Grey seal usage maps for MORL/BOWL developments

Phase 2 delivery

Project Name:	BOWL EIA Work Marine Mammals
Reference:	13.03.11.RPS
Project Manager:	Dr Kate Grellier

Drafted by:	Esther Jones and Jason Matthiopoulos
QA by:	Dr Kate Grellier
Date:	Wednesday, 07 December 2011

VAT reg. No. GB 607 6064 48

SMRU LIMITED is a limited company registered in Scotland, Registered Number: 296937.
Registered Office: 5 Atholl Crescent, Edinburgh EH3 8EJ

TABLE OF CONTENTS

GREY SEAL USAGE MAPS FOR MORL/BOWL DEVELOPMENTS PHASE 2 DELIVERY

AUTHORS: **ESTHER JONES**[1] & **JASON MATTHIOPOULOS**[1,2]

1. *Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, KY16 8LB*
2. *Centre for Ecological & Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, St Andrews, KY16 9LZ*

SUMMARY

Grey seal (*Halichoerus grypus*) telemetry data from 1992-2008 were combined with aerial survey data from 1988-2009 to produce maps of estimated total, at-sea, and hauled-out usage in a study area surrounding the proposed Moray Offshore Renewables Ltd (MORL) and Beatrice Offshore Wind Farm Ltd (BOWL) wind farm developments.

Temporal and demographic (age-class and sex) information on seals were aggregated to produce the spatial maps.

The usage maps in this (Phase 2) report utilise developments in the software used to produce the maps. As such, there have been several key changes:

1. The Phase 1 report showed tracks that had zero and non-zero aerial survey counts associated with them. To improve transparency between the number of tracks shown in Table 1 and Figures 1 and 2, and the usage maps (Figure 6, 7 and 8), only tracks that were matched to non-zero aerial survey counts were used in the final analysis. This explains why 7 tracks shown in Table 1 and Figures 1 and 2 of the Phase 1 report do not feature in the corresponding Table 1 and Figures 1 and 2 of this (Phase 2) report.
2. Additional data have become available since the Phase 1 report was submitted. This has resulted in 11 additional tracks being incorporated into the final analysis (Table 1; Figures 1 and 2).

3. The second driver in the differences between total and at-sea usage maps (Figures 6 and 7) is the development of a new null usage model (see Null (Accessibility) Model). This is a bespoke model using only telemetry tracks from the analysis, and therefore represents a more accurate picture of null usage at the study site. In the Phase 1 report, a generalised null usage map was used.
4. Because the total and at-sea usage appear graphically similar, a hauled-out usage map was included to give a clearer understanding of where the differences in total and at-sea usage occurs.

METHODS

AVAILABLE DATA

AERIAL SURVEY

Aerial surveys are conducted each year by the Sea Mammal Research Unit (SMRU) and are funded by Scottish National Heritage (SNH) and the National Environmental Research Council (NERC). They take place throughout August and both grey and harbour seals (*Phoca vitulina*) are counted. At that time, harbour seals are moulting and are in aggregated groups. Grey seals are in dispersed haul-outs along the coast. Over a number of consecutive years the entire Scottish coastline is surveyed and counts are marked using OS Landranger maps (1:50,000) to an accuracy of 50m. Data from 1996-2009 surveys were used in the analysis.

Fixed-wing aerial surveys for grey and harbour seals are also carried out during the August moult each year in the Moray Firth, Tay Estuary and The Wash in East Anglia. These surveys are funded by NERC and Natural England. Counts from 1988-2009 were used in the analysis.

TELEMETRY

Telemetry data from individual grey seals have been collected by SMRU since 1988. These comprise two sources: Satellite Relay Data Logger (SRDL) tags developed by SMRU use the Argos satellite system and were deployed between 1988 and 2010. GPS phone tags that use the GSM mobile phone network with a hybrid Fastloc protocol (McConnell *et al*., 2004) have been deployed since 2007.

Telemetry data were selected from the SMRU database by species and processed through a set of data cleansing protocols to remove null and missing values, duplicated records and ineligible data. Tracks were then selected based on the criteria that if any part of a track passed within a 100km buffer zone of the proposed MORL/BOWL development sites, regardless of where tagging had taken place, that track was included. Forty four tracks were used in the final analysis (Table 1), from seals that were tagged between 1992 and 2008 (mostly using Argos tags). Thirty seven of the tagged animals were adults, four were juveniles and three were moulted pups. The male to female ratio was 26:18.

Year	Tag type	Number of tags	Sex ratio (m:f)	Mean tag lifespan (days)	Mean number of location fixes (per day)
1992	Argos	4	2:2	142	1.9
1995	Argos	2	1:1	111	0.4
1996	Argos	9	5:4	44	1.3
1997	Argos	2	1:1	106	0.8
1998	Argos	16	12:4	160	0.8
2001	Argos	1	0:1	73	0.6
2002	Argos	2	2:0	130	0.5
2003	Argos	1	1:0	215	4.7
2004	Argos	1	0:1	130	7.4
2008	Argos/GPS	6 - (2 Arg, 4 GPS)	2:4	208	2.7
	TOTAL	**44**	**MEAN**	**134**	**1.6**

Table 1. Summary of telemetry tracks used in the final analysis.

Figure 1 shows the geographical locations of tracks used in the analysis, split by tag type. GPS tags have a smaller spatial extent, concentrated to the south of the Moray Firth. Figure 2 shows the tracks split by year from 1992-2008.

