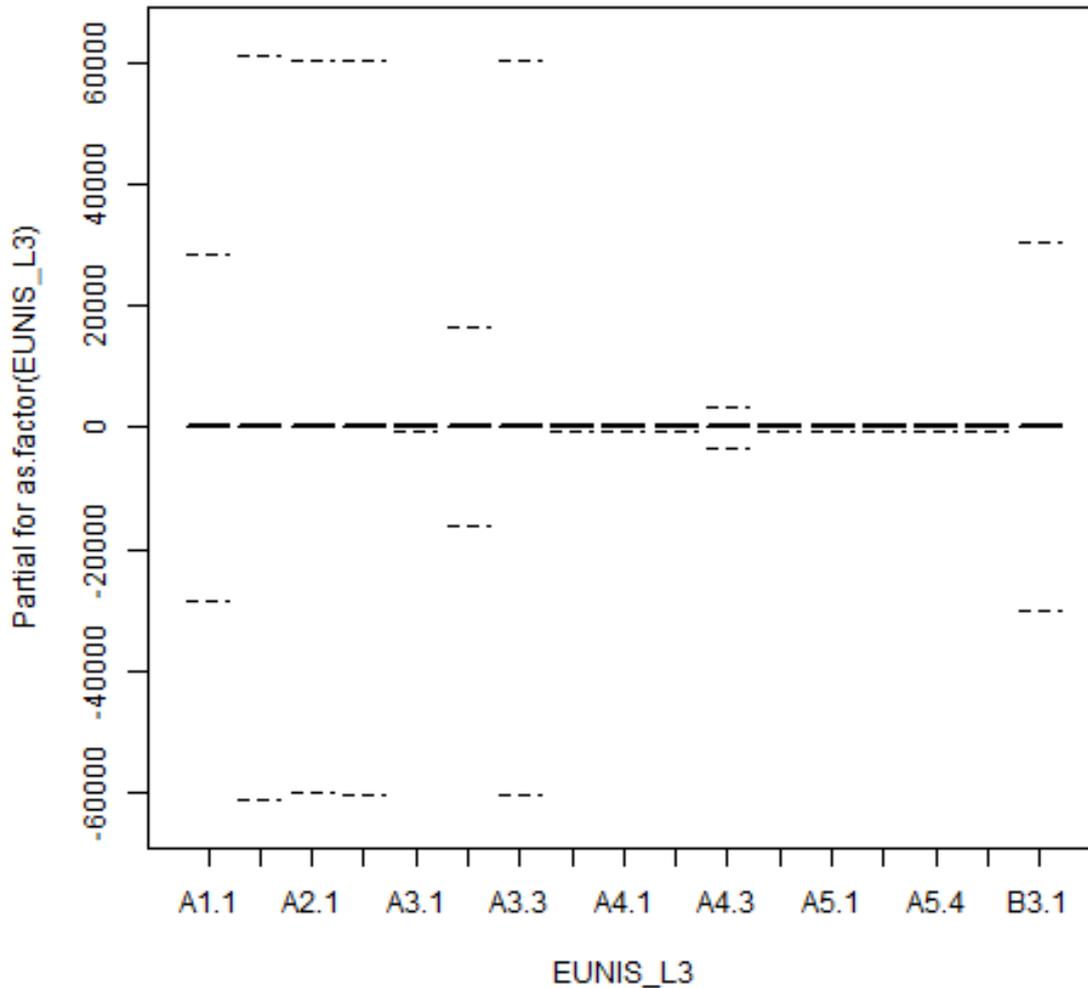
69Á The smooth term for depth and sdSST were required to have few parameters due to universally zero density values observed in some parts of the covariate ranges (Figure 6 & Figure 7). These functions were implemented using quadratic B-splines with a single knot at the mean for each covariate and even despite this simple functional form, the associated uncertainty about the fitted relationships resulting from observing a very large numbers of zero for some covariate values and just a few very large numbers at these covariate values can be seen in the partial fit plots (Figure 13 and Figure 14).

Figure 6   Observed density values across the depth range
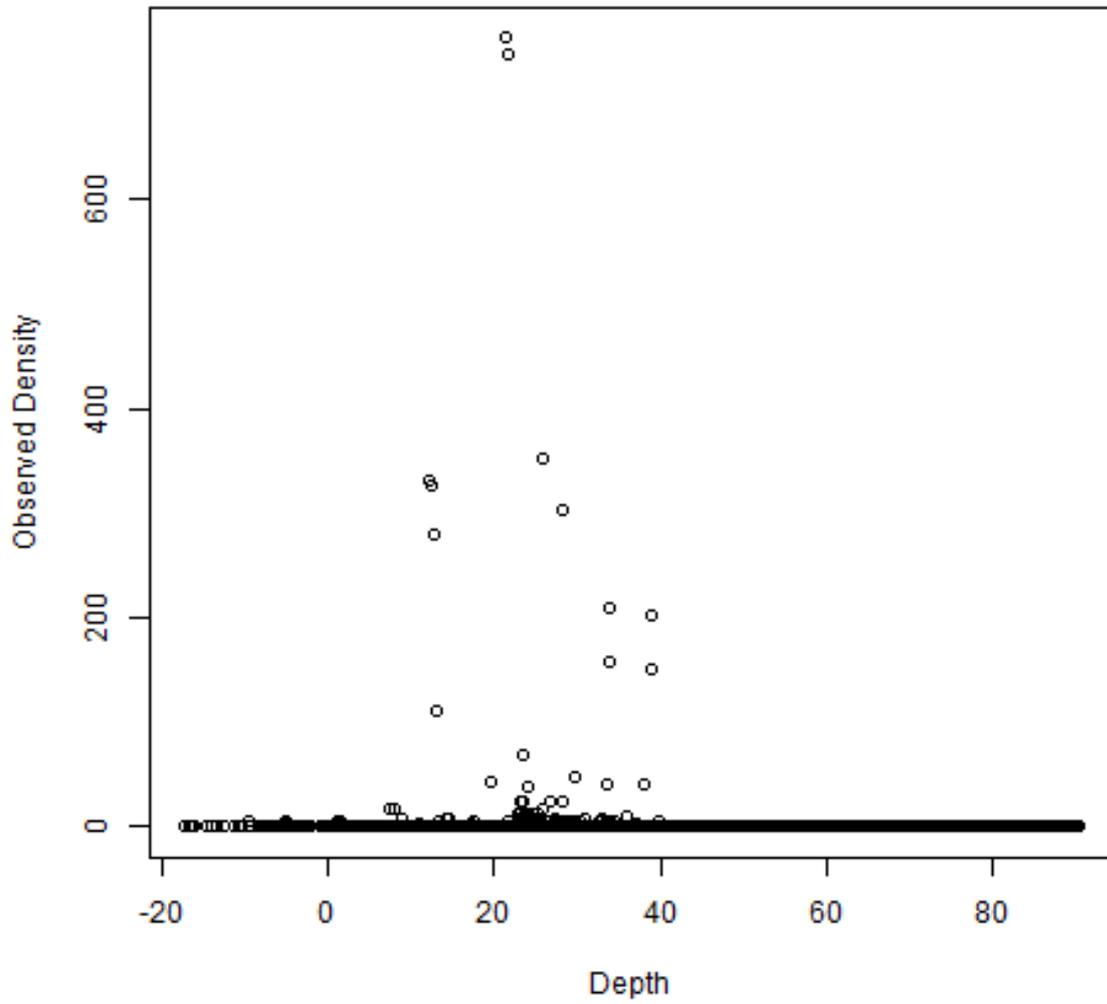
Figure 7   Observed density values across the sdSST range
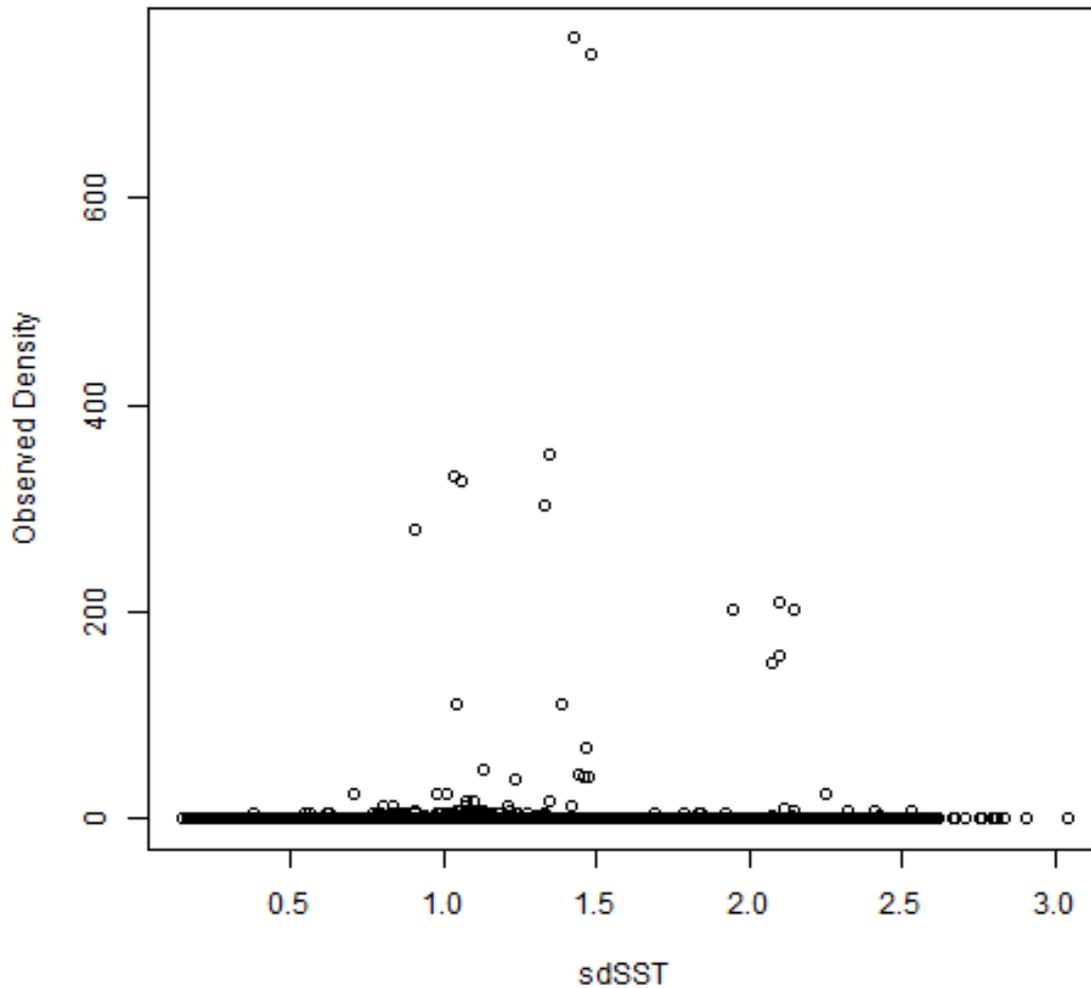
3.3.3˙ Model diagnostics

70Á   There were no concerning patterns apparent in the model residuals represented spatially. This is clearest when examining the deviance residuals (Figure 8). These residuals are closer to being Normally distributed and less affected by one or more very large values (as clear in the Pearson's residuals viewed spatially, Figure 9).

Figure 8 **Deviance residuals represented spatially. A good mix of positive and negative residuals indicates no systematic under-fitting issues**

Deviance residuals
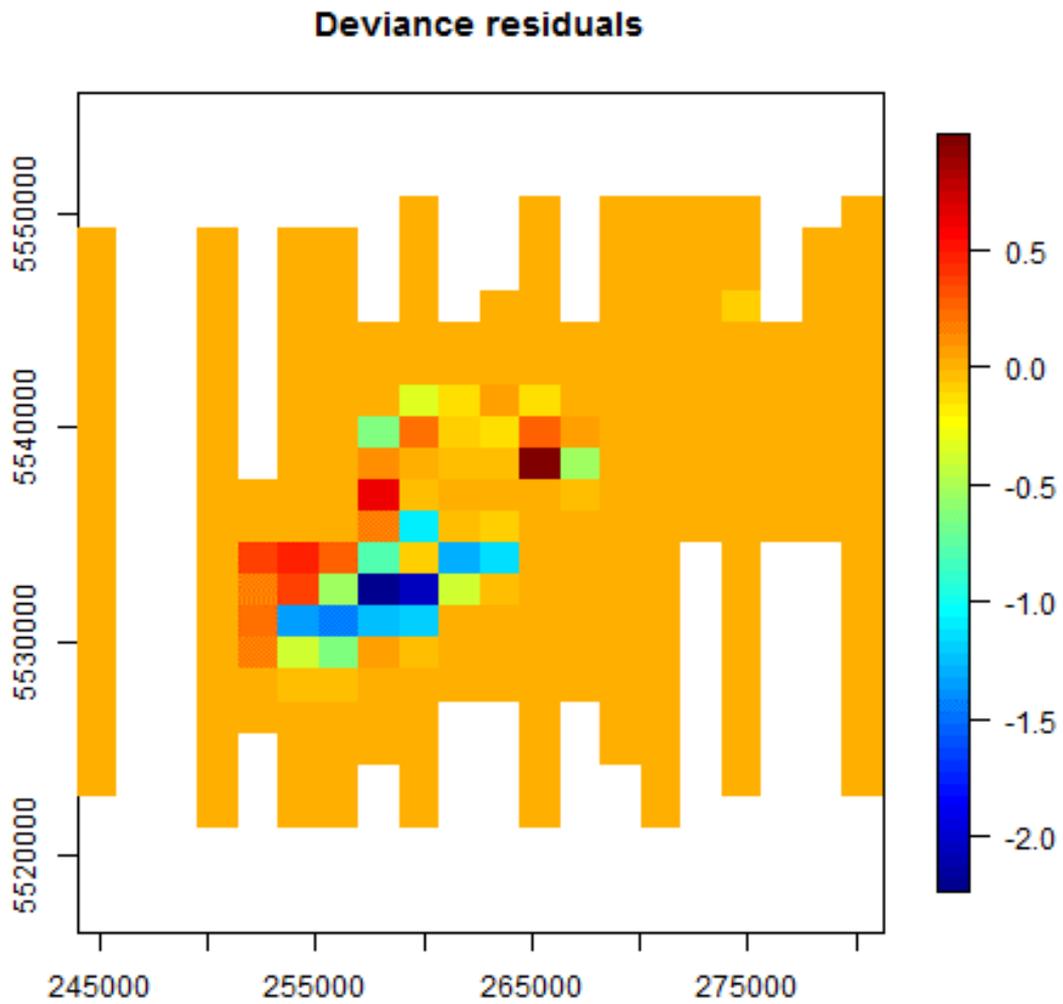
Figure 9 **Pearson's residuals represented spatially. A good mix of differenced size residuals indicates no systematic under-fitting issues.**
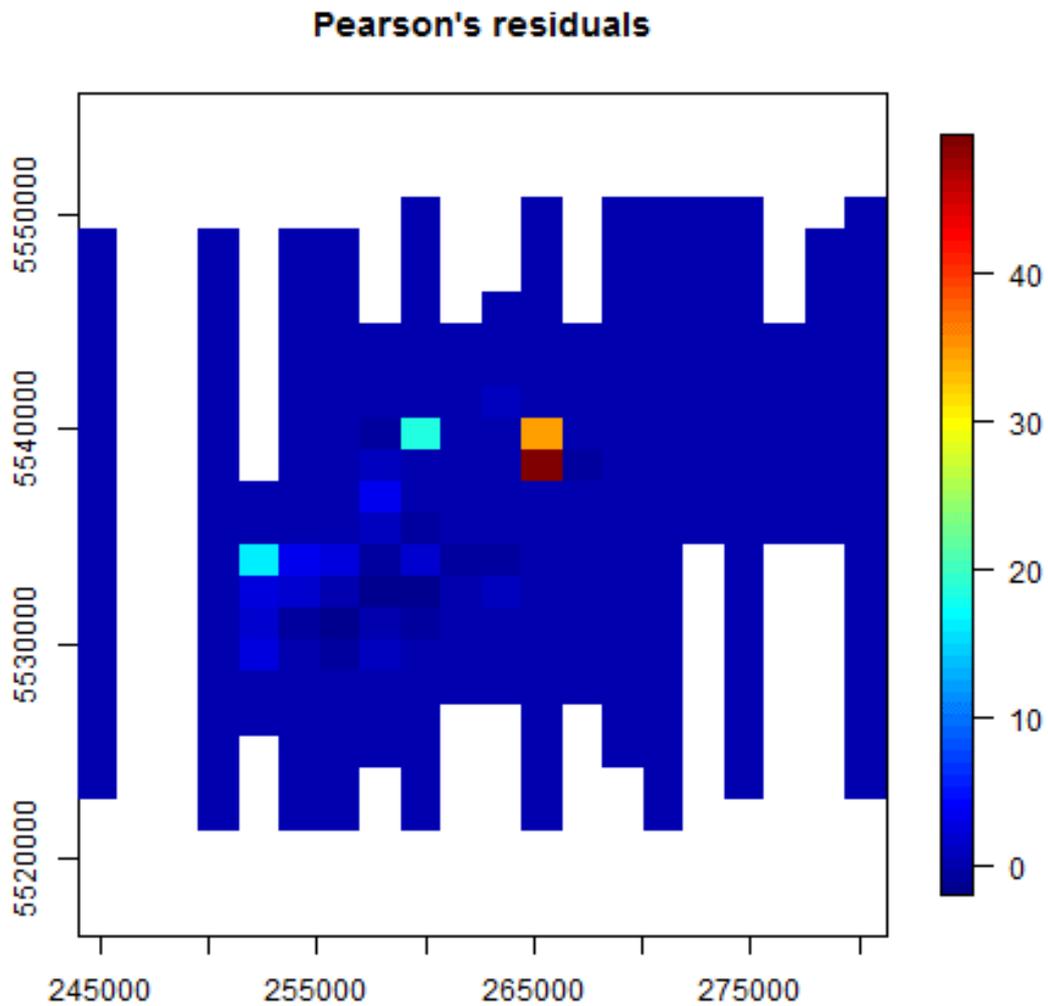
71Á   There was no evidence of any departure from the assumed mean-variance relationship (Figure 10) and there was no evidence for non-independence in model residuals (Figure 11). This latter point was also evidenced by the robust standard errors being very similar to model-based standard errors. Due to the non-significant empirical runs test results, model-based standard errors were used for model inference.

