Technical brief –

update of harbour porpoise summer distribution in the North Sea

University of Veterinary Medicine Hannover, Foundation

Institute for Terrestrial and Aquatic Wildlife Research (ITAW)

FINAL REPORT

Authors

Anita Gilles, Rémi Pigeault, Vincent Benoit, Nadya Ramirez-Martinez

University of Veterinary Medicine Hannover, Foundation, Institute for Terrestrial and Aquatic Wildlife Research, Buesum, Germany

Under contract from Aarhus University, Danish Centre for Environment and Energy (DCE), Roskilde

Contact:

Signe Sveegaard

Büsum, 30 Sep 2025

Table of Contents

INTRODUCTION *4*

METHODS *5*

Spatial modelling **5**

Data preparation and processing 5

Data analysis 7

Model structure, fitting and selection 7

Model evaluation 8

Gap analysis in the environmental space **8**

Hot spot analysis **9**

RESULTS and discussion *11*

Spatial modelling **11**

Searching effort and sightings 11

Model results and predicted density surfaces 13

Model diagnostic 14

Gap analysis in the environmental space **16**

Hot spot analysis **17**

References *21*

APPENDIX *25*

Model diagnostic 25

Hotspot analysis 25

INTRODUCTION

A critical step towards understanding how sensitive the harbour porpoise (*Phocoena Phocoena*) is to offshore renewable energy development in different regions of the North Sea is to predict its distribution and determine which areas can be characterized as important areas for the species, e.g. serving as foraging or reproduction habitats.

Predictive modelling of marine mammal habitat is a powerful tool in marine science as it integrates heterogeneity in marine ecosystems and provides important information for ecological studies, management purposes, mitigation of anthropogenic impacts (e.g. Forney et al. 2012) and for understanding the processes that influence interannual and seasonal variability in species distributions (Gilles et al. 2011, 2016, 2025; Becker et al. 2016, 2017, 2018; Pigeault et al. 2024a). Marine mammals exhibit discernible fluctuations in their distributions. However, the drivers responsible for these variations remain unclear, in part due to the dynamic and poorly understood ecological relationships between, for instance, harbour porpoises, their environment and their prey species. In the absence of a comprehensive understanding of these ecological relationships, the here applied species distribution modelling can facilitate the elucidation of these relationships for harbour porpoises in the North Sea. The objective of this study is to predict and map the long-term summer distribution of harbour porpoises by fitting a habitat-based density model to the most recent visual survey data collected for harbour porpoises in the "Greater North Sea" assessment unit (as defined for harbour porpoises in the OSPAR Maritime Area; Geelhoed et al. 2022). The model is essentially updated from that of Gilles et al. (2016) for the North Sea, using the most recent data.

This spatial modelling approach, described in Gilles et al. (2016), utilises physical and biological characteristics as proxies for prey abundance. The ecological theory of species distribution modelling assumes that distribution is at least partly related to environmental variables, and that the relationship between species occurrence and environmental parameters can therefore be used to predict the distribution of the species in question (Austin 2007). Although behavioural factors such as migration, predator avoidance and social interactions influence the distribution of cetaceans, many of the distribution patterns are determined by the foraging response of top predators in a dynamic system (Redfern et al. 2006). This is particularly evident in the case of the harbour porpoise in the North Sea, which exhibits a marked preference for predictable hotspots characterised by high food availability. In the absence of specific prey density data at the necessary spatial resolution for habitat prediction modelling, physical and biological characteristics of the sea are utilised as proxies.

The result of the project can be used for sensitivity mapping of harbour porpoises in the context of marine spatial planning or in assessing sensitivities towards a broad range of anthropogenic activities in the North Sea.

METHODS

Spatial modelling

The modelling process used in Gilles et al. (2016), applying the workflow described in Gilles et al. (2025) for the spatial modelling of the SCANS-IV survey data, was reproduced for this study. Briefly, the modelling framework is based on Generalized Additive Models (GAMs) to analyse the survey data, with the objective of establishing the relationship between the number of observed animals (the response variable) and the environmental variables (the explanatory variable) (Hastie & Tibshirani 1990, Wood 2017). This framework allows to model non-linear relationships between cetacean habitat and the marine environment.