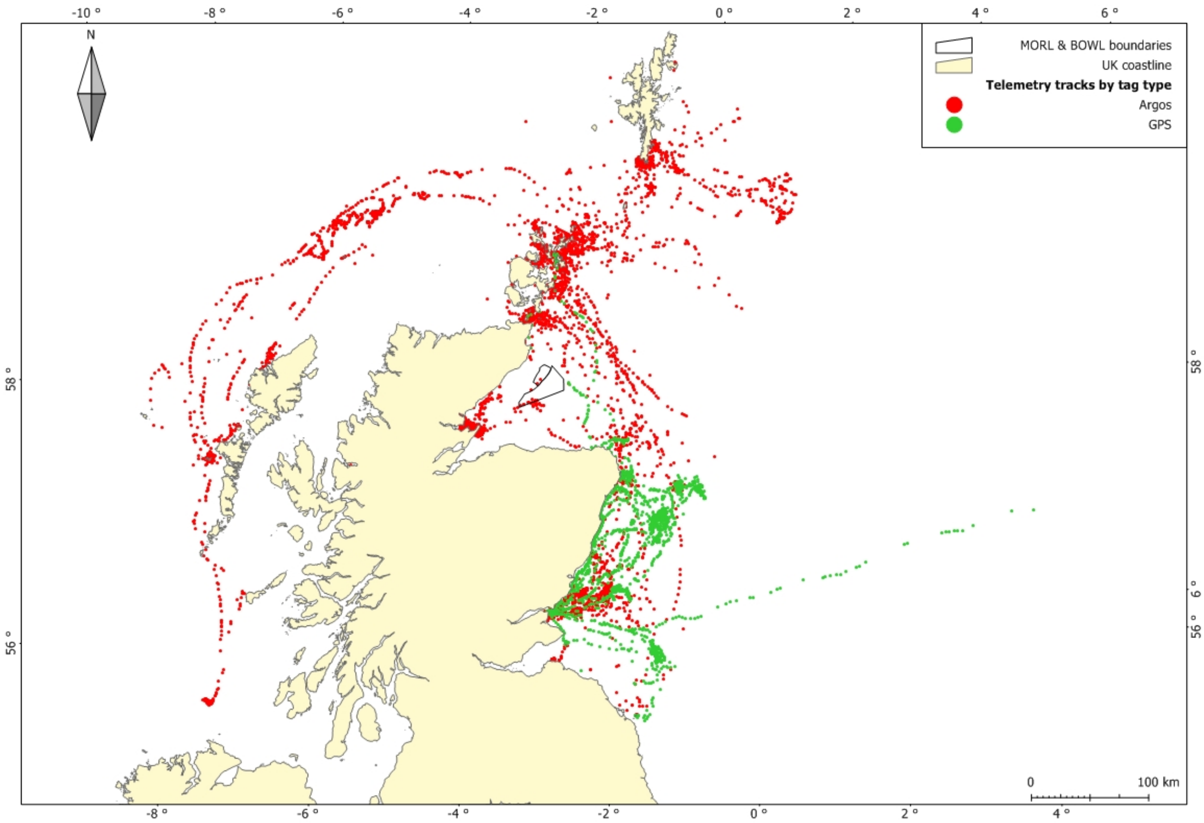

Figure 1. Map showing telemetry track locations by tag type.