Figure 10   Fitted mean variance relationship; a red line enclosed by the grey envelope indicates a well-fitting mean-variance relationship. The blue line and the black line represent the assumed relationship under the model. The black line is the generating process and the blue line is the observed process returned under this simulation based process (it's a diagnostic of the simulation method rather than a diagnostic of the mean-variance relationship for the particular data-model combination assessed).
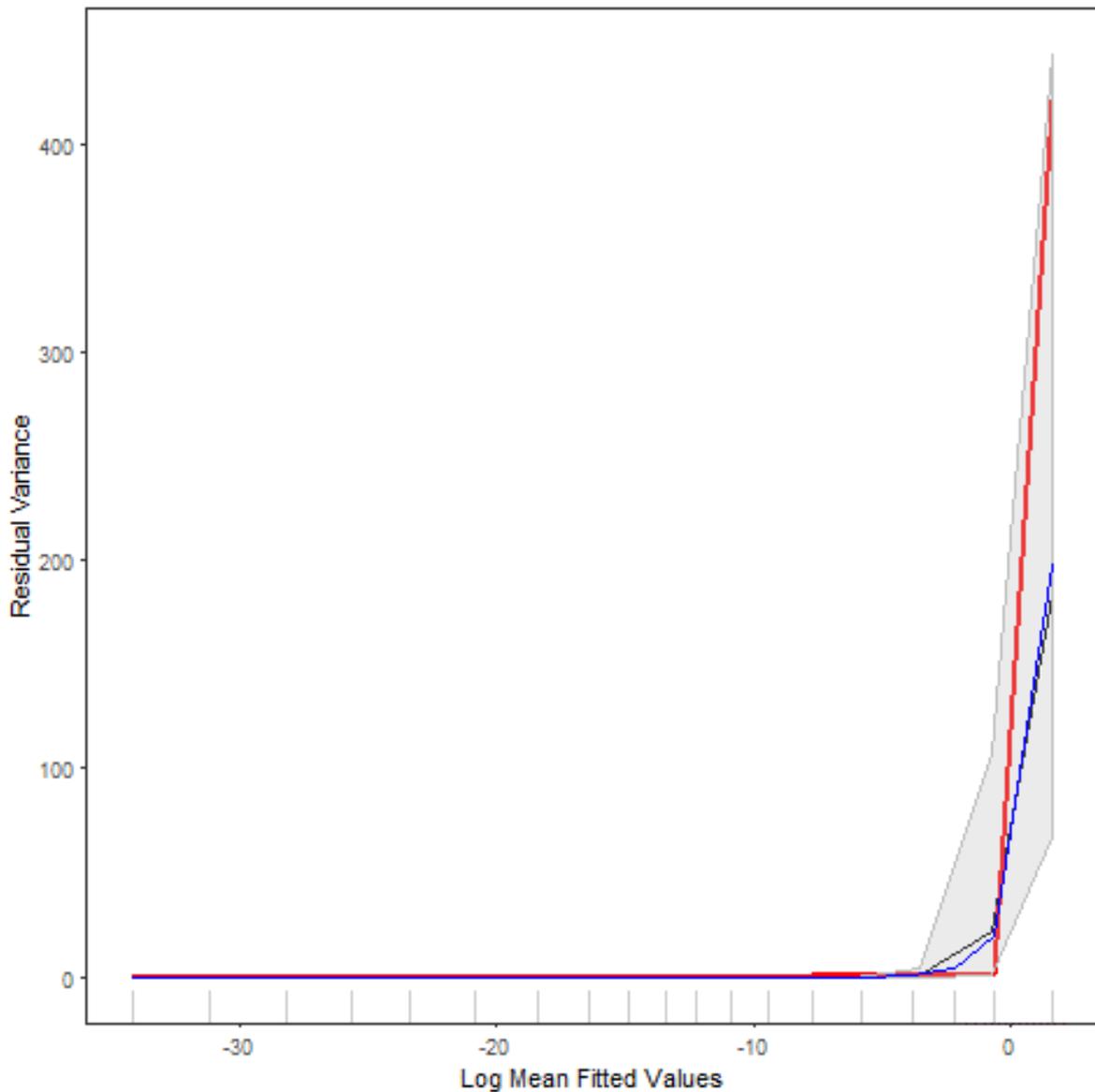
Figure 11   Autocorrelation function plot for the Pearson's residuals. The blue lines represent a 95% confidence interval for independence, which is what the empirical runs test concluded in this case

ACF plot for pearsons residuals
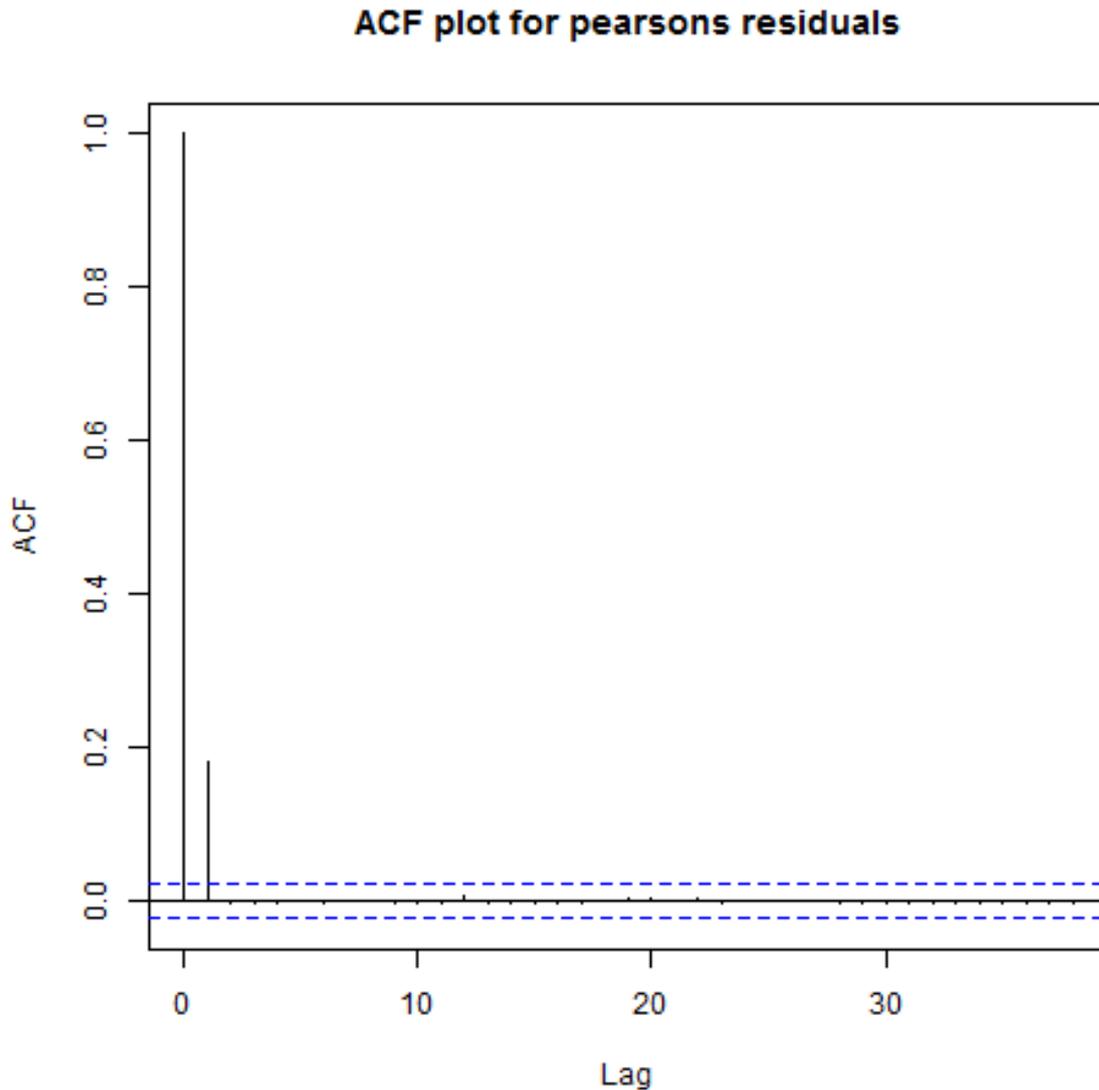
3.3.4 Fitted relationships

72Á   Average bird densities were similar in May 2014, June 2014, May 2015 and June 2015, but were lower in July 2014 and July 2015 compared with May 2014 (the baseline) (Figure 12).

Figure 12   Partial relationship for YearMonth. The May 2014 density was the baseline used
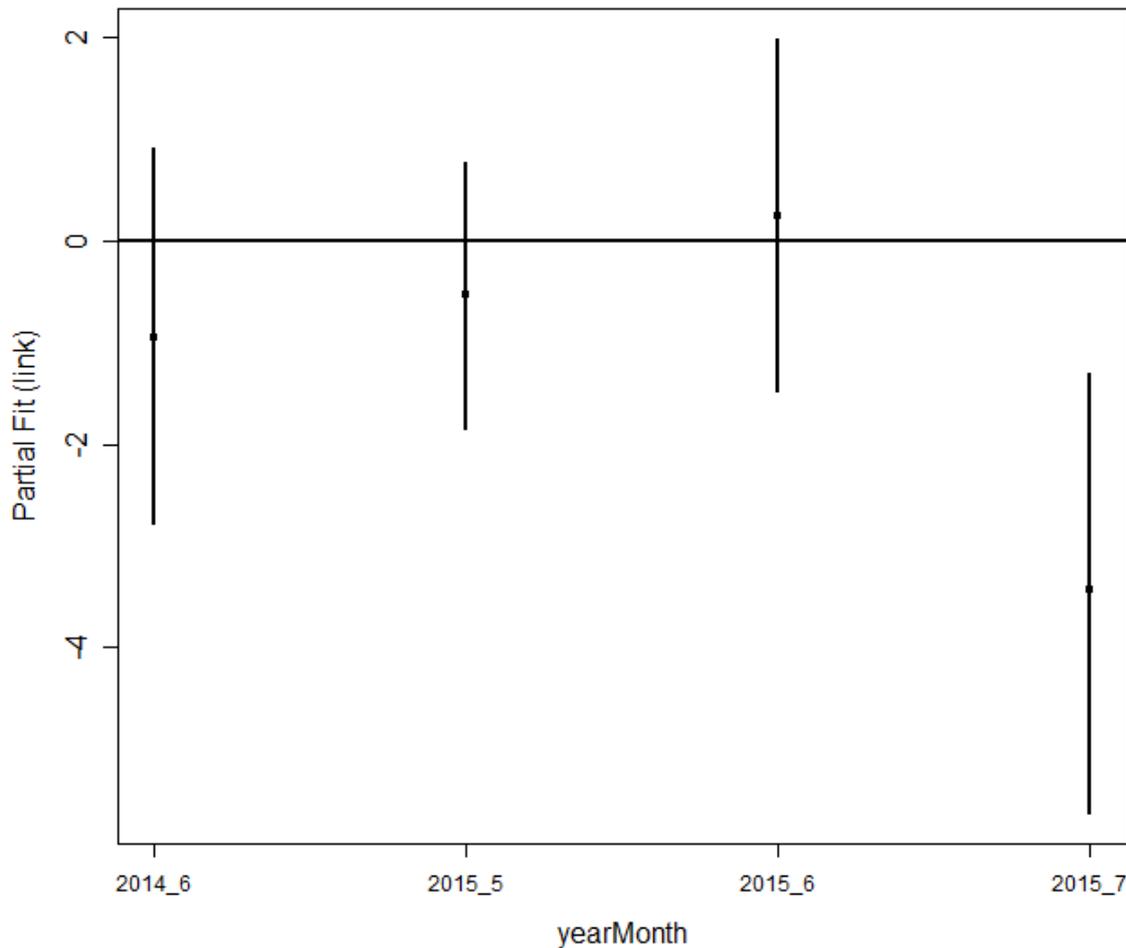
73Á   Bird density was predicted to peak at a depth of about 25 metres, but declined after this point (Figure 13). There was some uncertainty in this relationship, due to universally zero density values observed in some parts of the covariate ranges (see Figure 6 above). There was a great deal of uncertainty around the relationship with sdSST when examining the pointwise 95% confidence intervals (Figure 14), but this was not surprising given the very large range of densities observed for average levels of sdSST (Figure 7). Despite this uncertainty, the p-value associated with this covariate (considering all of the coefficients for s(sdSST) as a collection) was $p < 0.05005$.

Figure 13   Partial relationship for Depth with 95% confidence intervals
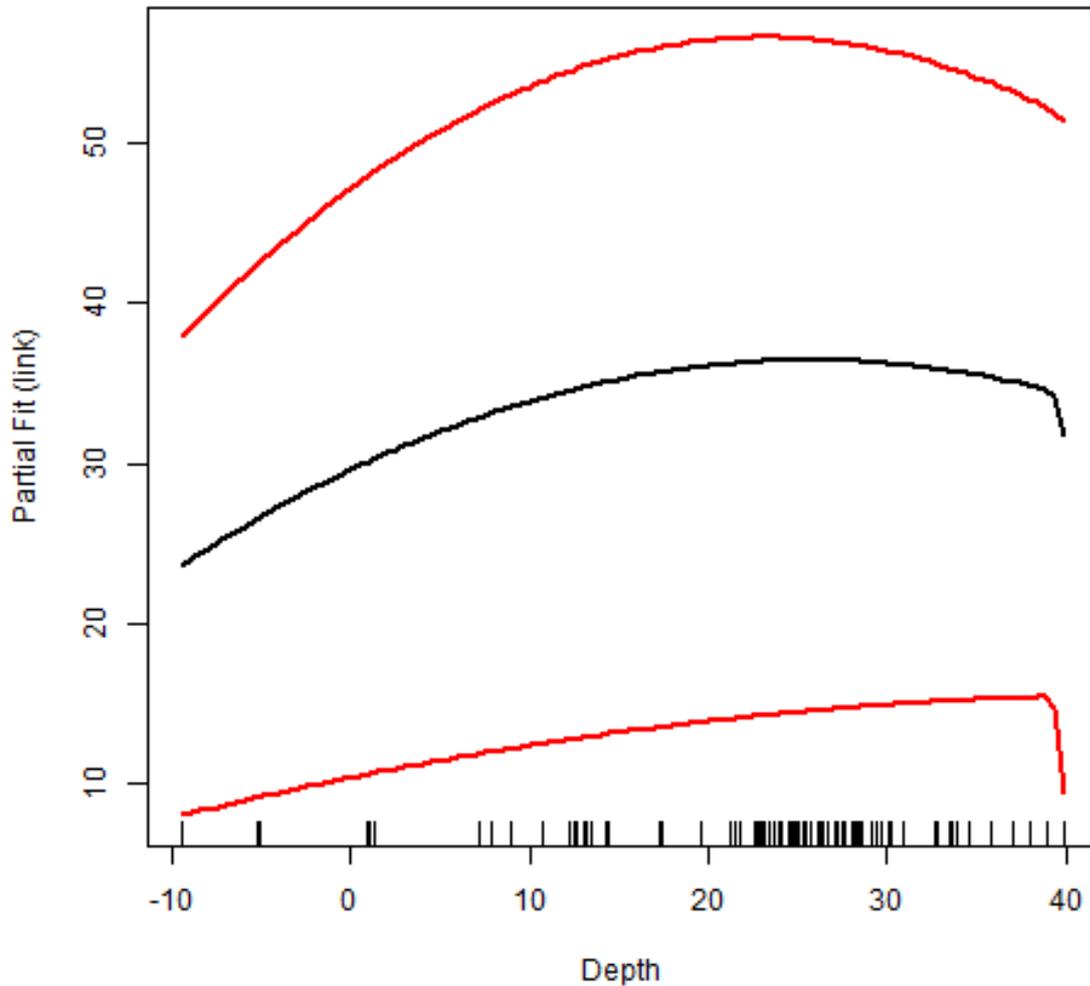
74Á   There was good agreement between the observed densities and the fitted values based on the model (Figure 15) and this is further evidenced when examining the observed densities alongside the fitted values more directly (Figure 16). The model predictions for the entire prediction grid (not just at surveyed locations) are also very well aligned with observed densities (Figure 17 & Figure 18).