Data preparation and processing

A comprehensive dataset was compiled from multiple dedicated line-transect aerial surveys, encompassing the "Greater North Sea" assessment unit (as delineated for harbour porpoises in the OSPAR Maritime Area; Geelhoed et al. 2022), or parts thereof, during the period spanning from 2020 to 2024. The data presented herein were obtained from two sources: the international large-scale SCANS-IV survey (Gilles et al. 2023), conducted in the summer of 2022, and the national monitoring of the North Sea waters by Germany (2020-2024) and Denmark (2020-2024), e.g. the Danish surveys of southern Danish North Sea and Skagerrak, and summer surveys from the pre-investigations by the planned North Sea Energy Island and North Sea I (from 2022-2024) as well as the survey of the complete Danish North Sea and Skagerrak in the summer of 2023. Further data, submitted to OSPAR's QSR 2023 (Geelhoed et al. 2022) were explored, but in this case of modelling data from the most recent period, only survey data from one year (2020) were available for Belgium waters. No survey data were available for analysis from the national monitoring of the Netherlands in this 4-year period.

All surveys followed the standardized SCANS field protocol and used the same data collection software (Scheidat et al. 2008; Gilles et al. 2009, 2016, 2025; Hammond et al. 2013, 2021; SAMMOA 2022). Survey effort and sightings data were quality-checked and several plausibility checks were performed. Data were then aggregated into the common dataframe.

The line-transect effort data were segmented into continuous portions of effort of approximate 10 km mean length, conforming with Becker et al. (2020), Gilles et al. (2016, 2025) and Virgili et al. (2019). Harbour porpoise sighting data were assigned to each segment. The effective area searched was estimated for each segment, based on the survey-specific effective strip widths (including g(0)). The effective area searched was subsequently included as an offset in the model structure. This procedure accounts for both varying segment lengths and the different detection probabilities recorded during

the surveys. The covariates were extracted at a daily resolution within a buffer of 5 km around the segments (see **Table 1** for candidate covariates). A suite of environmental covariates, i.e. spatial, static and dynamic covariates, were considered. It can be argued that the selected habitat predictors are most likely proxies for unmeasured underlying ecological processes driving species distributions rather than direct drivers.

***Table 1**. Candidate environmental covariates used in the density surface models for the Danish project, shown to be important in previous modelling exercises (e.g. Gilles et al. 2016, 2025; Lacey et al. 2022; Pigeault et al. 2024b).*

Covariate	Description	Source
X	Longitude converted to ETRS89 (EPSG:3035)	
Y	Latitude converted to ETRS89 (EPSG:3035)	
Water depth (depth)	Mean water depth (m)	EMODnet Digital Bathymetry (DTM 2020).
Slope	Slope of the seabed (°) calculated with R package *raster*, version 3.6-26 (Hijmans 2013).	Derived from bathymetry.
Aspect	The direction that the slope faces (°) calculated with R package *raster*, version 3.6-26 (Hijmans 2013).	
Distance to coast	Distance at segment centroid (km)	EMODnet Bathymetry World Coastline, estimated at highest astronomical tide.
Mean sea surface temperature (SST)	Daily temperature (°C)	Global Ocean Physics Reanalysis. E.U. Copernicus Marine Service Information (CMEMS).
Spatial sea surface temperature gradient on a radius of one cell (gradSST)	Spatial gradient in daily temperature (°C) within the radius of one cell, calculated with R package *grec,* version 1.6.0 (Lau-Medrano 2024)	
Spatial sea surface temperature gradient on a radius of two cells (grad2SST)	Spatial gradient in daily temperature (°C) within the radius of one cell, calculated with R package *grec,* version 1.6.0 (Lau-Medrano 2024)	
Bottom temperature (bottomT)	Sea water potential temperature at sea floor (°C)	
Difference of temperature between surface and sea floor (ΔTemperature)	Difference between SST and bottomT (°C)	

Eddy kinetic energy (EKE)	Eddies calculated as the current velocity (m/s)	
Mixed layer thickness (MLD)	Ocean mixed layer thickness defined by sigma theta (m)	
Net primary productivity (NPPV)	Expressed as carbon in sea water (mg/m^3 /day)	Global Ocean Biogeochemistry Hindcast. E.U. Copernicus Marine Service Information (CMEMS)
Distance to sandeel grounds	Distance at segment centroid (km)	Jensen et al. 2011

Data analysis

All data processing was undertaken in software R version 4.4.0 (R Core Team 2024), and modelling was conducted using R package *mgcv*, version 1.9-1 (Wood 2017).