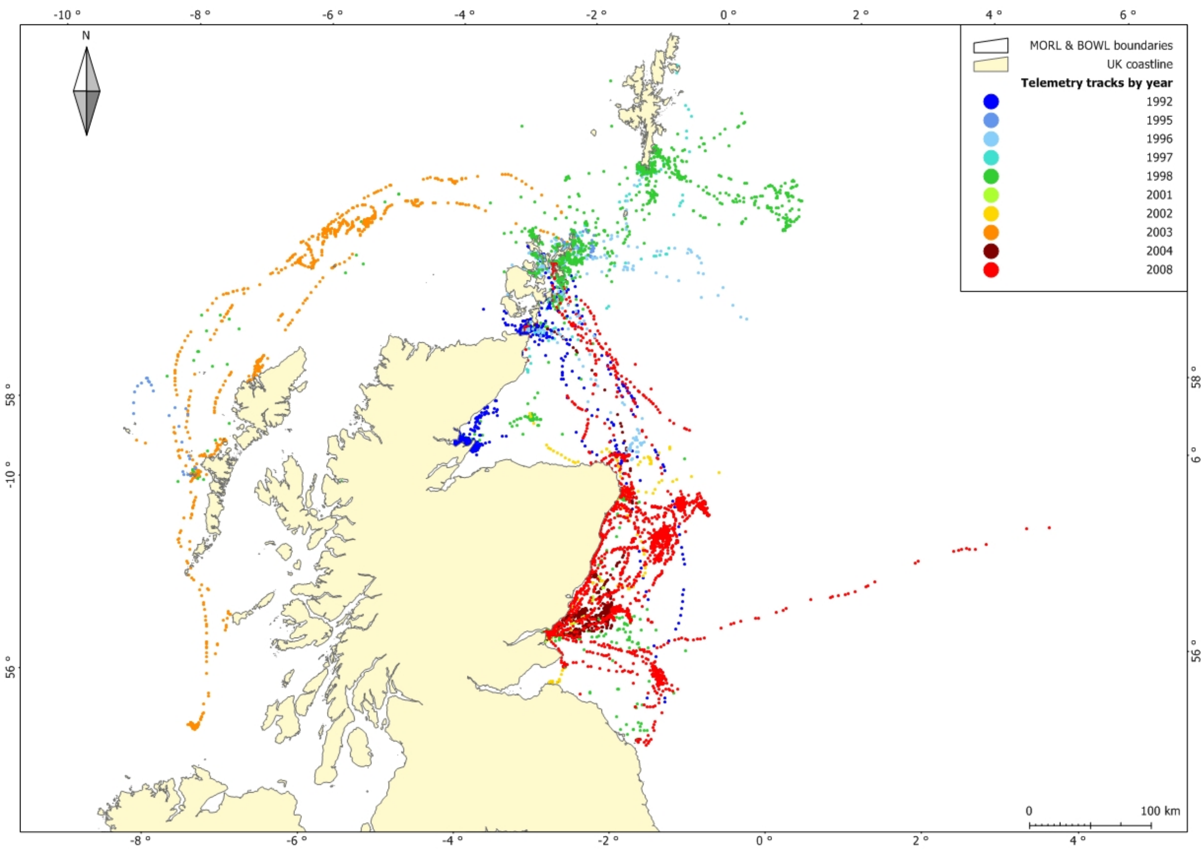

Figure 2. Map showing telemetry track locations by year.

UK COASTLINE

GSHHS 2.2.0 fine (f) resolution L1 data (Wessel & Smith, 1996) available to download from NOAA was used as the UK coastline layer in the usage maps.

SOFTWARE

The statistical package R (R Development Core Team, 2011) was used for analysis and GIS software Manifold version 8.0 was used to produce the maps. Both maps are in Universal projection Transverse Mercator zone 30^0 North (UTM30N), datum World Geodetic System 1984 (WGS84).

SPATIAL EXTENT

To provide continuity within the EIA, the same $4km^2$ grid was used in the analysis as was used by Bailey & Thompson (2011). However, the spatial range was extended to incorporate all of a telemetry track if one or more locations in that track were located within 100km of the boundaries of the proposed offshore wind farm developments.

TREATMENT OF POSITIONAL ERROR

Positional error, varying from 50m to over 2.5km (Argos User's Manual, 2011), affects all Argos telemetry points leading to a loss in fine-scale detail. The range of positional error is defined by the number of uplinks received during a satellite pass. Errors are assigned to six location classes: '0', '1', '2' and '3' indicate four or more uplinks have been received for a location, 'A' denotes three uplinks, and 'B' denotes two uplinks (Vincent *et al*., 2002). Because seals spend the majority of their time underwater, uplink probability is reduced and so over 75% of the telemetry data have location class error 'A' or 'B'.

There are many approaches to addressing the problem ranging from simple moving average smoothers to elaborate state-space models, but none have offered a comprehensive solution combining automation, computational speed, precision and accuracy. Since we are interested in large-scale population-level inferences rather than high-resolution individual-based insights we opted for a Kalman filter (Royer & Lutcavage, 2008; Patterson *et al.*, 2010; Roweis & Ghahramani, 1999) using a linear Gaussian state space model to obtain estimates with error accounted for. This has been developed in-house to give flexibility and fast processing times. Argos data were first speed-filtered (McConnell *et al*., 1992) at $2ms^{-1}$ to eliminate locations that would require an unrealistic travel speed between locations (Russell *et al.*, 2011). Observation model parameters were provided by the location class errors described above, and process model parameters were derived from Vincent *et al*. (2002).

GPS tags are generally more accurate than Argos tags, and 95% of these data have a distance error of less than 50m. However, occasional errors do arise and these data were excluded from the analysis by removing data with residuals that were either 0 or greater than 25, and removing locations with less than 5 satellite fixes (Russell *et al.*, 2011).

HAUL-OUT DETECTION

SRDL and GPS telemetry tags record the start of a haul-out event once the tag sensor has been continuously dry for 10 minutes. This event ends when the tag has been continuously wet for 40 seconds. Haul-out event data were combined with positional data and assigned to geographical locations. In the intervening period between successive haul-out events, a tagged animal was assumed to be at sea (if the tag provided such information) or in an unknown state (if the tag did not).

HAUL-OUT AGGREGATION

Haul-out sites were defined by the telemetry data as any coastal location where at least one haul-out event had occurred, aggregated into $4km^2$ grid cells.

KERNEL SMOOTHING

Kernel smoothing is a statistical technique, which fits a smooth spatial usage surface to a set of positional data (Matthiopoulos, 2003). The KS (Chacon & Duong, 2010; Duong & Hazelton, 2003; Wand & Jones, 1994; Wand & Jones, 1995) library in R was used to estimate the spatial bandwidth of the 2D kernel applied to the telemetry data.

HAUL-OUT DENSITY

Hauled-out usage was calculated by multiplying the proportion of telemetry points at each haul-out site by the estimated kernel smoothed at-sea density.

AERIAL SURVEY WEIGHTING

Aerial survey counts were used to scale up the telemetry data. Aerial survey data were weighted linearly, giving increasing importance to more recent data, to produce a single count incorporating all available years for each 4km^2 grid cell where animals had been counted.