Figure 15   Observed density vs fitted density at survey locations.
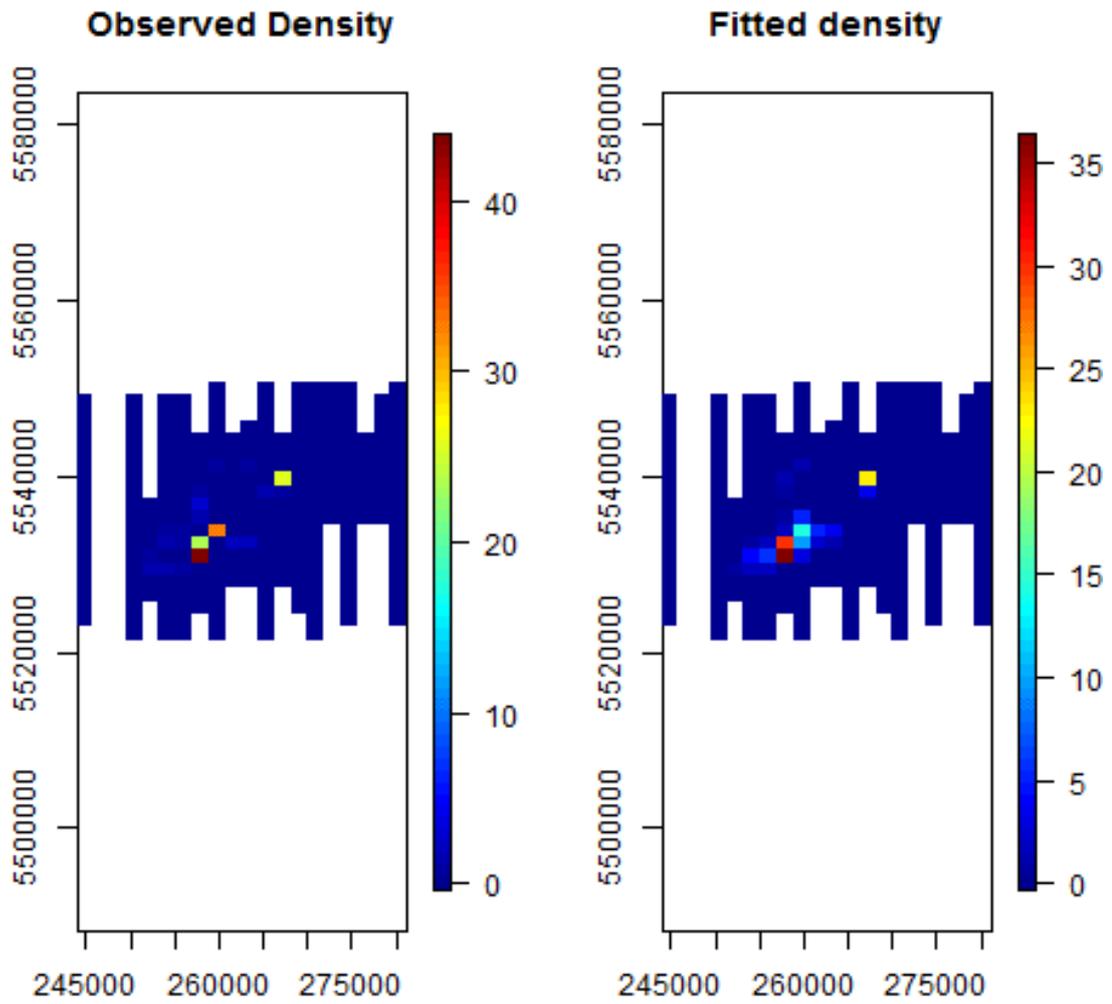
Figure 16 Observed density vs fitted density at survey locations with the observed densities overlaid. The size of the circles are scaled by the log(observed density) for additional information.
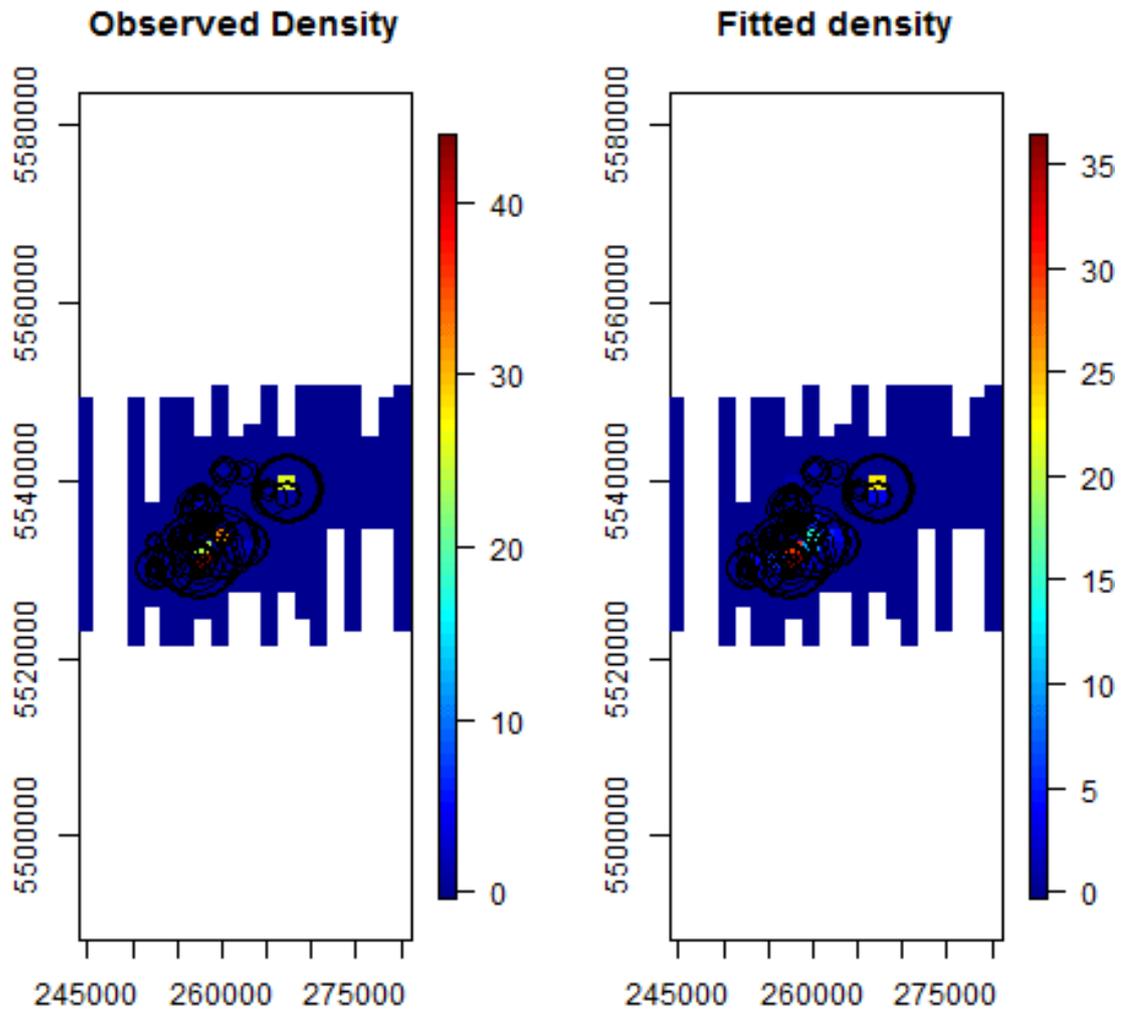
Figure 17   Predicted shag density (birds/km^2) across the prediction grid
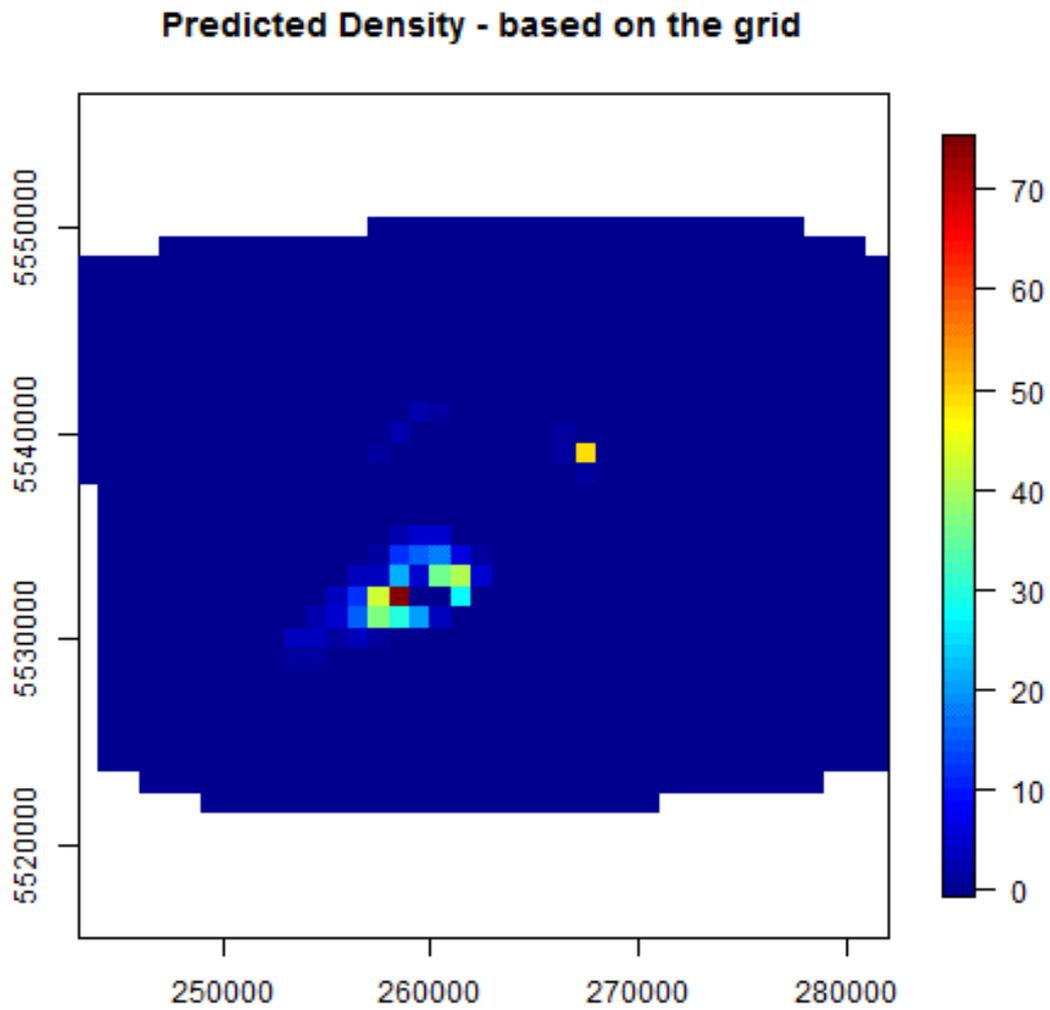
Figure 18 **Predicted density of shags (birds/km^2) across the prediction grid with the observed densities overlaid. The size of the circles was scaled by the log (observed density) for additional information**

Predicted Density - based on the grid
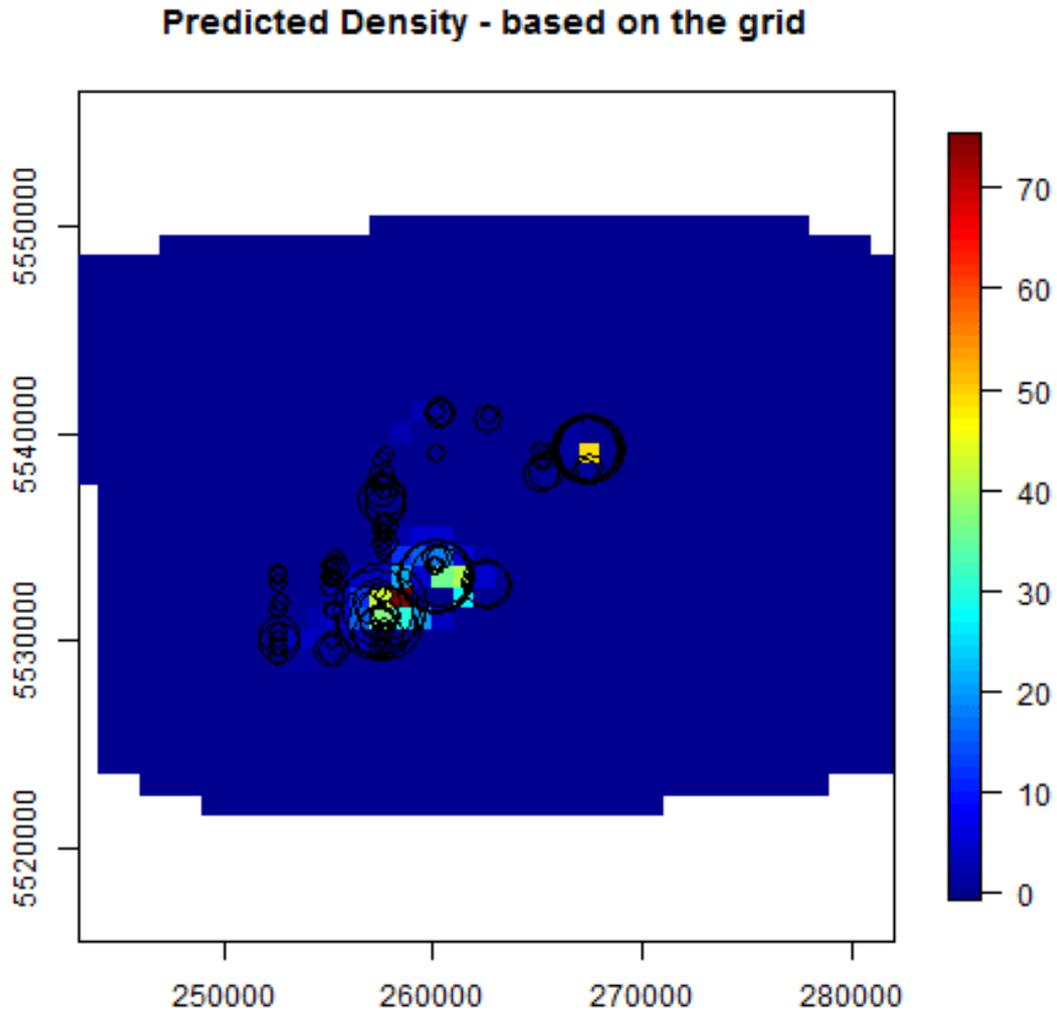
75Á Finally, the predicted densities of sitting shags were mapped relative to the islands in the archipelago (Figure 19). This clearly showed two main hotspots to the south-west and to the north-east, as the KDE data showed in 2014. Unlike the KDE prediction, the DSM predictions were much more coastal and more closely reflect the raw data (Figure 18).

Figure 19 **Predicted density of shags (birds/km²), from the DSM, across the prediction grid in relation to the islands in the archipelago (top figure). The KDE for 2014 (bottom left) and 2015 (bottom right) are provided for reference (Figure 3 and Figure 4 respectively).**
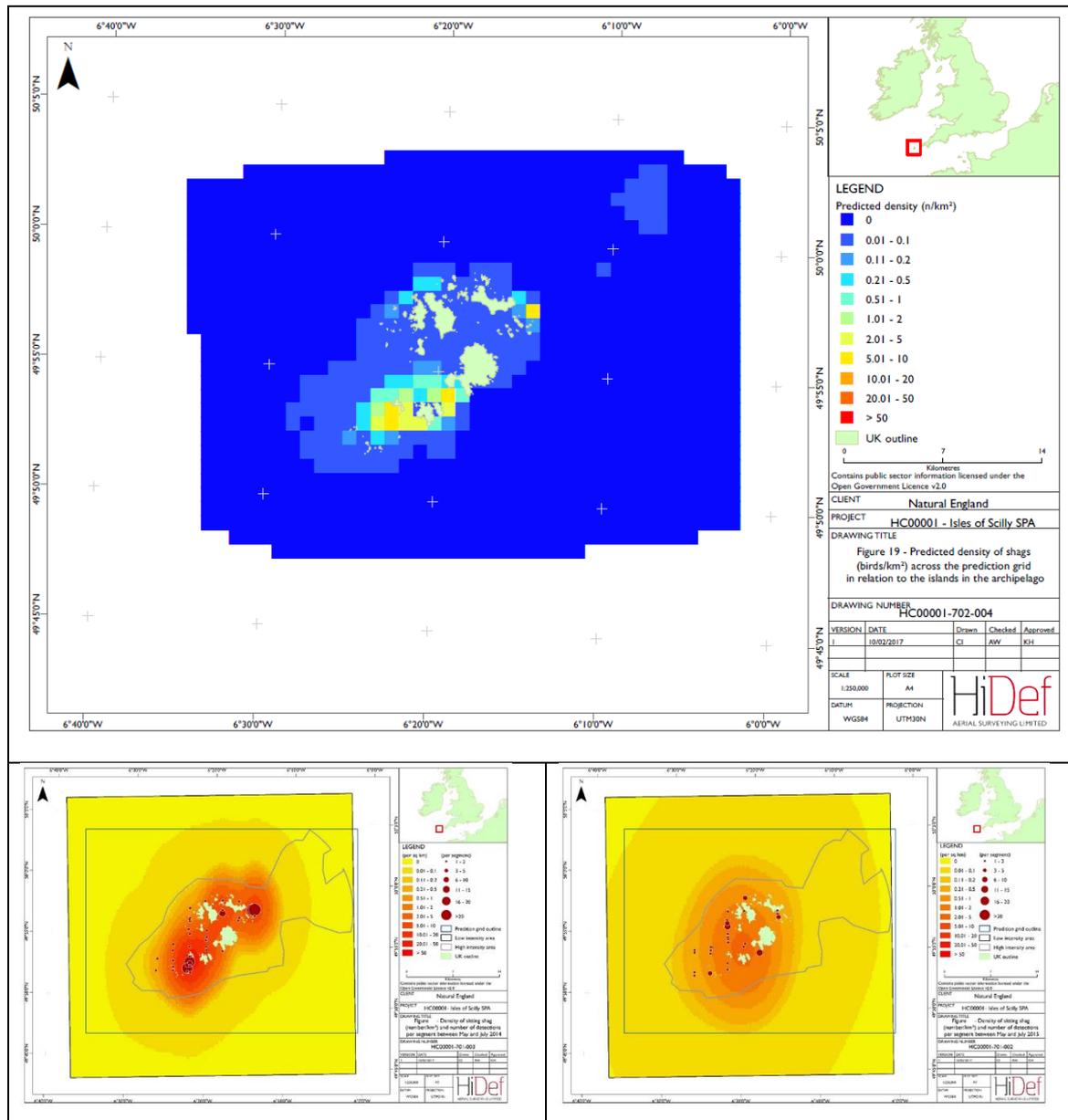
3.4ˈ GPS logger tracking

76Á   Since habitat selection by foraging shags may be particularly valuable to know in selecting most suitable territories for shags breeding in the Isles of Scilly, the EUNIS (level 3) covariate was considered particularly important. However, there were some missing values in the prediction grid for this covariate and the combination of certain categories for this covariate and others resulted in predicted probabilities close to zero and close to one. As such, we proceeded with parallel modelling with and without this covariate.

3.4.1 Results including the habitat covariate

77Á Smooth functions for depth and distance from colony were added to the initial model which included only the factor variables (YearMonth, breeding status and site). However, this resulted in large associated p-values and worse CV scores, until both were dropped from the model (in turn). Specifically, s(Distance from colony) returned a p-value=0.923 while s(depth) returned a p-value of 0.264. For this reason, these covariates were not considered further in this model. Two candidate models were considered further (Table 4).

Table 4   Model selection for the candidate models including EUNIS (level 3) habitat categories

Model	Covariates	p-value	CV Score
1	YearMonth	$p < 0.0001$	0.02236964
	breed	$p < 0.0001$	
	site	$p < 0.0001$	
	EUNIS (level 3)	$p < 0.0001$	
2	YearMonth	$p < 0.0001$	0.01060297
	breed	$p < 0.0001$	
	site	$p < 0.0001$	
	EUNIS (level 3)	$p < 0.00001$	
	s(x,y df=6)	$p < 0.0001$	
	site x s(x,y df=6)	$p < 0.0001$	

78Á In the chosen model, the average probability of presence for breeding birds was significantly higher than the associated values for both immature and unknown categories (Figure 20). There was a great deal of uncertainty associated with the coefficients for all categories of the EUNIS (level 3) covariate (Figure 21). There were missing values in the prediction grid for the EUNIS (level 3) covariate, and the combination of certain categories for this covariate, and others, resulted in predicted probabilities close to zero and close to one. However, this is not problematic where only the relative probabilities are considered. However, in addition, models excluding the EUNIS (level 3) habitat covariate (due to these predictions close to the boundary) were also run and are discussed below.