Model structure, fitting and selection

A multi-stage modelling approach was implemented with the objective of reducing bias in the density estimates generated from the habitat-based models. Methods largely followed the one described in Gilles et al. (2025), using Generalised Additive Models (GAM) to link environmental covariates to observations and we refer to this source for detailed information. Briefly, smooth functions were fitted using restricted maximum likelihood (REML) with automatic term selection (Marra & Wood 2011). Cubic regression splines were used for all covariates, with a maximum number of knots set to 10. For the covariate 'aspect', a cyclic cubic regression spline was used. This model-fitting method helps to avoid overfitting of the smooth functions by including a penalization (Marra & Wood 2011). The method can reduce the estimated degrees of freedom of a covariate term towards zero if it does not contribute sufficiently to account for the variability in the data.

In addition to the spatial-autocorrelation smooth *te(X, Y)* of locational covariates, a deviation from this main spatial-autocorrelation smooth was allowed by year including the covariate *te(X, Y, year)*. This was done using the factor smooth *bs = 'fs'*, allowing the spatial autocorrelation to be adjusted to the yearly observations if there were surveys. If no surveys were conducted during one of the years, or only a part of the study area was surveyed, then the factor smooth in this area for this year matches the effect of the main spatial-autocorrelation smooth *te(X, Y)*, so that no extrapolations are involved in this non-surveyed spatio-temporal frame. This is therefore a flexible auto-correlation smooth that allows for yearly deviations from the main effect *te(X, Y)* (Pedersen et al. 2019).

Models were fitted for each possible combination of two to five uncorrelated covariates (that is covariates with a Pearson's pairwise correlation coefficient < 0.50). The five models with the best goodness of fit, based on leave-one-out cross-validation, were

selected and their respective predictions were stacked (Yao et al. 2017) for further investigation, with their respective contribution to the final prediction estimated with the loo R-package (Vehtari et al. 2017).

For accurate uncertainty quantification, a pseudo-posterior approach was taken (e.g. King et al. 2000). The pseudo-posterior approach allows for seamless quantification of uncertainty for any derived quantities (e.g. abundance, CV, 95% confidence interval) from model parameters. Maximum likelihood estimates of parameters and their associated covariance matrix were extracted from fitted models (using the function rmvnorm from package *mvtnorm,* version 1.3-3, Genz & Brentz 2009) and used to generate a sample of 1,000 values from a pseudo-posterior, assuming a multivariate normal distribution for the parameters (King et al. 2000). This sample was used to carry out predictions at a daily level over the survey period. As the variability of some smooth functions was high and the upper values could reach extreme densities that are ecologically unrealistic due to the over-dispersion parameter, a threshold in densities was set to the 99.9% quantile of the initial predicted density (i.e. by sample, cell and day). Densities above this limit were removed from the samples as they could not be considered as ecologically realistic. Finally, the predicted densities were averaged over the survey period for each cell and sample, providing a pseudo-posterior distribution per cell.

The same prediction grid (10x10 km) was used as for SCANS-IV (Gilles et al. 2025), but here the spatial extent was limited to the Greater North Sea. On the temporal extent, daily prediction grids of 10x10 km were created from the first to the last summer survey date, specific for each survey year, with each grid populated using daily covariate mean values.

Model evaluation

The performance of the model was evaluated using several established metrics. These included the percentage of explained deviance, deviance residuals, information criteria and visual inspection of predicted and observed distributions.

QQ plots, degrees of freedom, fitted relationships, predicted species distributions and abundances were inspected for this selection of models. Goodness-of-fit and model performance diagnostics were overall consistent among selected models, and the model contributing the most to the final prediction was finally selected.

Gap analysis in the environmental space

The extrapolations and number of calibration data supporting the prediction of harbour porpoise density (aka nearby data) were analysed using the *dsmextra* R-package, version 1.1.5 (Bouchet et al. 2020). In this context, extrapolations can be defined as univariate (i.e., outside the range of a covariate) or combinatorial (when the combination of the covariate values is outside the range of sampled environments). Nearby data are defined

as calibration data that have similar environments to those of a prediction, thus allowing for a more reliable estimate of harbour porpoise density (King and Zeng 2007). These two metrics allow to examine the robustness of environmental variables for modelling and identify areas where prediction reliability can be impaired (Pigeault et al. 2024b). This so-called 'gap analysis' also assesses the environmental similarity between the calibration data and the predictions, and results can be mapped to visualise uncertainties given the environmental variables used for modelling.

In this work, we have estimated the extrapolations and nearby data for every predicted day. Results were then summarised by calculating the percentage of days over which predictions were extrapolated, and the mean nearby data percentages per prediction over the study period.