INFORMATION CONTENT WEIGHTING

To account for individual variation in the telemetry points collected from each animal, indexes of information content were devised from 60 remaining tracks (see Appendix – data waterfall). Generalised Additive Models (GAMs) using the R library MGCV (Wood, 2011; Wood, 2006) were built separately for total and at-sea usage. The response variable was rate of discovery, defined by the number of new 4km^2 grid cells an animal 'discovers' in the lifespan of the telemetry tag. This rate was modelled as a function of the number of received telemetry locations for an animal, tag lifespan and whether the tag was Argos or GPS. The intercept was set to zero and a Poisson distribution with a log link function was used.

Figure 3a shows a boxplot of tag type vs. discovery rate for total usage (at-sea data produced very similar results and is therefore not shown). The mean number of grid cells discovered throughout a tag's lifespan are shown by red triangles (Argos = 163, GPS = 489). A Welch two-sample t-test gave a significant difference between the means at the 90% confidence level. This was driven by a significantly higher tag lifespan (Figure 3b; Argos = 2775 hours, GPS = 5474 hours), and higher uplink rate per hour (Figure 3c; Argos = 0.194, GPS = 0.484).

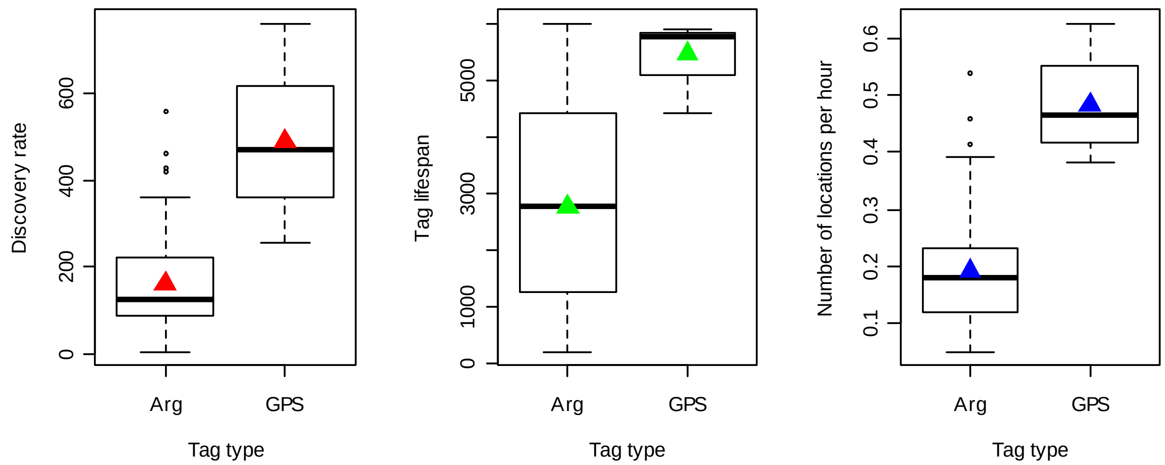

Figure 3. Boxplots showing significant differences between tag types. Coloured triangles represent mean values, thick black lines are median values, boxes are interquartile ranges, and dotted lines show minimum and maximum values. (L-R): 3a. Discovery rate; 3b. Tag lifespan; 3c. Number of locations per hour.

Number of locations, tag lifespan, and tag type (Argos or GPS) were significant and explained 87.3% and 88.2% of variation in the data for total and at-sea usage models respectively. Figure 4a shows total usage fitted values vs. observed discovery rate. Figures 4b and 4c show the GAM smoothing curves for tag lifespan and number of telemetry locations.

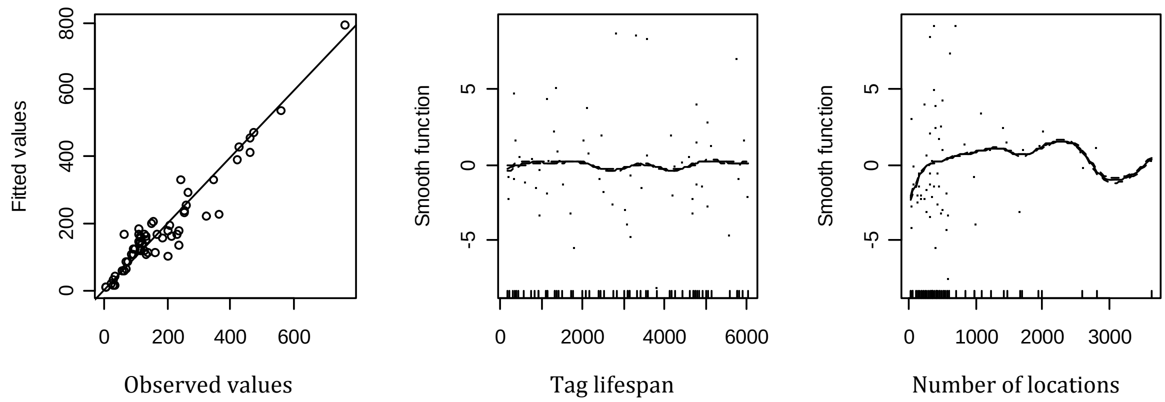

Figure 4. GAM model deriving 'information content' by individual. (L-R): 4a. Observed vs. fitted values; 4b. Tag lifespan smoothing curve; 4c. Number of locations smoothing curve.

Fitted values were normalised and used to weight the contribution of different animals to estimated usage associated with each haul-out location. This approach reduced the importance of data-poor animals, whilst simultaneously not overstating the contribution of animals with heavily auto-correlated observations.