Figure 20   Partial relationship for the "breed" covariate. Here, breeding birds (B) is the baseline level and is compared with immature (I) and unknown (U)
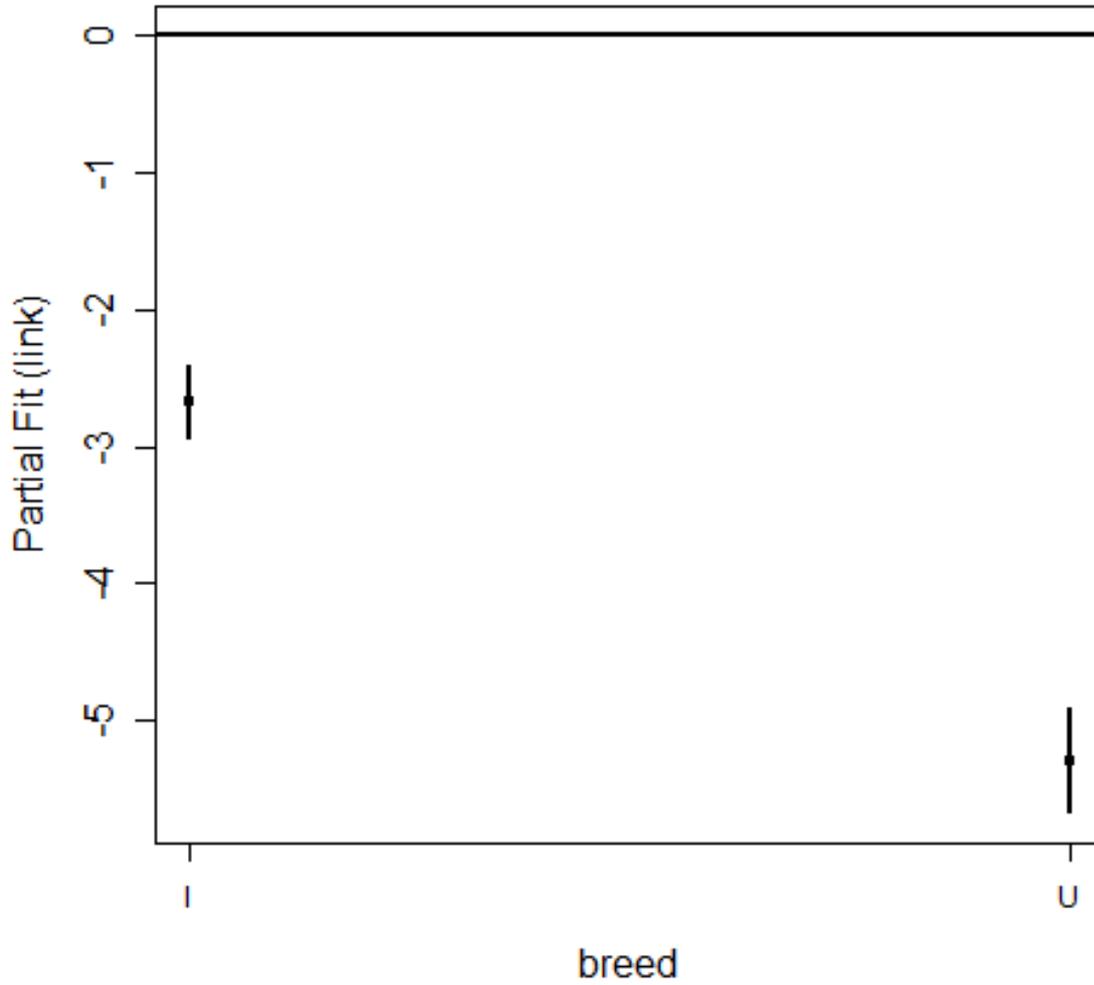
Figure 21 **Partial relationship for the EUNIS (level 3) habitat covariate. Here, the A1.1 level is the baseline**
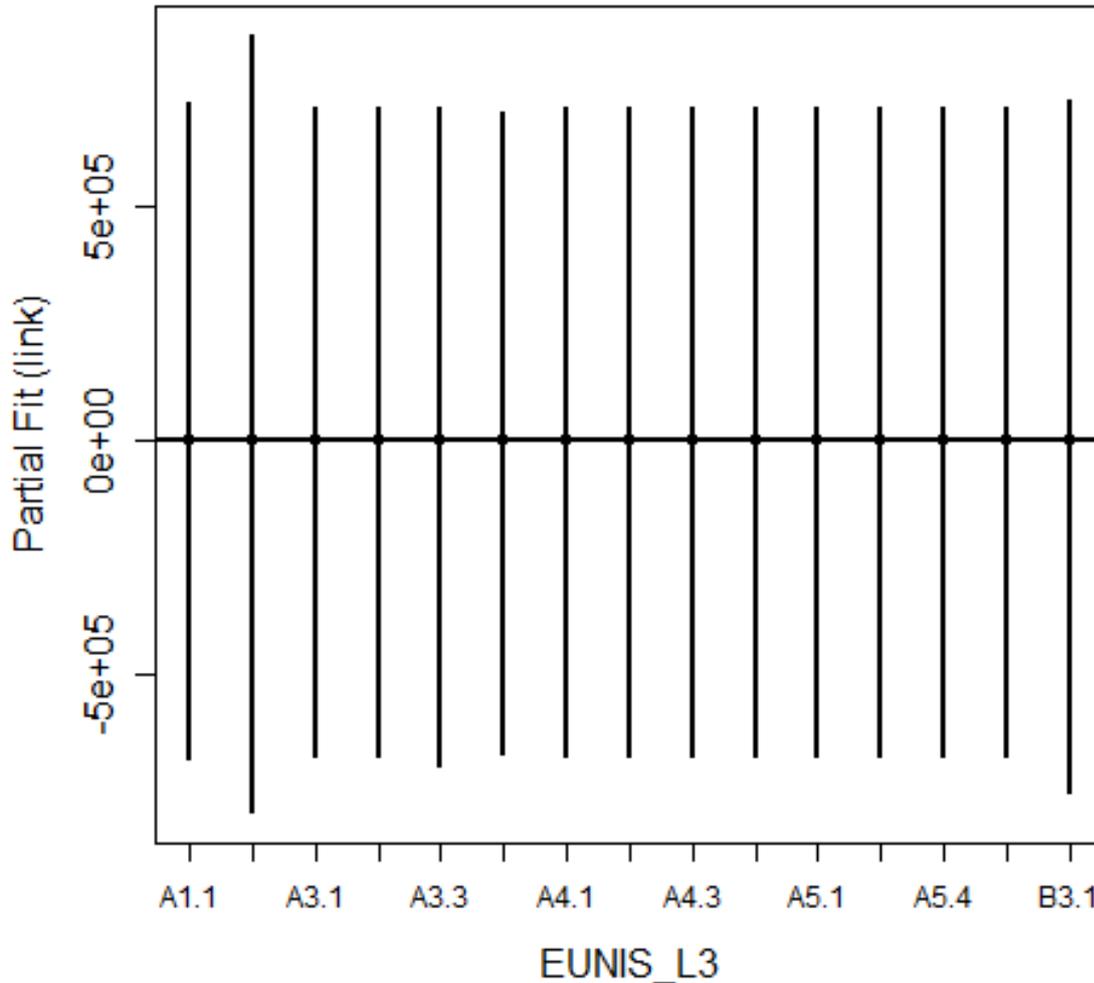
79Á   The predicted probabilities of occurrence across the prediction grid for each site, breeding status and year-month combination were plotted (Figure 22 & Figure 23). In each plot, the colours indicate the relative probability of presence which has its overall level somewhat artificially set based on the pseudo-absences sampled. Consequently, the purpose of these plots is to compare the probability of presence in each location relative to the others. In each plot, the locations at which birds were located given the combination of covariates chosen (i.e. YearMonth, breeding status, and site) for each plot are also displayed as a sense check. The pseudo absences are not shown on each plot but were randomly sampled across the grid. For all plots, the observed presences aligned well with the fitted values and there were no concerning patterns apparent.

Figure 22 **Predicted relative probability of occurrence of breeding shags from tracking from the Annet and Samson colonies. Rows one and two are from the Annet colony and row three is from the Samson colony. Tracking from 2010 is in the left-hand column, tracking from 2011 is in the middle column and tracking from 2012 is in the right-hand column. White cells represent locations where habitat data were absent**

Site:SCI:ANN;Breed=B;Yearmonth:20116

Samson

Site:SCI:SAM;Breed=B;Yearmonth:20105

Site:SCI:SAM;Breed=B;Yearmonth:20115

Site:SCI:SAM;Breed=B;Yearmonth:20125
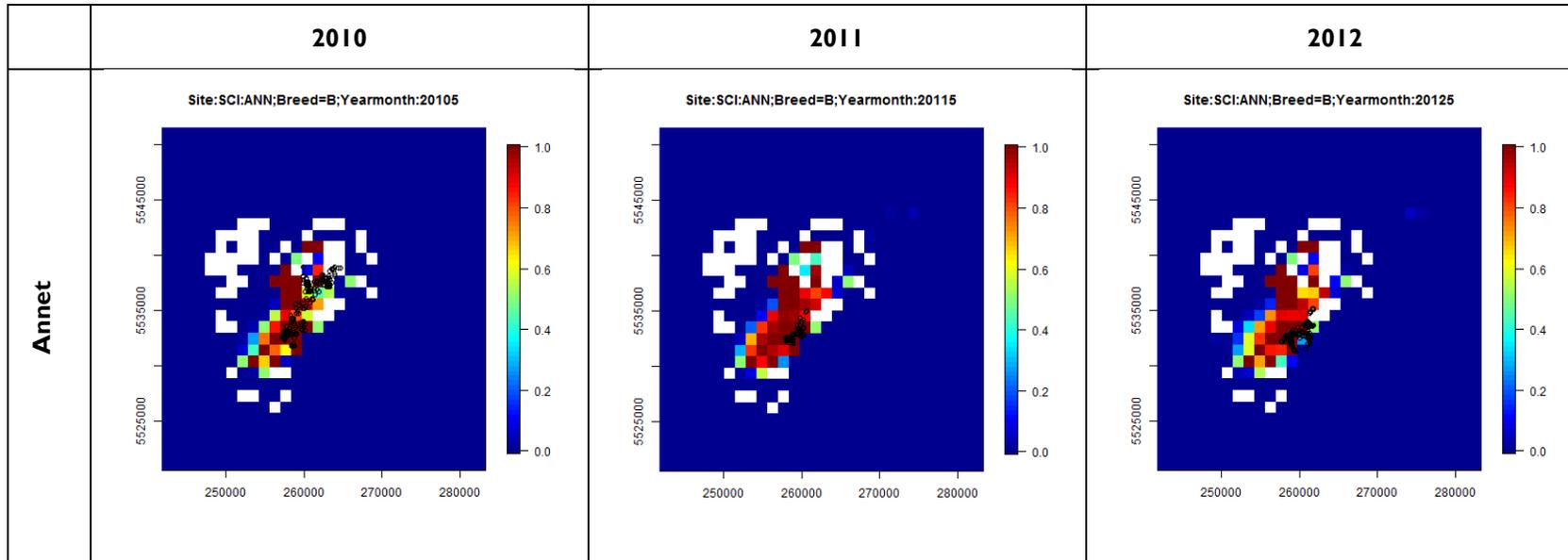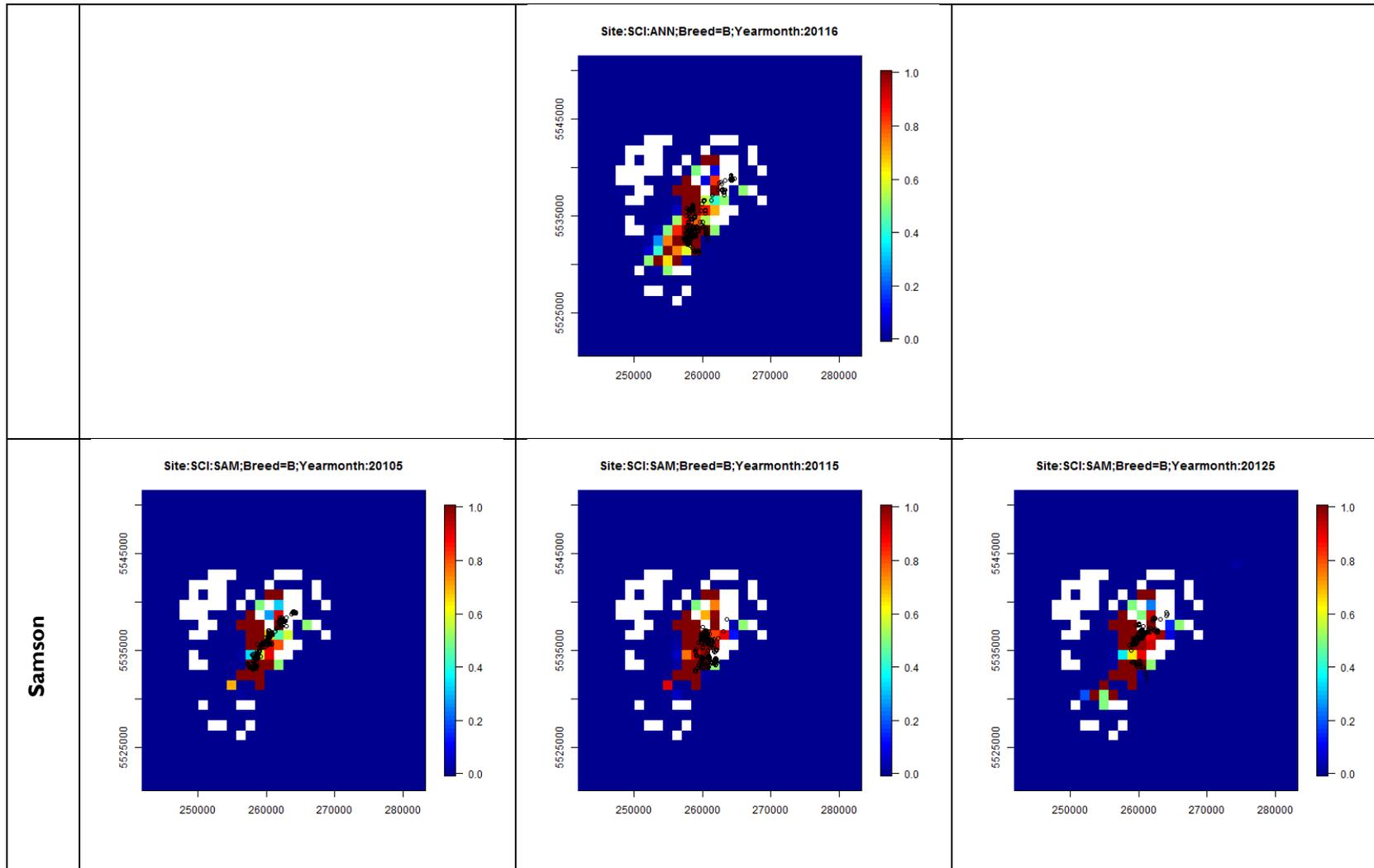
Figure 23 **Predicted relative probability of occurrence of immature shags or shags of unknown breeding status. One bird was tracked from the Annet colony (left) and one bird was tracked from the Samson colony (right). Both birds were tracked in 2011. White cells represent locations where habitat data were absent**
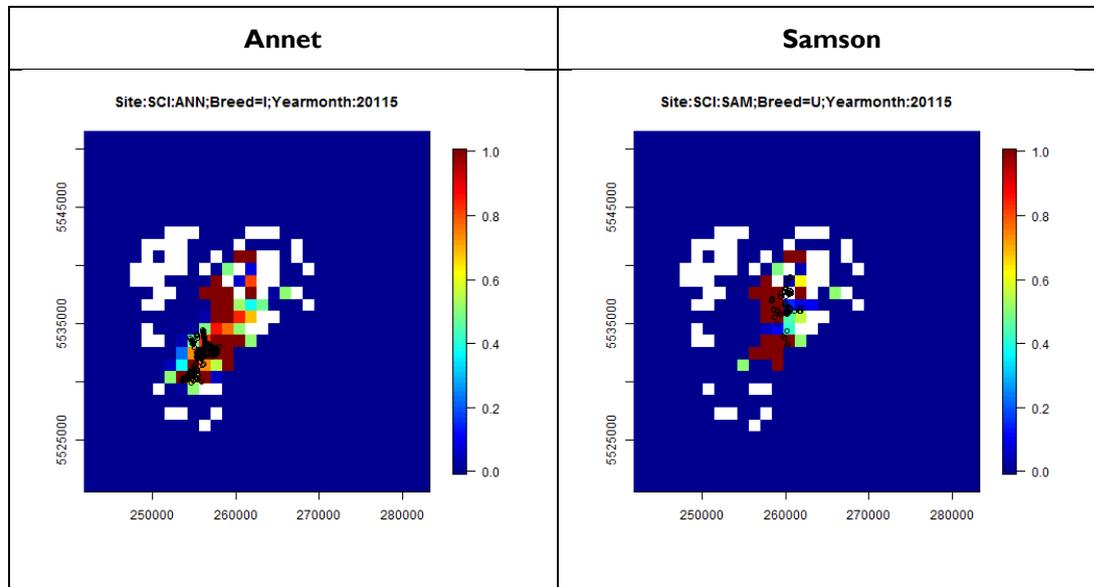
3.4.2 Results excluding the habitat covariate

81Á   Smooth functions for depth and distance from colony were added to the initial model with factor variables. Again, this resulted in large associated p-values and larger CV scores, until both were dropped from the model (in turn). This indicates that the 2D representation of distance from colony is more appropriate than a 1D representation of this relationship, which indicates that the rate of decline in the probability of sighting these birds with increasing distance from the colony differs with the direction from the colony. The fact that a 2D spatial term has been selected indicates spatial patterns are justified in the model and it is evident that since this spatial term also significantly differed with site (via the site x s(x,y) interaction term) that these spatial patterns were materially different across the two sites. However, the very low dimensional smooth spatial term (with just four parameters) improved the CV score and lifted the pseudo R^2 value. This smoother based term was also statistically significant at the 5% level.

Table 5 **Model selection for the candidate models excluding EUNIS (level 3) habitat categories**

Model	Covariates	p-value	CV Score
1	YearMonth	$p < 0.0001$	0.1028579
	breed	$p < 0.0001$	
	site	$p < 0.0001$	
2	YearMonth	$p < 0.0001$	0.01186756
	breed	$p < 0.0001$	
	site	$p = 0.3187$	
	s(x,y, df=7)	$p < 0.0001$	
	site x s(x,y, df=7)	$p < 0.0001$	

82Á   In an identical way to the model results including the EUNIS (level 3) habitat categories, the predicted probabilities of occurrence across the prediction grid for each site, breeding status and year-month combination were plotted (Figure 24). Again, for all plots, the observed presences aligned well with the fitted values and there were no concerning patterns apparent.