Hot spot analysis

Determining "hotspots" or seasonal core habitats, which can be defined as areas of higher-than-average biodiversity or abundance (Harvey et al. 2017), is an important step in comprehending the ecological significance of these regions and facilitating the optimisation of conservation and management strategies. Here, in an exploratory approach, we evaluated different methods on their performance to identify such "hotspots", the most common aspatial approach being the use of arbitrary thresholds of top percentages of grid cells (Parviainen et al. 2009, Tolimieri et al. 2015): the top 5% (i.e., threshold set at the 95th percentile value), 10%, 15% and 20%. We also trialled the use of a spatial statistical method, Getis-Ord Gi* (Getis 2010), implemented with the *spdep* R-package, version 1.3-13 (Bivand et al. 2017). This approach allows to detect spatial clustering of high and low densities (i.e. "hotspots" and "coldspots") as was shown in Harvey et al. (2017) for marine mammals in British Columbia.

The Gi* is a promising metric as it also includes a test hypothesis, which yields p-values and allows to measure the statistical significance of hot and coldspots. Computed on the output of a density surface model (DSM), the Gi* metrics takes a focal cell on the prediction grid and compares it to its neighbourhood to delineate hotspots. The latter are identified when a focal cell and its neighbouring cells are associated with high density values relative to all density values within the predicted grid. A p-value can be computed under a null hypothesis of spatial homogeneity (Harvey et al. 2017 but see also Bivand & Wong 2018). However, a p-value computed on a map of predicted densities from a DSM cannot be interpreted easily as the predictions are under a model which purpose is to estimate an inhomogeneous point process. The null hypothesis assumed for computing a p-value has already been rejected during the model selection process unless the selected model used for predictions includes only an intercept. To identify hotspots, we calculated the Gi* metric first on the DSM output. To compute and interpret a p-value, we also ran this analysis on the effort-corrected observation data. These data were corrected

to take effort into account to ensure that we highlighted the areas with high densities of sightings (i.e., sightings per effort) and not just the areas with the most effort.

Additionally, the hotspots were also overlaid with the observations of mother-calf pairs during the surveys to evaluate whether the detected hotspots are consistent with the presence of calves, therefore indicating areas of high ecological importance for the harbour porpoise during the reproduction period.

RESULTS AND DISCUSSION

Spatial modelling

Searching effort and sightings

Survey data collected in the summer months (June-August) constituted the largest proportion of available data and, accordingly, predictions are valid for the summer. The total number of effort segments, groups and individuals sighted per segment is reported in **Table 2** for each year. The maps of spatial distribution of realised line-transect effort and sightings for the years 2020 to 2024 are presented in **Figure 1**.

We aggregated 46,879 km of on-effort survey data with 4,026 sightings of harbour porpoise groups. From this, a total of 4,739 effort segments, with a final segment mean length of 9.9 km (SD = 1.4 km), were included in the modelling.

***Table 2**. Summary of 2020-2024 aerial survey data, showing number of effort segments, number of groups and individuals sighted in the summer of each year, as used for model fitting.*

Year	Total no. of effort segments	No. of effort segments with groups	% effort segments with groups	Number of groups	Number of individuals
2020	537	270	50	600	755
2021	373	160	43	335	407
2022	2,284	820	36	1,537	2,072
2023	998	429	43	1,019	1,405
2024	547	244	45	535	674
Total	**4,739**	**1,923**	**41**	**4,026**	**5,313**