POPULATION SCALING

Grey seals haul-out for approximately 35% of their time (with 95% confidence intervals between 32% and 38%) during the summer irrespective of sex, length (as a proxy to age), region (i.e. location), or survey timing (Lonergan *et al*., 2010). Therefore, to scale the weighted aerial survey counts up to a population estimate, a scalar multiplier of 2.85 (100/35.05) was applied.

NULL (ACCESSIBILITY) MODEL

To account for areas in the maps where aerial survey data were present but telemetry data were not, null maps of estimated density were produced. GLMs were used to model the number of telemetry locations associated with each haul-out. This count was modelled using at-sea distance from the haul-out to represent accessibility by animals to each haul-out, and the distance to the shore to represent accessibility to the coast. A random sub-sample of 25 grey seal tracks were selected. A Poisson distribution with a log link function was used. Figure 5 shows the observed vs. fitted number of telemetry locations associated with each haul-out for grey seals.

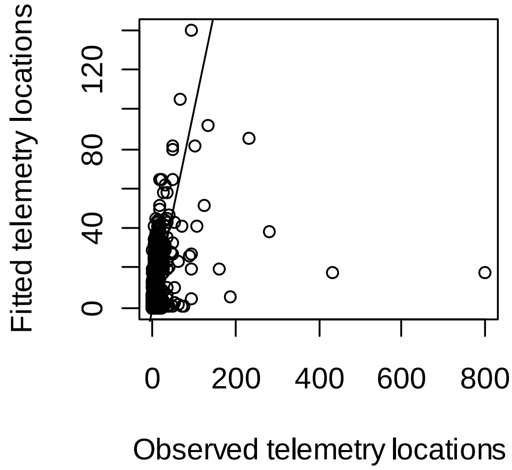

Figure 5. GLM model deriving null usage. Observed vs. fitted locations for grey seals.

CONFIDENCE INTERVALS

Uncertainty within haul-outs was propagated through the analysis using two sources: by estimating the variability in the telemetry data and using variability in the null usage models.

TELEMETRY DATA VARIANCE

Linear Models (LMs) were built to estimate variance. All haul-outs with more than 6 animals associated with them were used. The response variable was logged variance and covariates were sample size (number of animals associated with a haul-out), logged estimated mean density of seals weighted by information content, and the interaction between them. At-sea kernel smoothed densities were bootstrapped 500 times for each haul-out, and sample size was sampled with replacement and logged, to produce estimated logged variance and logged mean densities.

NULL USAGE VARIANCE

Estimated mean densities in the null maps were produced using a Poisson log link distribution. Therefore, the variance in these maps was equal to the mean.

According to the central limit theorem, the aggregated variance maps were normally distributed and so were scaled up to confidence intervals using a scalar multiplier of 1.96.

ANALYSIS

To create single maps of total usage and at-sea usage, all grey seal telemetry data from the SMRU database was put through a series of data cleansing protocols to remove unusable data (Appendix – data waterfall). Argos data were spatially interpolated using a Kalman filter and merged with GPS data. Tracks were then selected based on the criteria that any part of a track passed within a 100km buffer zone of the proposed MORL/BOWL development sites, regardless of where tagging had taken place.

A $4km^2$ grid was created to extend to the limits of the telemetry tracks and overlaid onto the data. Haul-out detection and aggregation were applied to the data at 4km resolution. After spending time at sea an animal could either return to its original haul-out (classifying this part of the data as a return trip), or move to a new haul-out (giving rise to a transition trip). Return trips were attributed to the departure haul-out. Transition trips were divided temporally into two equal parts and the corresponding telemetry data were attributed to departure and termination haul-outs.

At-sea data (i.e. when animals were not hauled-out) were then kernel smoothed. A bandwidth was estimated for each animal. Each animal/haul-out combination was kernel smoothed using the estimated bandwidth to produce separate animal/haul-out association distribution maps.

For total usage, each animal/haul-out map was multiplied by a normalised Information Content Weighting to correct for individual animal bias. All maps connected to each haul-out were aggregated and hauled-out density was added onto each map. Each map was then scaled to the estimated number of animals using that haul-out using the weighted aerial survey counts and then further scaled to the population estimate. A null usage map was derived for each aerial survey site without corresponding telemetry data. Each map was normalised, scaled to aerial survey counts and population estimates, and added to the total usage map.

For at-sea usage, each animal/haul-out map was multiplied by the normalised at-sea Information Content Weighting. Each map was normalised and multiplied by the proportion of telemetry locations not hauled-out. All maps connected to each haul-out were aggregated and scaled to weighted aerial survey counts and then population estimate. Null usage maps were derived using the same process as total usage, but were multiplied by the total proportion of time animals spent not hauled-out (see Population Scaling above) before being added to the at-sea usage map.

Variance in the telemetry data was then estimated for each total and at-sea usage. For total usage, the uncertainty models predicted variance by grid cell for the animals associated with each haul-out, which were then aggregated over all haul-outs. The models were applied in the same way to at-sea usage and both sets of variance maps were scaled to aerial survey counts and population estimates. For the null usage maps, variance was equal to estimated density. Each grid cell was normalised and scaled appropriately to population estimates for total and at-sea usage and added to the telemetry data variance maps. The maps were then scaled up to confidence intervals.

Hauled-out usage and variance was calculated by subtracting the at-sea usage and variance from the total usage and variance.