Figure 24 Predicted relative probability of occurrence of breeding shags from tracking from the Annet and Samson colonies. Rows one and two are from the Annet colony and row three is from the Samson colony. Tracking from 2010 is in the left-hand column, tracking from 2011 is in the middle column and tracking from 2012 is in the right-hand column

Site:SCI:ANN;Breed=B;Yearmonth:20116

Samson

Site:SCI:SAM;Breed=B;Yearmonth:20105

Site:SCI:SAM;Breed=B;Yearmonth:20115

Site:SCI:SAM;Breed=B;Yearmonth:20125
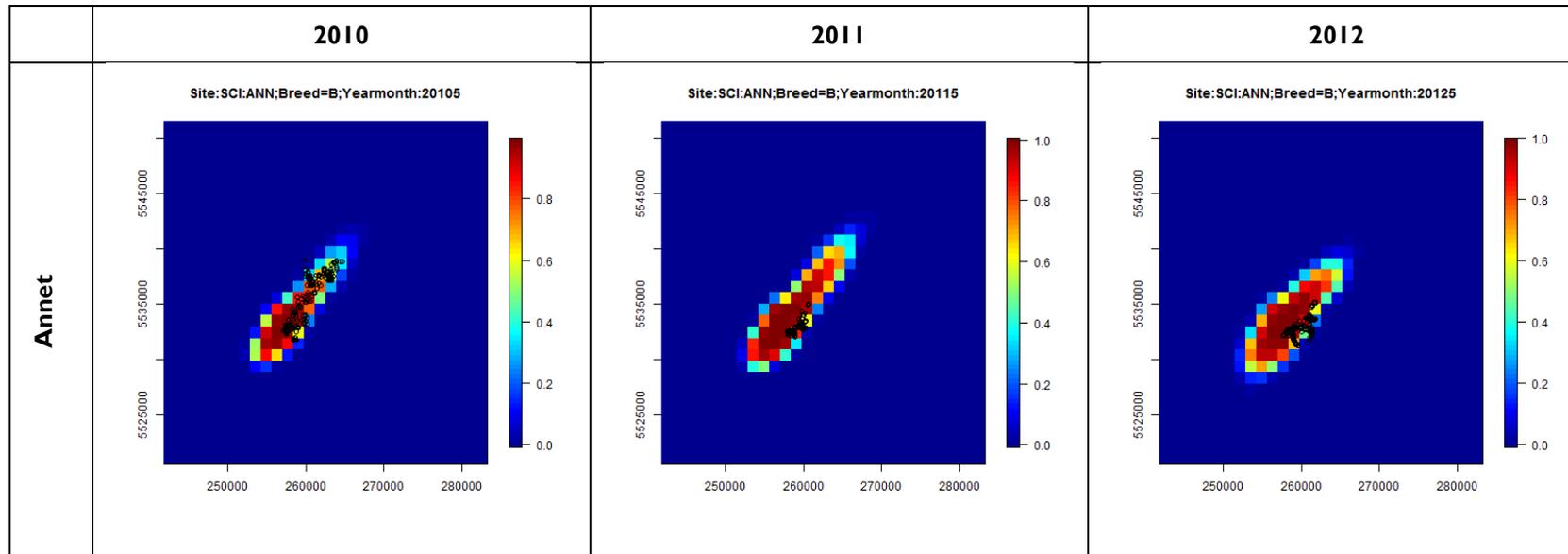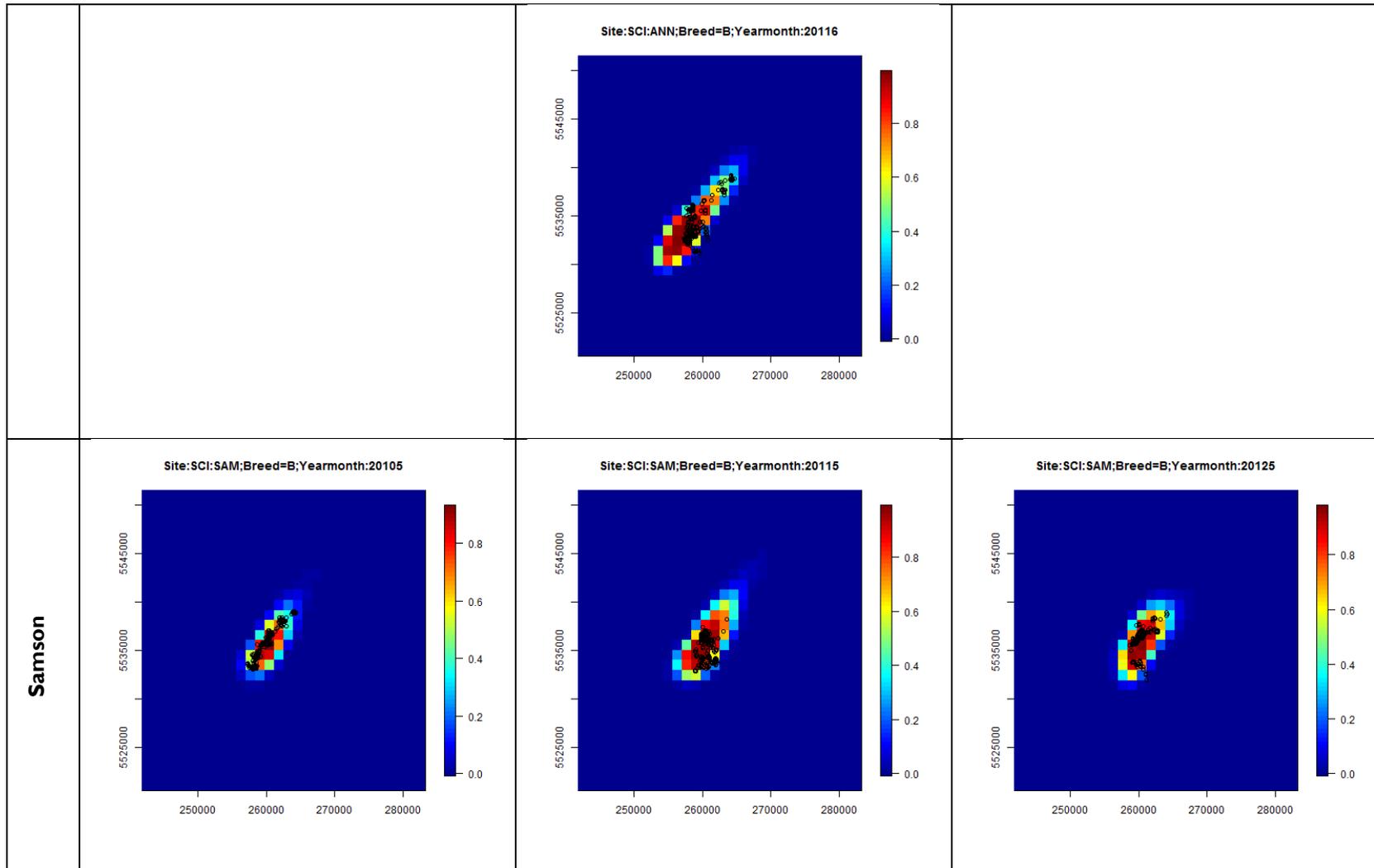
Figure 25 **Predicted relative probability of occurrence immature shags or shags of unknown breeding status. One bird was tracked from the Annett colony (left) and one bird was tracked from the Samson colony (right). Both birds were tracked in 2011**
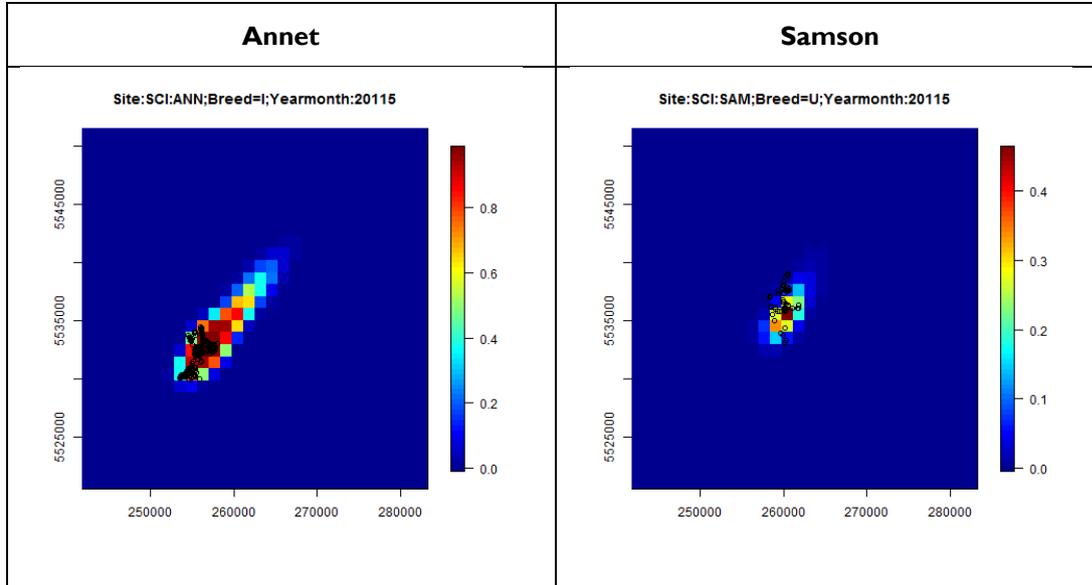
83Á   Finally, the predicted relative probability of occurrence was mapped relative to the islands in the archipelago using the model that included the EUNIS level 3 habitat covariate (Figure 26). Unlike the KDE or DSM predictions, the predictions from the tracking data suggested that birds were only concentrated in the south-west of the archipelago, with no prediction of another hotspot of birds to the north-west around the Eastern Isles. This is likely a reflection of the available data, with tracking mainly from Annet and Samson, both in the south-west of the archipelago. Only a single bird from the colonies in the north-east of the archipelago was tracked, in 2012.

Figure 26 **Predicted relative probability of occurrence of shags across the prediction grid in relation to the islands in the archipelago (top figure), compared with the KDE estimates from DAS data from 2014 and 2015 (both left and right respectively) (Figure 3 and Figure 4)**

3.5 Non-targeted visual boat-based and aerial survey data

84Á   Data for sitting shags during the breeding season were mapped from SeaMaST ESAS data (Figure 27). These showed a fairly simple pattern of coastal abundance evenly around and between the islands. The co-variates used for this DSM were x, y and distance to coast using a 3km x 3km grid cell. This results in relatively low resolution information when viewed at relatively small spatial scales (such as Figure 27). Unlike the analysis of the DAS data using either KDEs or DSM there was no suggestion of hotspots in the south-west and north-east of the islands.

Figure 27   Densities of sitting shags from SeaMaST ESAS data (Bradbury et al. 2014) around the Isles of Scilly
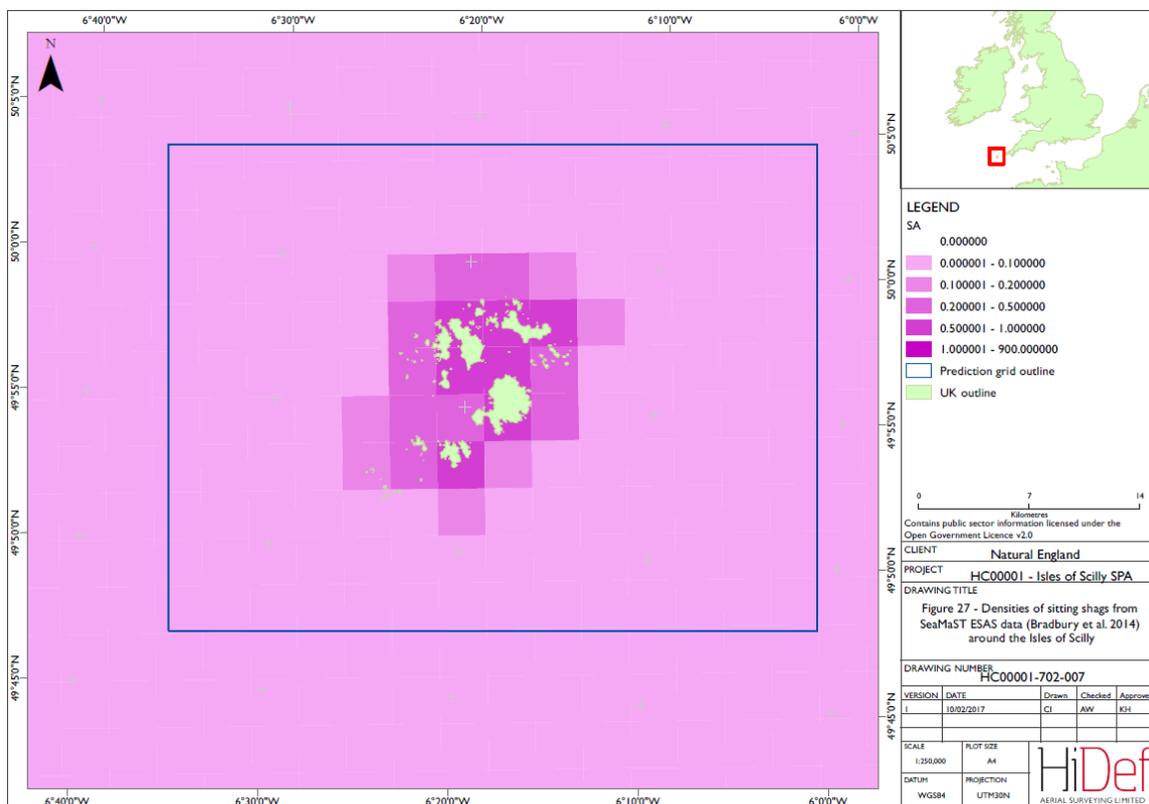
85Á   Data for sitting shags during the breeding season were also mapped from SeaMaST WWT visual aerial survey data (Figure 28). The pattern was broadly similar to the ESAS data, but with birds predicted to be more widely distributed further from shore all around the archipelago. Since these data were of low resolution and also based on a simple interpolation technique it was considered that the predicted distribution was not likely to be representative of the finer scale pattern of use needed for boundary determinations. This was never the primary purpose of these data. These data have been used in the past for very broad scale predictions of areas of high sensitivity to offshore wind farm development (Bradbury et al. 2014), so it is perhaps unsurprising that at the necessary scale here, the data do not perform particularly well. They do however provide useful context, and finer resolution data should be expected to occur within these predictions.

Figure 28   Densities of sitting shags from SeaMaST WWT data (Bradbury *et al.* 2014) around the Isles of Scilly
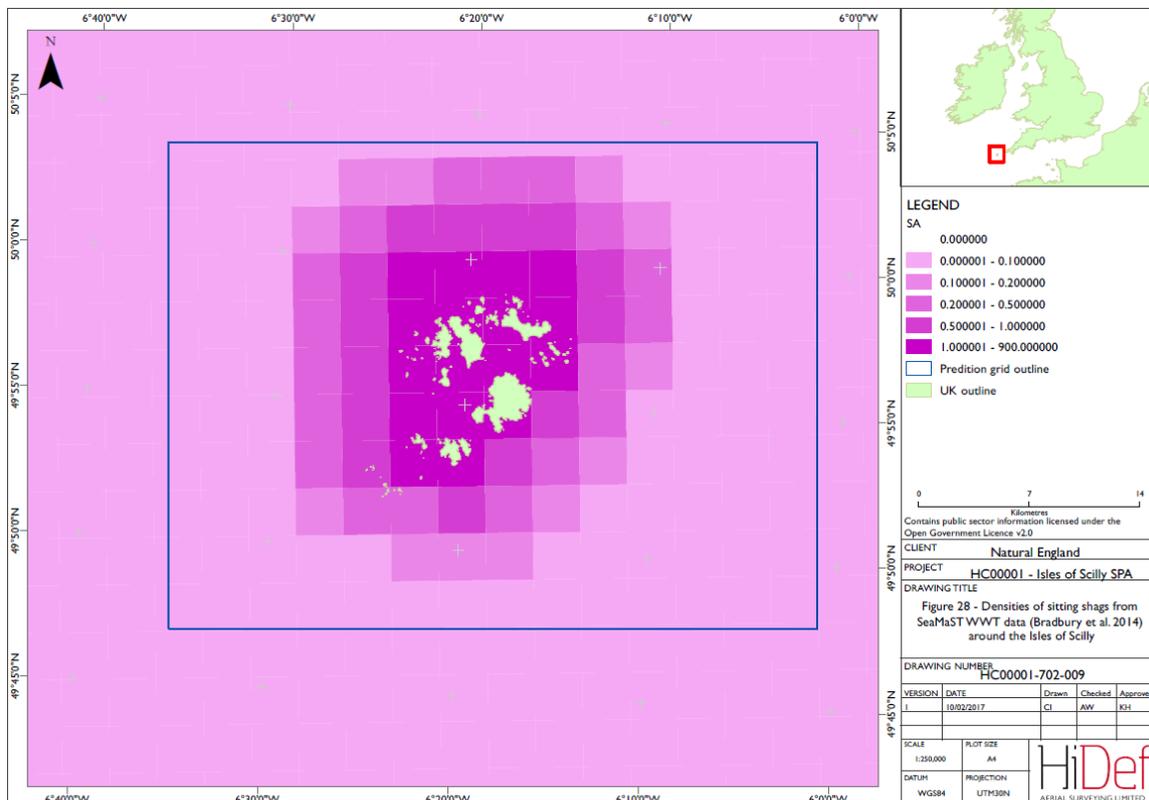
3.6 Boundary determination

86Á   Boundary determination was attempted for the results of the KDE analysis (for 2014 and 2015 survey data separately, for DSMs of the same data (with lower and upper 95% CIs), for models of predicted usage for data from tracked breeding birds only and for all tagged birds (with lower and upper 95% CIs). Boundary determination was not attempted for models predicting usage of tracked birds in which EUNIS codes were selected as a covariate, because the predictions were found to be unreliable (see Section 3.4.1). Similarly, boundary determination was not done for SeaMaST data, because the prediction grids were found to be too coarse to be able to provide accurate boundaries.

3.6.1 Maximum curvature from KDE outputs

87Á   The relationship between the cumulative population of shags and the cumulative area occupied in the KDE outputs are presented in Figure 29 with plots of maximum curvature. The relationship showed a relatively long slope typical of a smoothing method based on a nearest neighbour analysis. The two different methods for fitting the relationship between the cumulative number and cumulative area gave different results, with a much more pronounced curvature for the single exponential model (Figure 29). However, when the residuals for both models were plotted against the proportionate area (Figure 30), the double exponential formula gave the better fit with a much lower value for the sum of squares (Table 6). The preferred sum of squares value was considered sufficient to not require further modelling

options to be attempted. The results and metrics of the maximum curvature analysis are presented in Table 6.