USAGE MAPS

Figure 6 shows spatial usage of grey seals around the proposed MORL/BOWL development sites. The map can be interpreted as the average number of seals in each 4km^2 grid cell at any point in time. For example, a green square denotes that, on average, between 1 and 5 grey seals will be within that grid square at any point in time. A red square denotes that over 50 animals will be in that grid square at any point in time.

White contour lines denote standard deviation from the mean as a measure of uncertainty magnitude around estimated usage. Labels show the value of standard deviation at each contour as the square root of the estimated variance. This in turn was a combination of two modelling processes: null usage and telemetry data. Variance from the null model was larger than for estimates informed by telemetry data. Therefore, in regions that received considerable usage from haul-outs for which no telemetry data are available, it is often the case that uncertainty contours appear smoother than the usage density map. This is a desirable feature of the model: it inflates uncertainty in regions where the ratio of data/usage is likely to be low.

Within the study area, highest usage is located in the Dornoch Firth, Pentland Firth, and around the Orkney Islands. Possible offshore foraging patches can also be seen throughout the study area, denoted in orange and yellow.

Figure 7 shows estimated grey seal at-sea usage with white contour lines denoting standard deviation. Total and at-sea usage display similar characteristics, although at-sea usage is 28% lower due to the removal of hauled-out usage.

Figure 8 shows estimated grey seal hauled-out usage with white contour lines denoting standard deviation. The highest usage occurs in the Dornoch Firth and at some of the islands in the Pentland Firth and around Orkney.

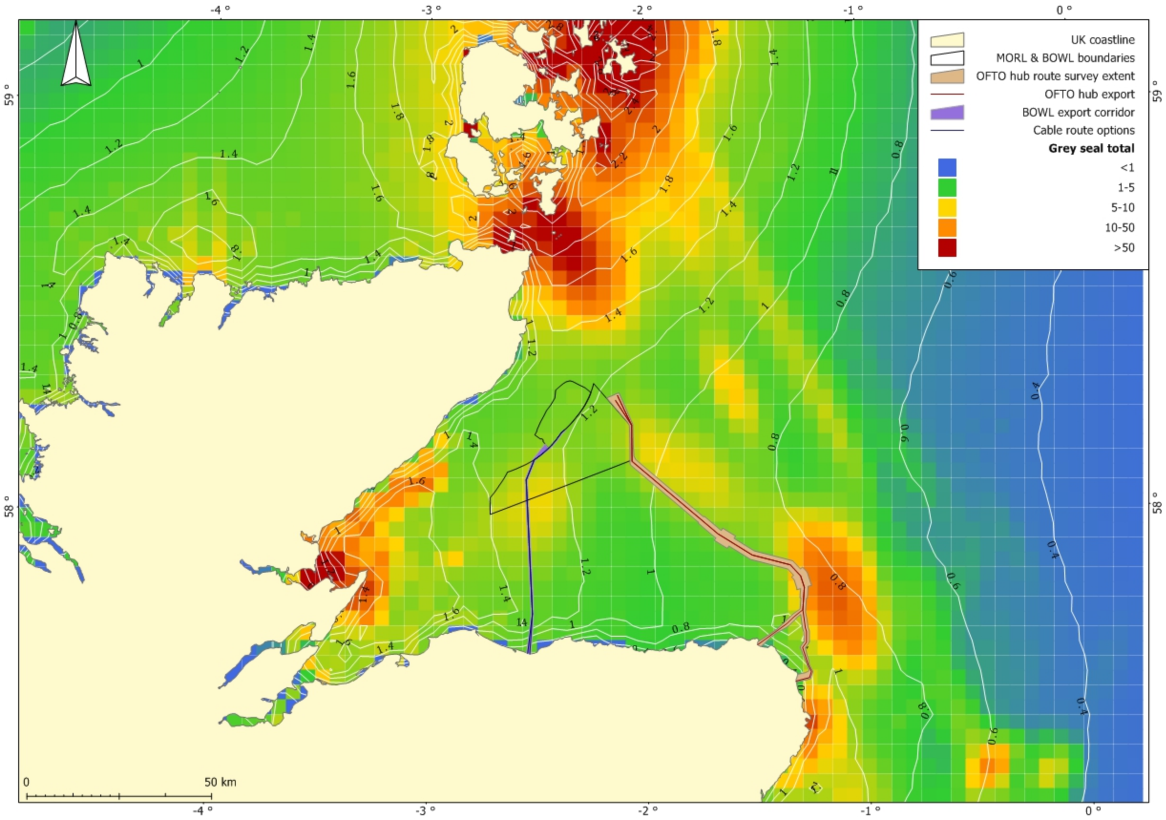

Figure 6. Estimated grey seal total (at-sea and hauled-out) usage around the proposed MORL/BOWL development sites. White contours show standard deviation from mean usage as a measure of uncertainty.