Figure 29   Cumulative number and area for sitting shags from KDE density predictions in 2014 and 2015 with maximum curvature
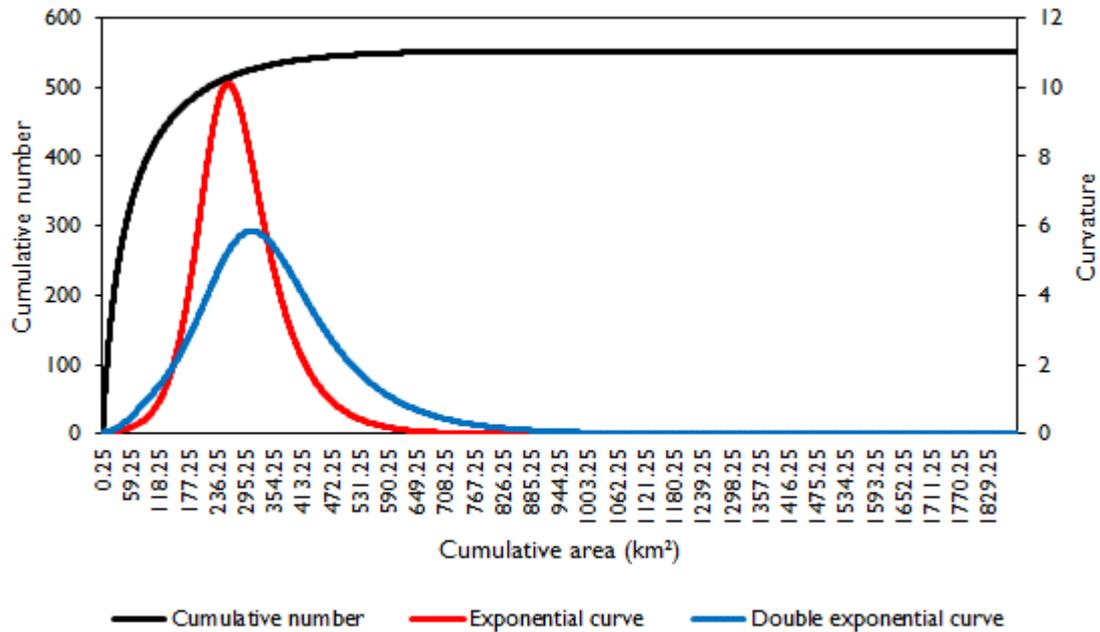
Figure 30   Residuals (fitted minus cumulative proportionate number) for the exponential and double exponential curvature models for KDE outputs compared to cumulative proportionate area
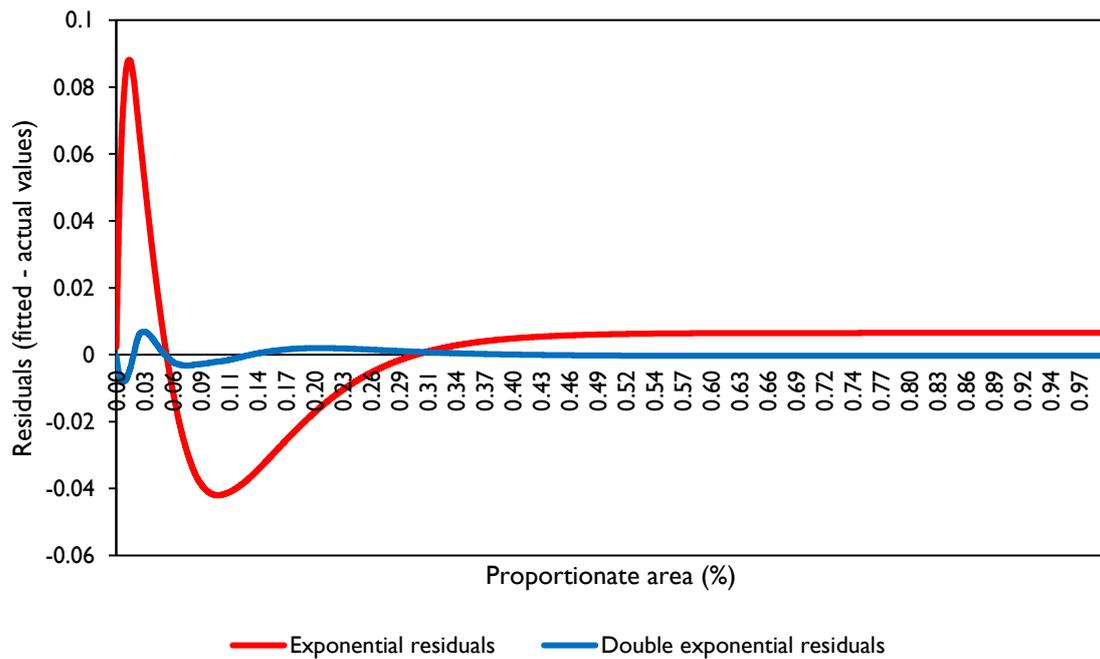
Table 6   **Outputs from maximum curvature analysis of KDE density predictions for sitting shags in 2014 and 2015**

Data source	Model type	Maximum curvature	Sum of squares of residuals	Cumulative number threshold of preferred model	Cumulative area threshold of preferred model	Density threshold of preferred model
KDE predicted density	Single exponential	15.77	2.6305	n/a	n/a	n/a
KDE predicted density	Double exponential	5.85	0.0169	525.5487	308	0.1968

3.6.2 Maximum curvature from DSM outputs

88Á   The relationship between the cumulative population of sitting shags using the point density estimate and the cumulative area occupied in the DSM outputs, and the single and double exponential maximum curvature are presented in Figure 31. These plots showed a relatively steep relationship between cumulative number and cumulative area and a very sharp maximum curvature relationship, as might be expected from smoothed interpolation based on several spatial and habitat covariates. When the residuals for both models were plotted against the proportionate area (Figure 32), the double exponential formula gave slightly the better fit with a much lower value for the sum of squares (Table 7). The sum of squares value was sufficient to not require other more complex curve modelling options to be required. The results and metrics of the maximum curvature analysis are presented in Table 7.

Figure 31   Cumulative number and maximum curvature in relation to cumulative area for sitting shags from DSM outputs using the point density estimate

Figure 32   Residuals (fitted minus cumulative proportionate number) for the exponential and double exponential curvature models for DSM outputs compared to cumulative proportionate area
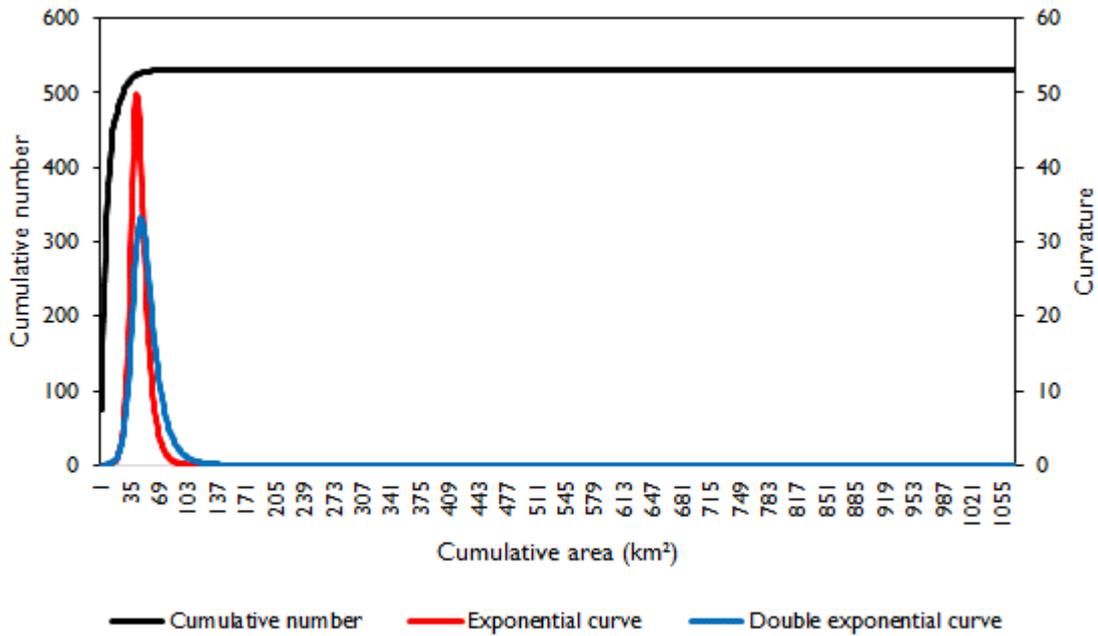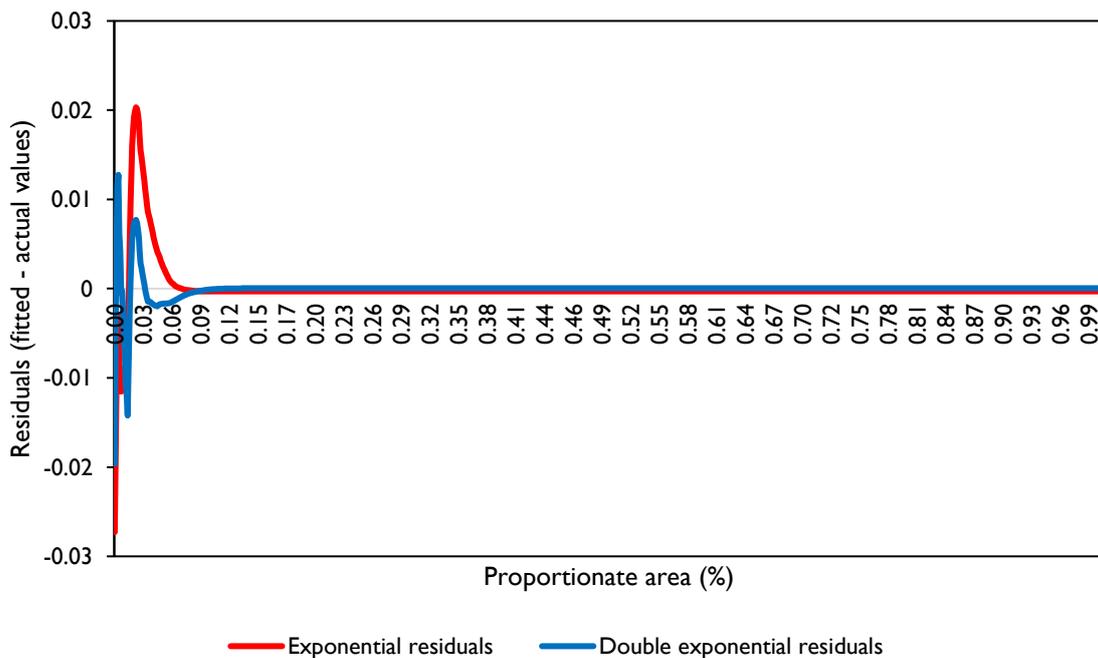
Table 7   Outputs from maximum curvature analysis of DSM density predictions for sitting shags in 2014 and 2015

Data source	Model type	Maximum curvature	Sum of squares of residuals	Cumulative number threshold of preferred model	Cumulative area threshold of preferred model	Density threshold of preferred model
DSM predicted density	Single exponential	10.10	0.0007	n/a	n/a	n/a
DSM predicted density	Double exponential	49.80	0.0004	526	48	0.33

3.6.3 Maximum curvature from tracking data

89Á   The relationship between the cumulative probability of all shags (i.e. including the data from one non-breeding shag and one of unknown breeding status) from the tracking model using the point probability of occurrence estimate and the cumulative area occupied in the DSM outputs, and the single and double exponential maximum curvature are presented in Figure 33. These plots showed a relatively steep relationship between cumulative probability and cumulative area and a very sharp maximum curvature relationship, as might be expected from smoothed interpolation based on several spatial and habitat covariates. When the residuals for both models were plotted against the proportionate area (Figure 33), the double exponential formula gave considerably better fit with a much lower value for the sum of squares (Table 8). The sum of squares value for the single exponential was sufficient to not require other more complex curve modelling options to be required. The results and metrics of the maximum curvature analysis are presented in Table 8.

Figure 33   Cumulative probability and maximum curvature in relation to cumulative area for all shags from tracking model outputs using the point estimate of the probability

Figure 34   Residuals (fitted minus cumulative proportionate number) for the exponential and double exponential curvature models for tracking outputs compared to cumulative proportionate area
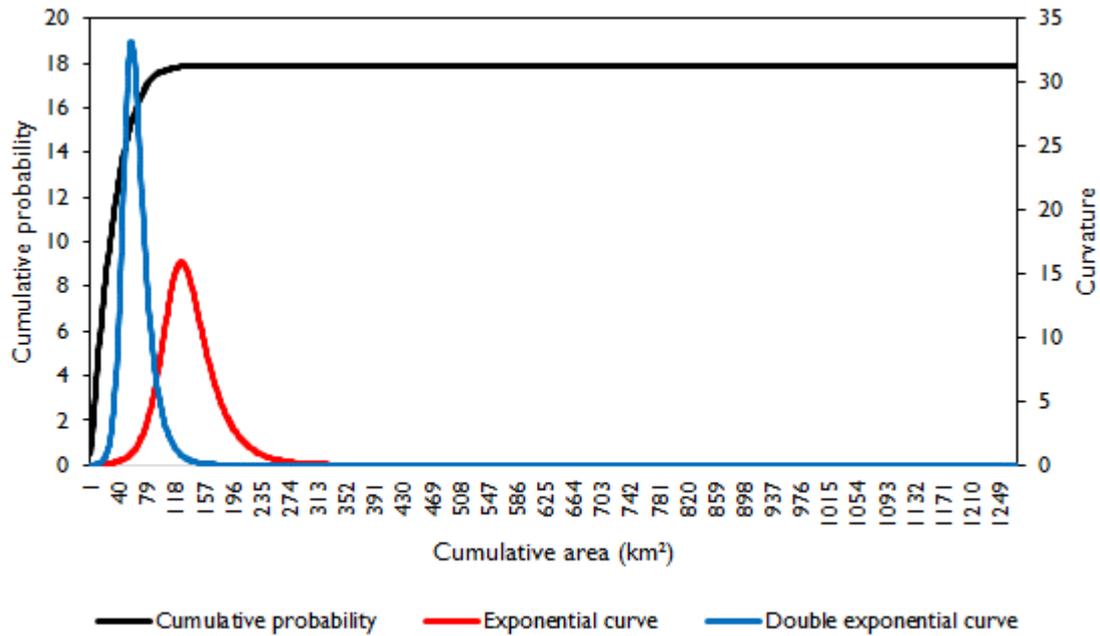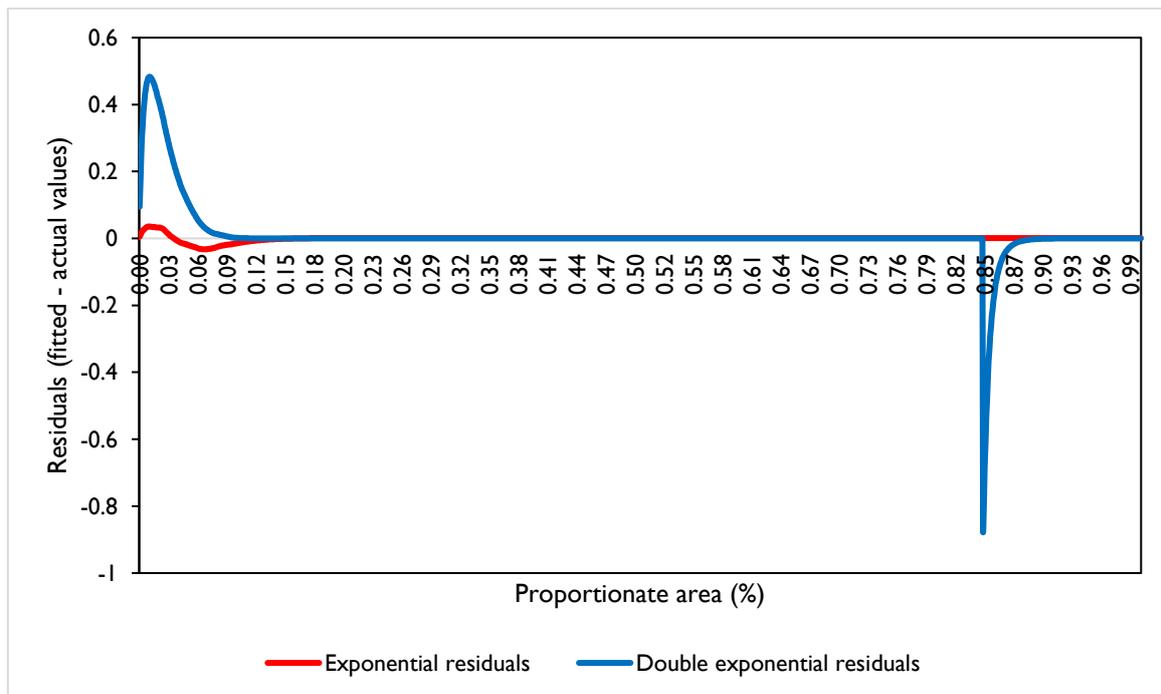
Table 8 **Outputs from maximum curvature analysis of tracking model predictions of all shag probability of presence.**

Data source	Model type	Maximum curvature	Sum of squares of residuals	Cumulative number threshold of preferred model	Cumulative area threshold of preferred model	Density threshold of preferred model
Tracking predicted probability	Single exponential	15.86	0.0803	17.83	125	0.00292
Tracking predicted probability	Double exponential	33.19	10.4895	n/a	n/a	n/a

3.6.4 Boundary options for a shag marine SPA proposal

90Á   The density threshold selected by the double exponential maximum curvature model, when applied to the KDE average density grid describes a broad area around the Isles of Scilly, which extend up to 6.5km south of the Western Islands, Isles of Scilly and 5.5 km north-east (Figure 35). There is little evidence from the raw observations from the digital aerial survey that this extent for a boundary option is realistic.

Figure 35 **Extent of grid cells with density exceeding the threshold determined by double exponential maximum curvature for sitting shag KDE outputs from survey data in 2014 and 2015**
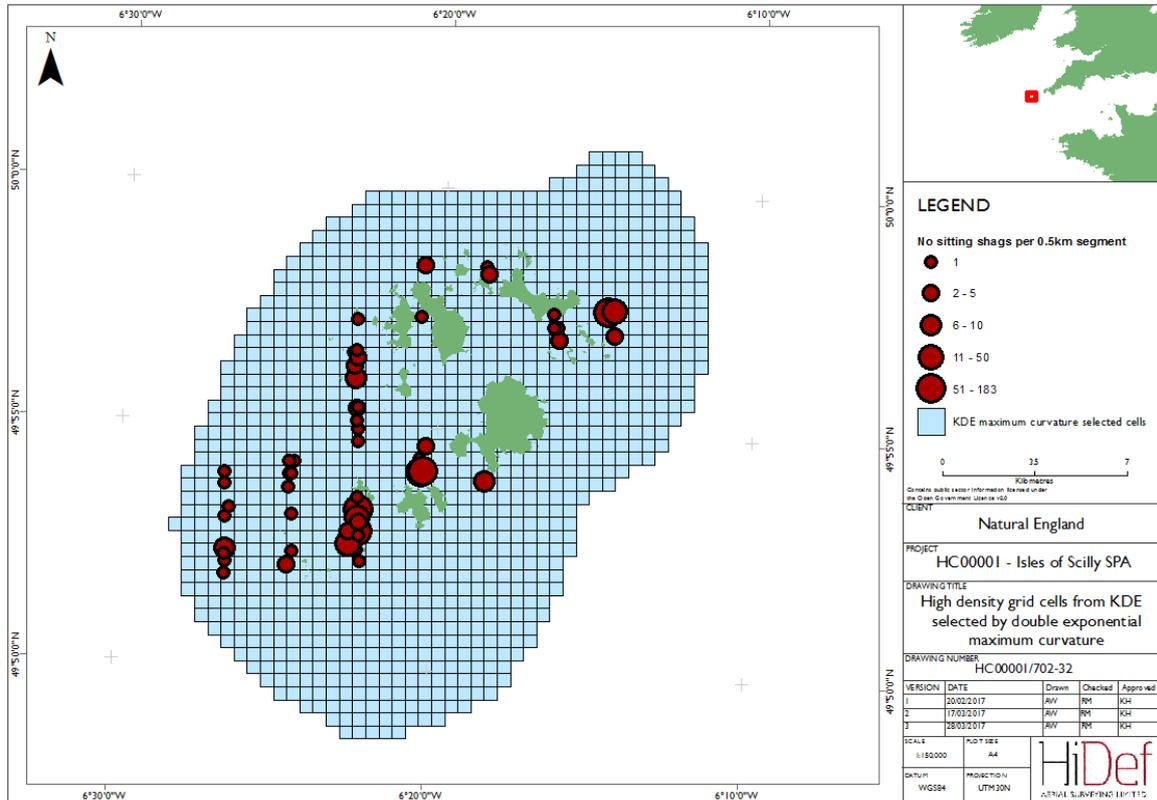
91Á The density thresholds derived from maximum curvature analysis, when applied to the DSM grids for the sitting shags from digital aerial survey data describe very tight areas around the raw observations of sitting shags (Figure 36). Greatest representation of selected cells is around the largest aggregations of raw observations, especially in the south, but some of the smaller aggregations in the raw observations to the west of the islands are not represented by the selected cells. Although there were few raw observations of shags, there were no selected cells in the region between the islands of Tresco, St Martins and St Marys (Figure 36).

Figure 36   **Extent of grid cells with density exceeding the threshold determined by double exponential maximum curvature for sitting shag DSM average density estimates from digital aerial survey data with raw observations**
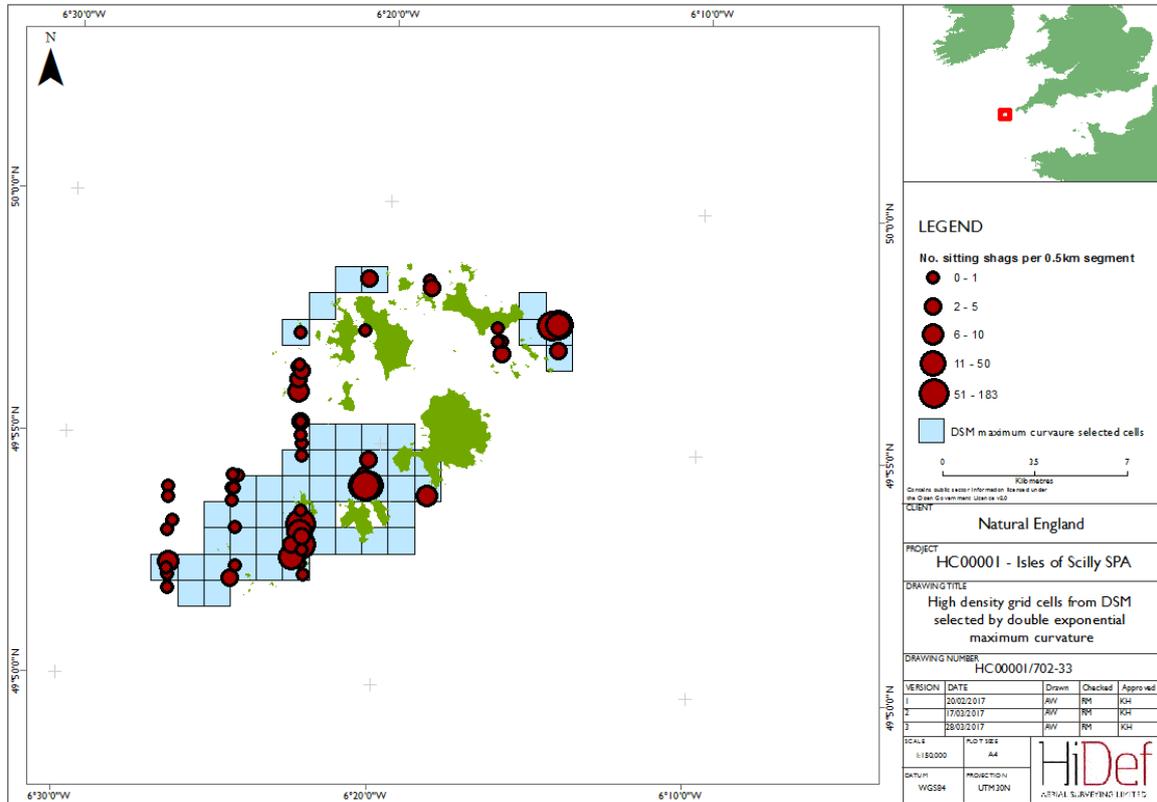
92Á   The probability thresholds derived from maximum curvature analysis, when applied to the predictive grids for all shags from tracking survey data describe fairly tight areas around the Isles of Scilly (Figure 37). The predicted usage is slightly more extensive than the raw observations of tracked shags, particularly to the north-east and south of both the archipelago and the raw observations, with no cells selected to the north-west of Samson and Tresco. The predictions, like the raw tracking data, identified the area between the islands of Tresco, St Martins and St Marys for inclusion within the SPA proposal, unlike the outputs from the DSMs of digital aerial survey data (Figure 36 and Figure 37).

Figure 37   Extent of grid cells with probability exceeding the threshold determined by single exponential maximum curvature for all tracked shag model output average probability estimates with raw tracking data
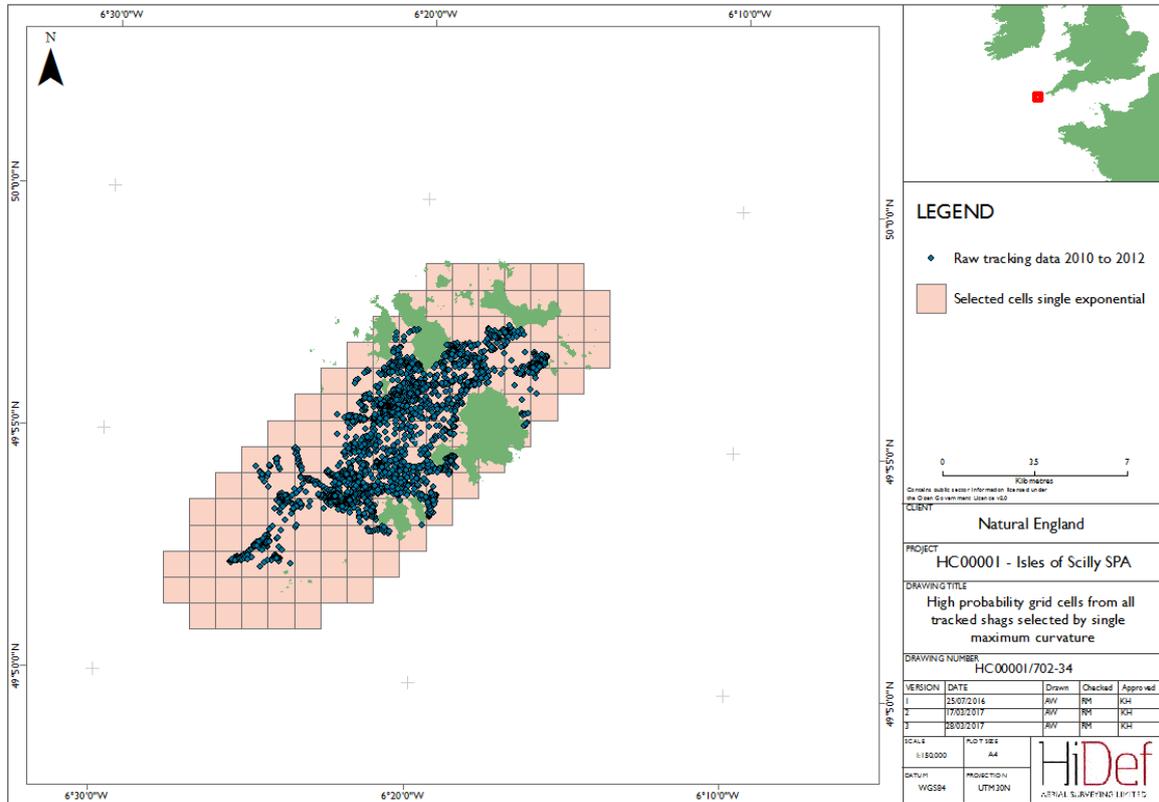
4 Discussion

93Á   The analysis of existing data sources (SeaMaST and GPS logger data) and data collected for determining the boundaries of a SPA around the Isles of Scilly for shags (DAS data) found broadly similar patterns, which largely met the expectation of shags being highly coastal birds.

94Á   All existing SPAs for shags are coastal breeding colonies, though recent SPAs in the marine environment have been proposed and are in the consultation phase. Four sites have been proposed in Scotland. Scapa Flow pSPA and North Orkney pSPA are proposed for shags as a winter feature, and the Moray Firth pSPA and Outer Firth of Forth and St Andrews Bay Complex pSPA are proposed for shags as both a breeding and winter feature. These sites have been proposed for shags occurring in coastal waters, even though some sites include more pelagic waters (for other seabird features).

95Á   The raw data collected using digital aerial surveys in 2014 and 2015 also showed a highly coastal distribution of shags sitting on the water. "Sitting" birds included birds diving and birds taking off or landing on the water. This selects birds interacting with the water surface and is a better indication of habitat use than birds in flight. The raw data indicated a higher occurrence in the south-west and north east of the archipelago and this was reflected in the analysis of these data using KDEs. The surveys in 2014 found more birds sitting on the water than the surveys in 2015, despite identical levels of effort between years. It is uncertain why this was the case. This difference resulted in the spatial abundance of shags being predicted by the KDEs as different between each year. However, the DSM model could predict a single density surface while controlling for temporal effects.

96Á   The DSM results predicted higher abundances in the waters around St Agnes and the Eastern Isles, between St Martin's and St Mary's. The breeding colonies on the Isles of Scilly are spread across many of the islands, but the bulk of the population occurs in the south-west of the archipelago and around the Eastern Isles (Figure 38). Thus, the prediction met the expectation that shags are highly coastal and have relatively short foraging ranges (Thaxter *et al.* 2012).

Figure 38 **Breeding colonies of shags in the Isles of Scilly. The size of the circles represents the size of the colony in 2015 or 2016.**
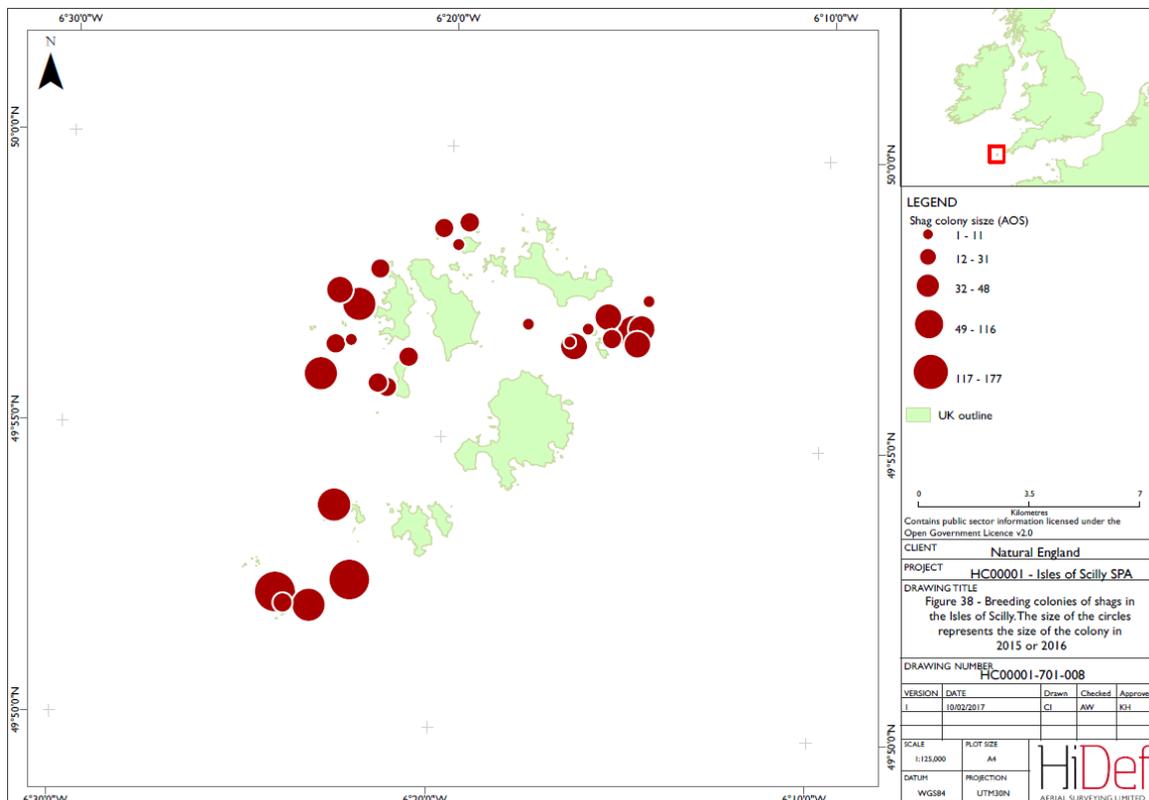
97Á   The analysis of shag tracking data needs to be interpreted with some care. The model provides outputs on the *relative* probability of occurrence. Two models were run, one with habitat covariates and the other without habitat covariates. Each produced quite different results, though both were well supported statistically. The model including the habitat information predicted occurrence across a much wider area than the model that excluded these data. We think that this is due to the model not constraining the distribution of shags by distance from the colony or by water depth (as these were not significant factors in the model). However, since we know that both distance from the colony and water depth are constraints on shag distribution (e.g. Daunt *et al.*, 2015), it was decided that this model may not provide a reasonable explanation of shag distribution. The strong relationship with habitat appeared to be predicting in occurrence only because of that habitat variable. On removing this variable from the model, the results were closer to the observed tracking data results (Evans *et al.* 2016), and to those found by the DSM model of DAS data. Thus, it was thought that these were a more likely to be a reasonable descriptor of shag distribution. One outcome of not being able to use distance from colony as a covariate in the models, perhaps due to the small sample size, was that it wasn't possible to extrapolate the model predictions to other shag colonies within the Isles of Scilly archipelago.

98Á   It is also important to consider the input data for this model from GPS loggers. Only three colonies were tracked and used in these models (Annet, Samson and Ganinick). Annet and Samson are both in the west and south-west of the archipelago, and Ganinick is in the north-east. Thus, the lack of predicted occurrence in the north-west is not unexpected. It is also important to consider that most birds were only tracked during chick rearing, a relatively short period of the whole breeding season. Thus, the

results may underestimate the spatial use of the area by breeding shags, as data are more limited and it may be that, as in other seabirds (Daunt *et al.* 2015), shag foraging is more extensive during incubation than chick rearing. This was not a constraint on other data sources. However, as Evans *et al.* (2016) showed, there was considerable overlap in the foraging ranges of the three colonies that were sampled. The intention of the approach to modelling shag spatial distribution was to extrapolate from other, un-sampled, colonies to predict use across the archipelago. However, while colony was a significant variable in the model, distance to colony was not. As a result, it was not possible to extrapolate the patterns of spatial use from the sampled colonies to the un-sampled ones. Given the broad similarities in the DSM and tracking predictions, it may be that a strong response of birds to forage in only a few particular hotspots may have been the reason for the low significance of distance from colony in the model.

99Á The other existing data set that was assessed was ESAS and WWT data from SeaMaST (Bradbury *et al.* 2014). The ESAS boat-based survey results show higher densities around the islands, with perhaps a higher concentration of birds in the north and east of the archipelago than in the south and west. These results differ from the predictions of both the KDE and DSM analysis of DAS data, although this is likely to be a consequence of the non-systematic way in which these data were collected (mainly from ferries approaching the main island from the north-east). These data support the findings of the digital aerial surveys and the tracking data that the core areas used by shags occur within the bounds of the Isles of Scilly archipelago.

100Á Maximum curvature analysis of the outputs from the different modelling outputs selected different groups of grid cells which could be included within the boundaries of a marine SPA proposal. For all of the density or probability prediction methods, either the single or the double exponential curvature method provided a close match to the relationship between the cumulative number or probability and the cumulative area, evidence by low sum of squares of the residuals. The meant that it was not necessary to resort to other more complex and less appropriate maximum curvature methods. The largest area was generated by the combined KDE analyses, as might be expected from a relatively simple process such as nearest neighbour methods. The selected grid cells included a large number at some considerable distance from the existing islands and most importantly from the raw sightings of the sitting shags during the digital aerial surveys and the foraging shags in the tracking data.

101Á Maximum curvature analysis used density as the input metric from the KDE and DSM analysis, which conforms with the approach described by O'Brien *et al.* (2012). However, the maximum curvature analysis from the tracking data used probability of presence as the input metric. While this doesn't conform exactly to the principles of O'Brien *et al.* (2012), a trade-off between the cumulative probability of presence and the size of area is still a useful principle to define a boundary and was used successfully by Wilson *et al.* (2014) to define proposed SPA boundaries for foraging terns.

102Á Selected grid cells that could be included within a marine SPA proposal using outputs from the DSM were tightly clustered around the island archipelago.

103Á The grid cells selected for inclusion in a marine SPA proposal based on the tracking data were only carried out using model outputs that did not include any habitat covariates, depth, or distance from shore. This selected a slightly more extensive core area to the north-east and south-west of those selected from the DSM analysis of DAS data. The satellite aggregation identified based on the analysis of 95% CIs is not realistic and not supported by any observations in this region from DAS, tracking or SeaMaST data.

104Á   Based on these observations, two outputs from the two modelling approaches appear to give greatest confidence and represent two options for drawing recommended boundaries for a shag marine SPA proposal. A boundary to contain the core selected grid cells based on the double exponential maximum curvature analysis using the DSM outputs, referred to as option 1 and a boundary to contain the core selected grid cells based on the single exponential maximum curvature analysis using the tracking data outputs referred to as option 2. Neither of these two options represents the full extent of the raw observations from the digital aerial survey and the those of the tracking data. The option based on the DSM of aerial survey data did not predict the inclusion of grid cells within the Scillies archipelago between the islands of Tresco, St Martins and St Marys, nor to the north of the Bishop Rock whereas the option based on the tracking data did not predict inclusion of cells to the north-west of the islands, yet extended beyond raw observations to the north-east and south-west of the archipelago. Based on this observation, a boundary option based on a composite of these two options would represent a potential third option.

105Á   The boundary should be drawn initially as a convex hull polygon around the selected cells for each option plus the location of all the shag breeding sites in the existing SPA to include the most likely flight lines between the colonies and the predicted feeding areas. This is because the models of shag distribution based on the DSM were based only on birds that were likely to be feeding or engaged in maintenance activities while sitting on the sea and did not include flying or roosting birds, the distribution of which were unlikely to correlate with habitat variables. Because the breeding location and the flight lines between these and shag feeding sites are still an important component of what might be considered a most suitable territory, it is important to ensure that any additional sea areas and the airspace above them are included in the final boundary.

5 Conclusions

106Á Modelling of the spatial distribution of shags around the Isles of Scilly resulted in broadly similar patterns of shags being highly coastal, even when multiple methods were used. KDE analysis of both DSM and SeaMaST data, DSM of DAS data, and spatial analysis of tracking data were all used. This is unexpected, as shags are known to be highly coastal.