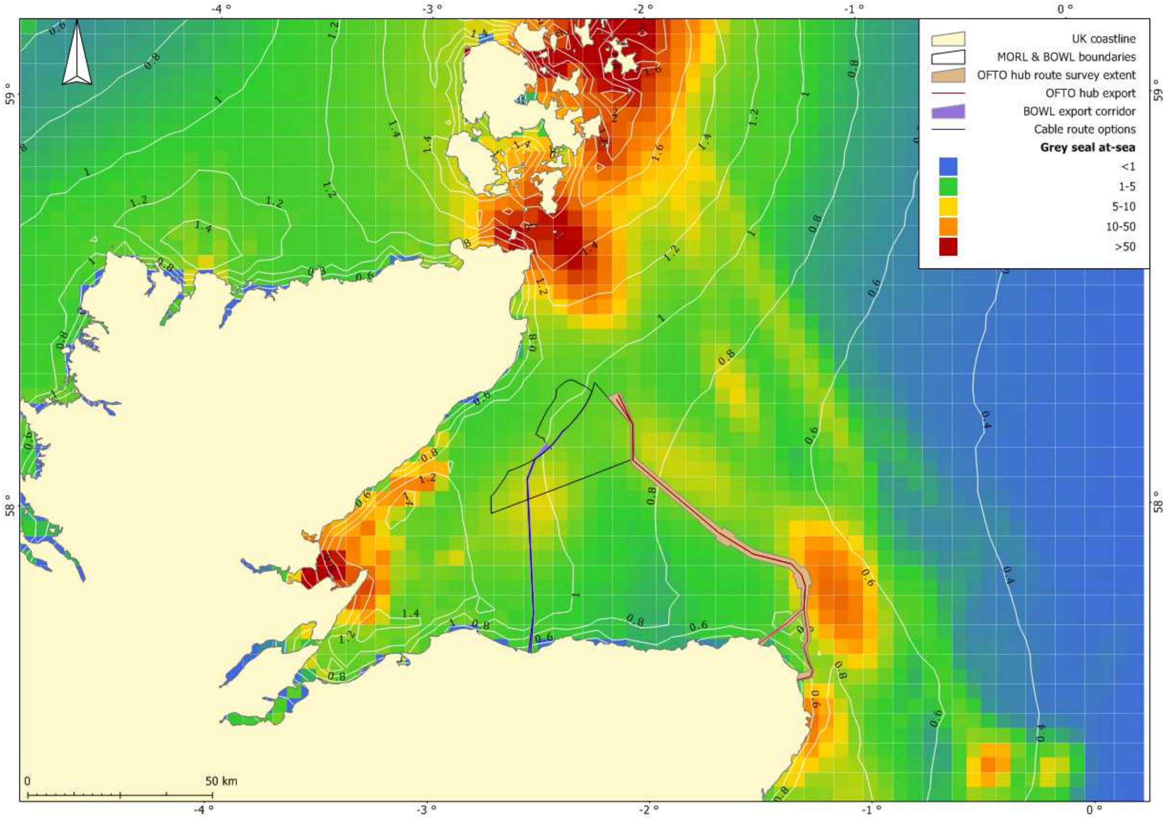

Figure 7. Estimated grey seal at-sea usage around the proposed MORL/BOWL development sites. White contours show standard deviation from mean usage as a measure of uncertainty.

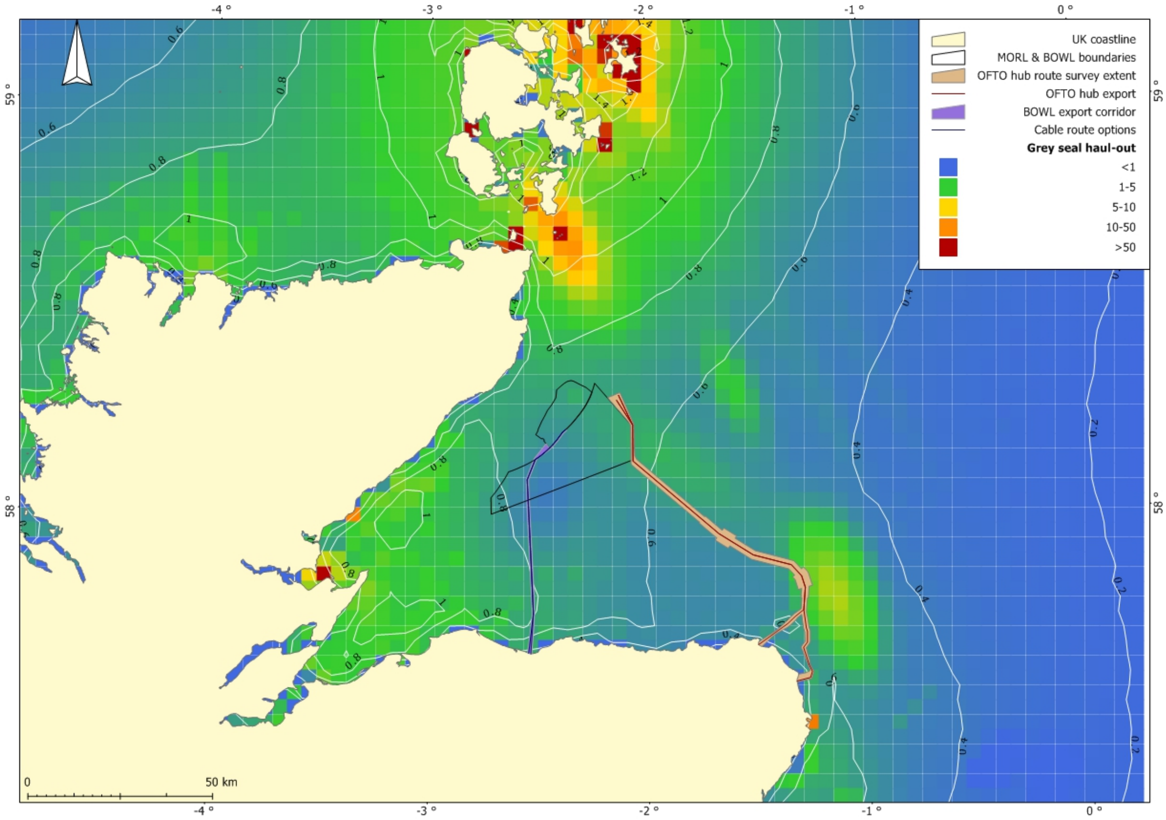

Figure 8. Estimated grey seal hauled-out usage around the proposed MORL/BOWL development sites. White contours show standard deviation from mean usage as a measure of uncertainty.