107Á Both the KDE analysis of DAS data and the SeaMaST data predicted shags to be much less coastal and more widely spread than either the DSM analysis of the DAS data or the modelling of the tracking data. This is most likely due to the analysis being a relatively simple interpolation model for the KDE, and because analysis was carried out on a coarse scale for the latter study.

108Á The results of the DSM of the DAS data showed a strong relationship with water depth, which was also unsurprising and is similar to results from other colonies (e.g. Daunt *et al.* 2015 and the references therein). The marginally significant standard deviation of sea surface temperature, was also as expected, as this has been found elsewhere (Virgili 2014). However, the shape of the relationship itself is hard to explain, and was not particularly strong.

109Á Modelling of tracking data also produced results that broadly fit the picture of birds occurring in coastal waters. Unfortunately, since few birds from only three colonies were tracked, the results of this analysis should take this into account. Despite this the predictions for modelling of tracking data were broadly similar to those of the DSM analysis.

110Á The DSM analysis of the DAS data appeared to provide the most robust prediction of the spatial abundance of shags across the Isles of Scilly.

111Á The use of maximum curvature analysis to define the most important high density or high probability grid cells to include within a marine SPA proposal for shags produced areas that varied only slightly in size. The cells selected largely reflected the outcomes of the modelling approaches, with the most robust method, the DSM of DAS data, selecting cells close to the islands in the archipelago.

112Á The two DSM and tracking results were broadly similar if a boundary were drawn around each. Therefore, it may be valuable to consider a single boundary based on the larger extent of both analyses. This can only be achieved by combining the two boundaries, rather than through a new boundary analysis of combined data, as they two data sources are not the same (one providing density information and the other probability information).

6 Recommended boundary options

113Á Based on the outcome of the modelling, the maximum curvature analysis and initial boundary analysis, two boundaries were produced that could inform the marine SPA proposal for shags:

- Á Option 1 uses the double exponential maximum curvature analysis of the DSM analysis of the DAS data (Figure 39). The mean of the different predicted values of density per 1km grid cell were selected using a double exponential maximum curvature analysis. A convex hull polygon was drawn around the centroid of these cells and the location of each shag breeding colony within the Isles of Scilly archipelago with a 707 metre buffer (the distance from the centroid to

the corner of each 1km x 1km grid square). The inclusion of the colony location and the buffer ensures the inclusion of flight lines between the colonies and the predicted feeding areas.

Figure 39   Location of boundary for a marine SPA proposal for shags based on selected DSM grid cells from DAS data and colony location (option 1)
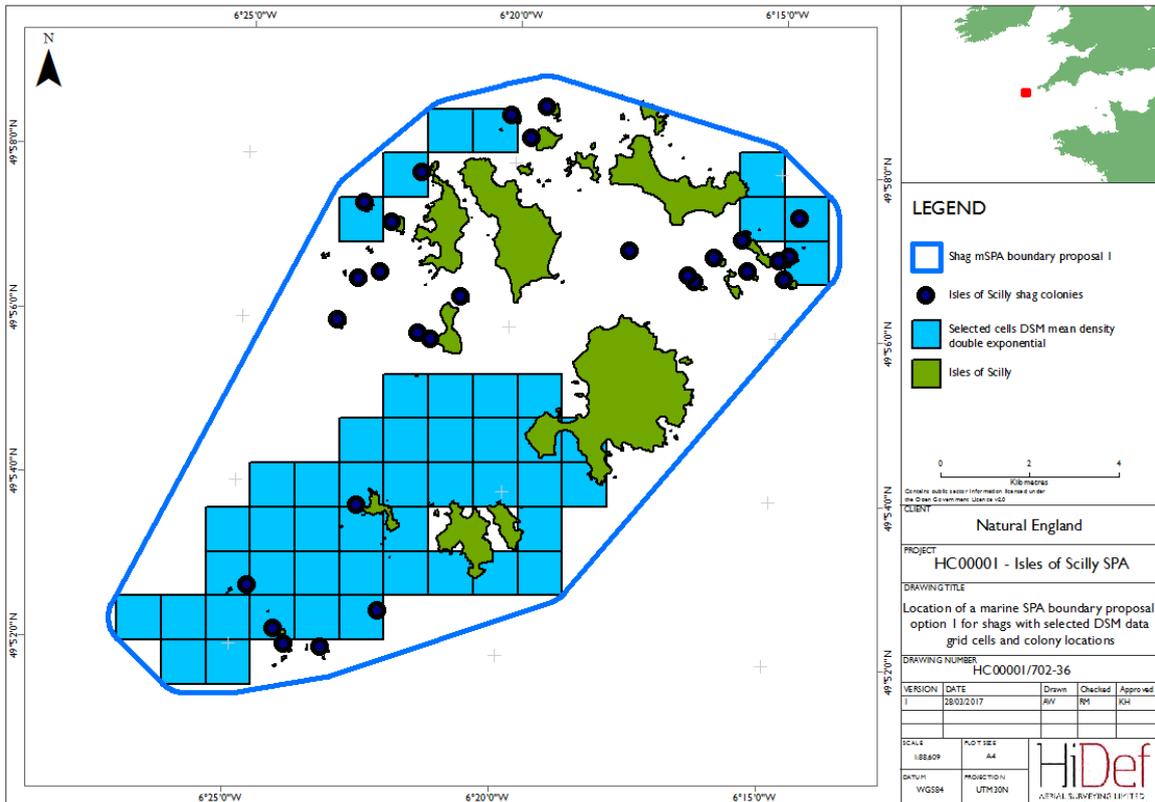
- Á Option 2 uses the single exponential maximum curvature analysis of the analysis of the GPS tracking data (Figure 40). The mean of the different predicted probability of presence per 1km x 1km grid cell were selected using a single exponential maximum curvature analysis. A convex hull polygon was drawn around the centroid of the selected cells and the location of each shag breeding colony within the Isles of Scilly archipelago with a 707 metre buffer. As with the option 1 boundary proposal, the inclusion of the colony data and buffer ensures the inclusion of flight lines between the colonies and the predicted feeding areas.

Figure 40   Location of boundary for a marine SPA proposal for shags based on selected grid cells from GPS tracking data and colony location (option 2)
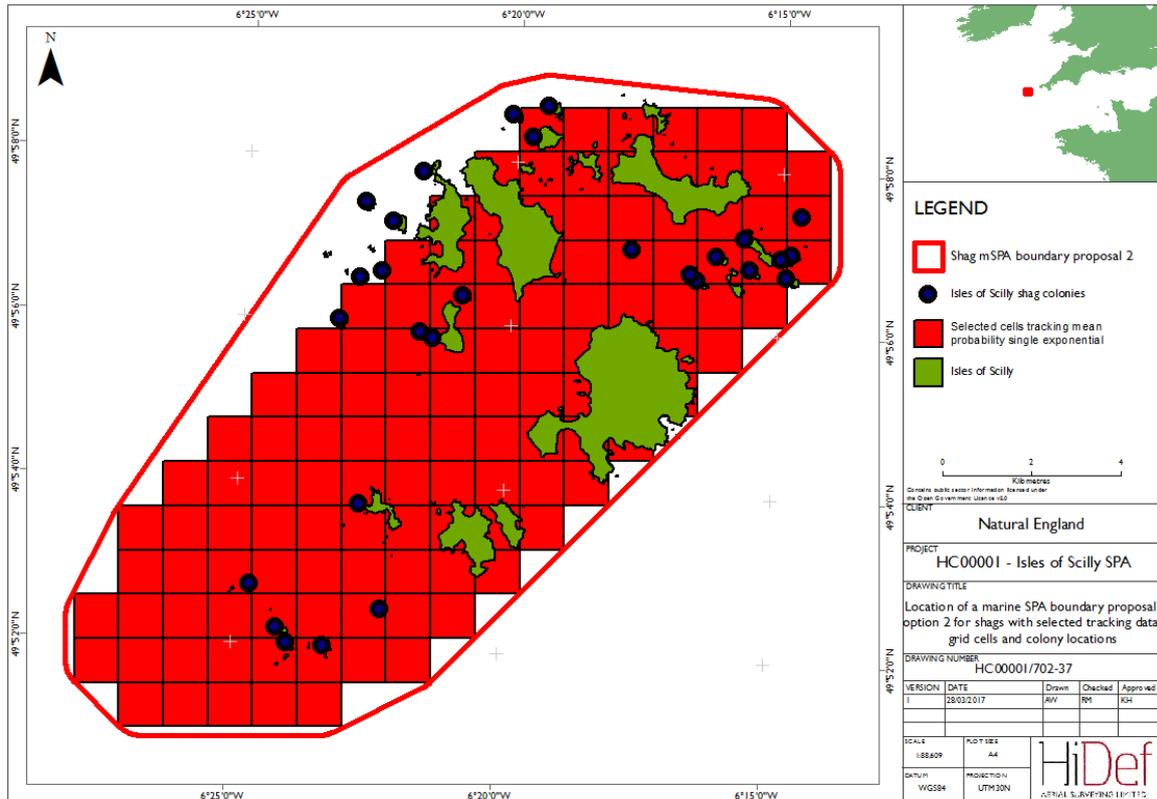

114   We recommend that the option 1 boundary represents the most suitable territory as a SPA for shags on the basis that this model that underlies the boundary appears to be the most robust of the two options and that there is no evidence that the shags feeding around their colonies in the Isles of Scilly were venturing significantly beyond this boundary.   Option 2 provides a slightly more extensive boundary than option 1 and includes some areas to the north-east and south-west where there is no evidence of these waters being used by feeding or flying shags but included some areas not well represented by the option 1 boundary, notably to the north of the Bishop Rock where there is evidence of low level usage from the DAS raw observations.   We do not recommend the option 2 boundary by itself, but if overlaid with the boundary for option 1 it could represent a third option which might represent a more precautionary option than that provided by option 1 and is provided in Figure 41.

115   Whichever boundary is ultimately used, a further step will be necessary to draw a boundary based upon straight lines of latitude and longitude as described by Johnston *et al.* (2002).

Figure 41   Location of a marine SPA boundary proposal for shags using both option 1 and option 2 boundaries with colony location
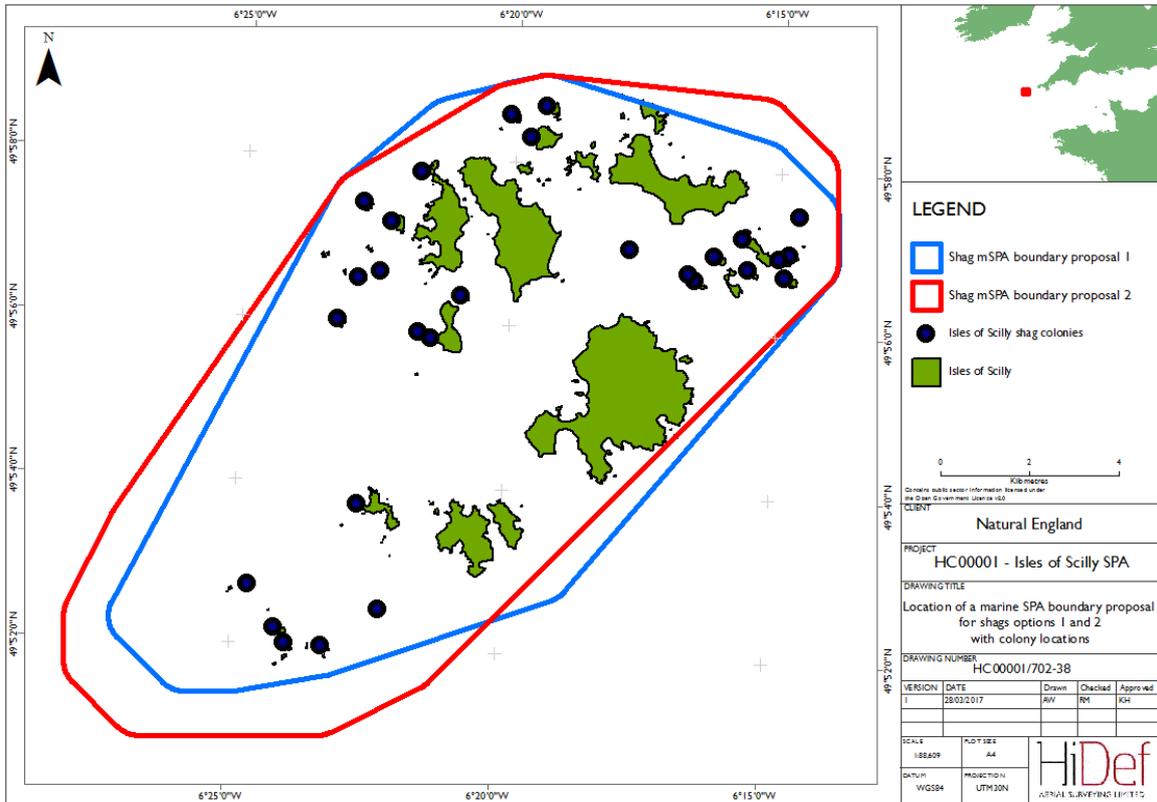
7 References

Barbet-Massin, M., Jiguet, F., Albert, C.H. and Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), pp.327-338.

BirdLife International. 2001. Important Bird Areas and potential Ramsar Sites in Europe. BirdLife International, Wageningen, The Netherlands.

BirdLife International. 2010. Marine Important Bird Areas toolkit: standardised techniques for identifying priority sites for the conservation of seabirds at sea. BirdLife International, Cambridge UK.

Bradbury G., Trinder, M., Furness, B., Banks, A.N., Caldow, R.W.G., Hume, D. 2014. Mapping Seabird Sensitivity to Offshore Wind Farms. PLoS ONE 9(9): e106366. doi: 10.1371/journal.pone.0106366. pmid:25210739.

Daunt, F., Bogdanova, M., McDonald, C. & Wanless, S. 2015. Determining important marine areas used by European shag breeding on the Isle of May that might merit consideration as additional SPAs (2012). JNCC Report No 556. JNCC, Peterborough

EUMETSAT/OSI SAF. 2015. GHRSST Level 3C North Atlantic Regional (NAR) subskin Sea Surface Temperature from SNPP/VIIRS (GDS V2) produced by OSI SAF. Ver. 1. PO.DAAC, CA, USA. Dataset accessed 2016-12-22.

Evans, J.C., Dall, S.R., Bolton, M., Owen, E. & Votier, S.C., 2016. Social foraging European shags: GPS tracking reveals birds from neighbouring colonies have shared foraging grounds. *Journal of Ornithology*, 157(1), 23-32.

Garthe, S., Markones, N., Mendel, B., Sonntag, N. and Krause, J.C., 2012. Protected areas for seabirds in German offshore waters: designation, retrospective consideration and current perspectives. *Biological Conservation*, 156, pp.126-135.

Heaney, V., Lock, L., St Pierre, P. & Brown, A. 2008. Breeding seabirds on the Isles of Scilly. *British Birds*, 101, 418-438.

IOC, IHO, and BODC, 2003, "Centenary Edition of the GEBCO Digital Atlas", published on CDROM on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans; British Oceanographic Data Centre, Liverpool.

Irwin, C. 2015. Digital video aerial survey at sea of European shag *Phalacrocorax aristotelis* around the Isles of Scilly in 2014. HiDef report HP00049-701 to Natural England.

Johnston, C.M., Turnbull, C.G. & Tasker, M.L. 2002. Natura 2000 in UK Offshore Waters: Advice to support the implementation of the EC Habitats and Birds Directives in UK offshore waters. JNCC Report, No. 325. JNCC, Peterborough.

Kober, K., Webb, A., Win, I., Lewis, M., O'Brien, S., Wilson, L.J., Reid, J.B., 2010. An analysis of the numbers and distribution of seabirds within the British Fishery Limit aimed at identifying areas that qualify as possible marine SPAs. JNCC Report, No. 431, Peterborough, UK.

Lawson, J., Kober, K., Win, I., Allcock, Z., Black, J. Reid, J.B., Way, L. & O'Brien, S.H. 2016. An assessment of the numbers and distribution of wintering red-throated diver, little gull and common scoter in the Greater Wash. JNCC Report No 574. JNCC, Peterborough.

Lawson, J., Kober, K., Win, I., Bingham, C., Buxton, N.E., Mudge, G., Webb, A., Reid, J.B., Black, J., Way, L. & O'Brien, S. 2015. An assessment of numbers of wintering divers, seaduck and grebes in inshore marine areas of Scotland. JNCC Report No 567. JNCC, Peterborough.

Mackenzie, M.L, Scott-Hayward, L.A.S., Oedekoven, C.S., Skov, H., Humphreys, E., and Rexstad, E. 2013. Statistical Modelling of Seabird and Cetacean data: Guidance Document. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

O'Brien, S.H., Webb, A., Brewer, M.J. and Reid, J.B., 2012. Use of kernel density estimation and maximum curvature to set Marine Protected Area boundaries: Identifying a Special Protection Area for wintering red-throated divers in the UK. Biological conservation, 156, pp.15-21.

Simonoff J. S. 1996. *Smoothing Methods in Statistics*. Springer, London.

Soanes, L.M., Bright, J.A., Angel, L.P., Arnould, J.P.Y., Bolton, M., Berlincourt, M., Lascelles, B., Owen, E., Simon-Bouhet, B. and Green, J.A., 2016. Defining marine important bird areas: Testing the foraging radius approach. *Biological Conservation*, 196, pp.69-79.

Stroud, D.A., Chambers, D., Cook, S., Buxton, N., Fraser, B., Clement, P., Lewis, I., McLean, I., Baker, H., Whitehead, S., 2001. The UK SPA Network: its Scope and Content, vols. 1–3. JNCC, Peterborough, UK.

Thaxter, C.B., Lascelles, B., Sugar, K., Cook, A.S., Roos, S., Bolton, M., Langston, R.H. & Burton, N.H., 2012. Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. Biological Conservation, 156: 53-61.

Virgili, A., 2014. Predicting variations in coastal seabird habitats to assess marine protected areas. Ingénieur des sciences agronomiques, agroalimentaires, horticoles et du paysage. Agrocampus Ouest, Rennes, 44p.

Webb, A. & Irwin, C. 2016. Digital video aerial survey at sea of European shag *Phalacrocorax aristotelis* around the Isles of Scilly in 2015. HiDef report HP00058 701 002 to Natural England.

Webb, A., McSorley, C.A., Dean, B.J., Reid, J.B., 2004. Modelling the distribution and abundance of black scoter *Melanitta nigra* in Carmarthen Bay in winter 2001/ 02: a method for identifying potential boundaries for a marine Special Protection Area. JNCC Report No. 330. JNCC, Peterborough.

Wilson L. J., Black J., Brewer, M. J., Potts, J. M., Kuepfer, A., Win I., Kober K., Bingham C., Mavor R. & Webb A. 2014. Quantifying usage of the marine environment by terns Sterna sp. around their breeding colony SPAs. JNCC Report No. 500

WWT Consulting. 2013. Spatial modelling, wind farm sensitivity scores and GIS mapping tool. Report to Natural England.