REFERENCES

Argos User's Manual. (2011) 2007-2011 CLS.

Bailey, H, & Thompson, P. (2011) Technical report on harbour seal telemetry and habitat model. Report to MORL. August 19th 2011.

Chacon J.E. & Duong T. (2010) Multivariate plug-in bandwidth selection with unconstrained pilot matrices. *Test*. **19**:375-398.

Duong T. & Hazelton M.L. (2003) Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*. **15**:17-30.

Lonergan, M, McConnell, B, Duck, C, & Thompson, D. (2010) An estimate of the size of the UK grey seals population based on summer haulout counts and telemetry data. *SCOS Briefing paper 10/4*.

Matthiopoulos J. (2003) Model-supervised kernel smoothing for the estimation of spatial usage. *Oikos*. **102**(2):367-377.

Matthiopoulos J, McConnell B.J, Duck C & Fedak M.A.(2004) Using satellite telemetry and aerial counts to estimate space use by grey seals around the British Isles*. Journal of Applied Ecology*. **41**(3):476-491.

McConnell B, Beaton R, Bryant E, Hunter C, Lovell P, Hall A. (2004) Phoning home - A new GSM mobile phone telemetry system to collect mark-recapture data. *Marine Mammal Science.* **20**:274-283.

McConnell B.J, Chambers, C, Fedak MA. (1992) Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the Southern Ocean. *Antarctic Science* **4**: 393-398.

Patterson T.A, McConnell B.J, Fedak M.A, Bravington, M.V, Hindell, M.A. (2010) Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. *Ecology*. **91**(1):273-85.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Royer F & Lutcavage M. (2008) Filtering and interpreting location errors in satellite telemetry of marine animals. *Journal of Experimental Marine Biology and Ecology*. **359**(1):1-10.

Roweis, S and Ghahramani, Z.(1999) A Unifying Review of Linear Gaussian Models. *Neural Computation.* **11**:305-345.

Russell, DJF, Matthiopoulos, J, & McConnell, BJ. (2011) SMRU seal telemetry quality control process. *SCOS Briefing paper (11/17).*

Vincent C, McConnell B.J, Ridoux V, Fedak MA. (2002) Assessment of Argos Location Accuracy From Satellite Tags Deployed on Captive Gray Seals. *Marine Mammal Science*. **224**(2):223-166.

Wand, M.P. & Jones, M.C. (1994) Multivariate plugin bandwidth selection. *Comp Statistics*. **9**:97-116.

Wand, M.P. & Jones, M.C. (1995) Kernel Smoothing. Chapman & Hall. London.

Wessel, P, and Smith, W.H.F. (1996) A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res., 101*, 8741-8743.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. Royal Stat Soc (B)*. **73**(1):3-36.

Wood, S.N. (2006) Generalized Additive Models: an Introduction with R, CRC.

APPENDIX – DATA WATERFALL

Description	**Counts**	**Drop-off #**	**Drop-off %**	**# animal tracks**
Haulout				
Species=grey or harbour seal	96,949			650
Remove missing / null values	96,934	15	100%	650
Argos				
Species=grey or harbour seal	507,300			603
Remove missing / null values	507,283	17	0%	603
Merging biological information	331,649	175,634	35%	434
Remove invalid / duplicated records	331,127	522	0%	428
Deduplication for Kalman filtering	236,264	94,863	29%	427
GPS				
Species=grey or harbour seal	362,040			182
Remove missing/null values & exclusion points	362,039	1	0%	181
Merging biological information	130,931	231,108	64%	181
Remove invalid/duplicated records	128,502	2,429	2%	181
Merge telemetry data				
Merge Argos & GPS	364,766	0	0%	484
Retain grey seals	155,250	209,516	57%	262
Retain animals whose tracks appear at least once within 100km buffer of MORL/BOWL sites	47,572	107,678	69%	76
Haulout detection				
Assign haulout events	47,572	0	0%	76
Only retain animals that have at least one haulout event associated with them	39,805	7,767	16%	60
Trip detection				
Assign trips & grid cell IDs to each animal / location	39,805	0	0%	60
Remove observations outwith location range of the overlying grid cells	39,524	281	1%	60
Kernel smoothing				
Remove haul-out events	32,341	7,183	18%	60
NULL usage				219
Information content weighting				
Total usage information content weighting	39,524	0	0%	60
At-sea information content weighting	32,341	0	0%	60
Usage maps				
(Total usage) Match haul-outs with aerial survey data	9,379	30,145	76%	44
(At-sea) Match haul-outs with aerial survey data	7,570	24,771	77%	44
(Hauled-out) Match haul-outs with aerial survey data	1,809	- 1,809		44