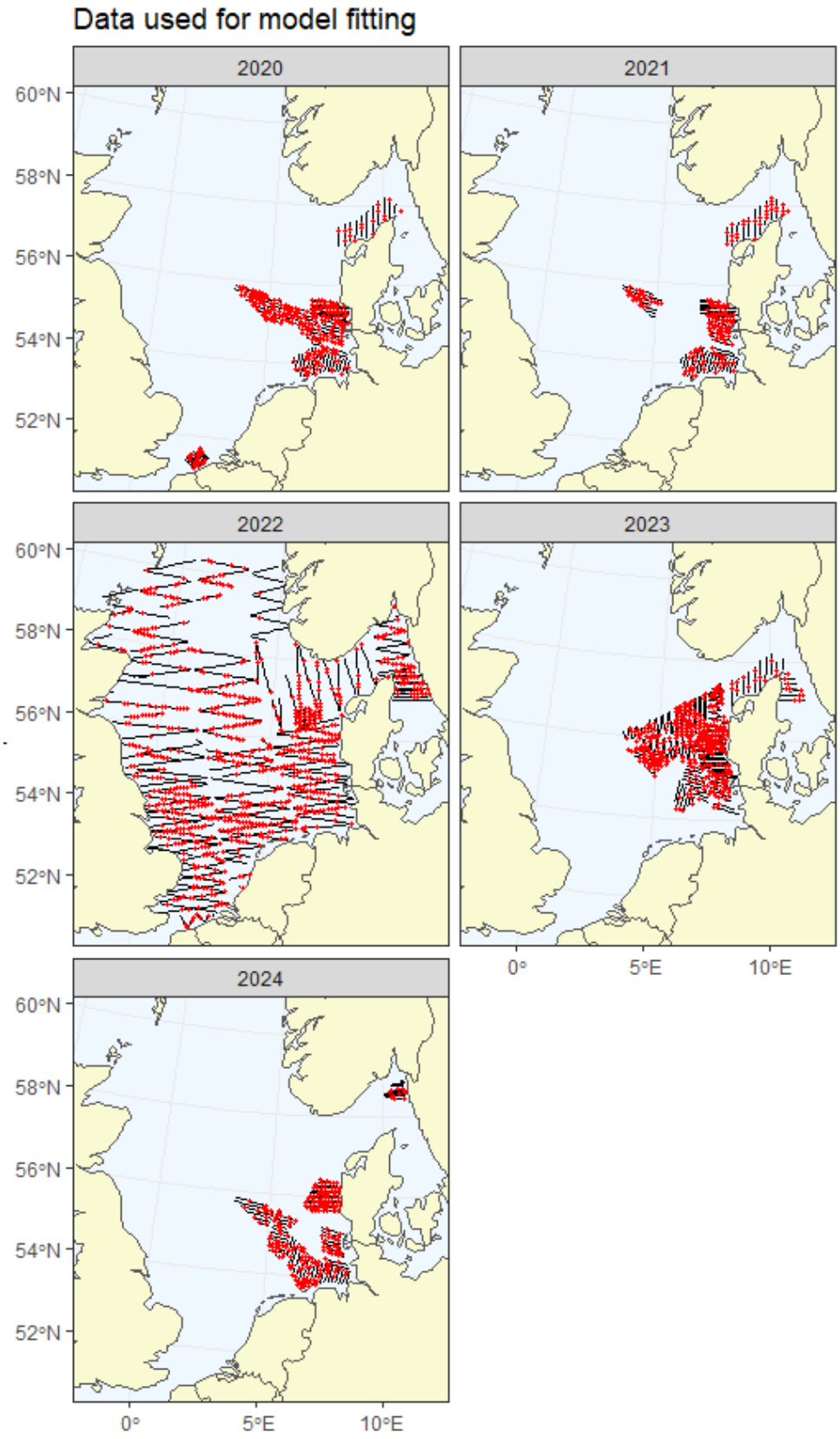

***Figure 1**. Aerial survey effort transects (black) and sightings of harbour porpoise groups (red) for the survey years 2020-2024.*

Model results and predicted density surfaces

The best model explained 14.2% of the deviance using a negative binomial error distribution and a two-dimensional tensor product smooth of location, and further covariates as seen in **Table 3**.

Table 3*. Model description and diagnostics for the final selected model of harbour porpoise individuals. Covariates in the models are described in* ***Table 1****.*

Distribution	Model covariates	Estimated degrees of freedom	Model degrees of freedom	% Deviance explained
Negative Binomial (0.605)	X,Y	17.3	46.5	14.2
	X,Y, year	11.6		
	SST	1.6		
	ΔTemperature	3.9		
	SDtime SST	5.2		
	slope	3.2		
	Distance to sandeel grounds	3.7		

The visualisation of the prediction generated from the GAM clearly delineates regions where the predicted density of porpoises is greatest. The maps showing surfaces of predicted density and associated estimated coefficient of variation (CV) are shown in **Figure 2**. The patterns of predicted density are influenced by the covariates retained in the models (see **Table 3**), the fitted smooth functions (see **Figure 3**), and spatial variation in the values of the covariates in the prediction grid.

The maps of CVs provide a measure of the confidence in predicted density across the survey area. Lower CVs are generally associated with areas of higher density when predictions are interpolations *sensu* Pigeault et al. (2024b). On the other hand, high CV values may betray either extrapolations, a high between-day variance in predictions, or be associated with areas of very low density. CVs for predicted harbour porpoise density (**Figure 2**) are relatively low across most of the survey area, whereas the confidence in the predictions in areas of low density is generally poorer. The magnitude of the CV is influenced by the number of sightings as well as by how well the models fit the data.

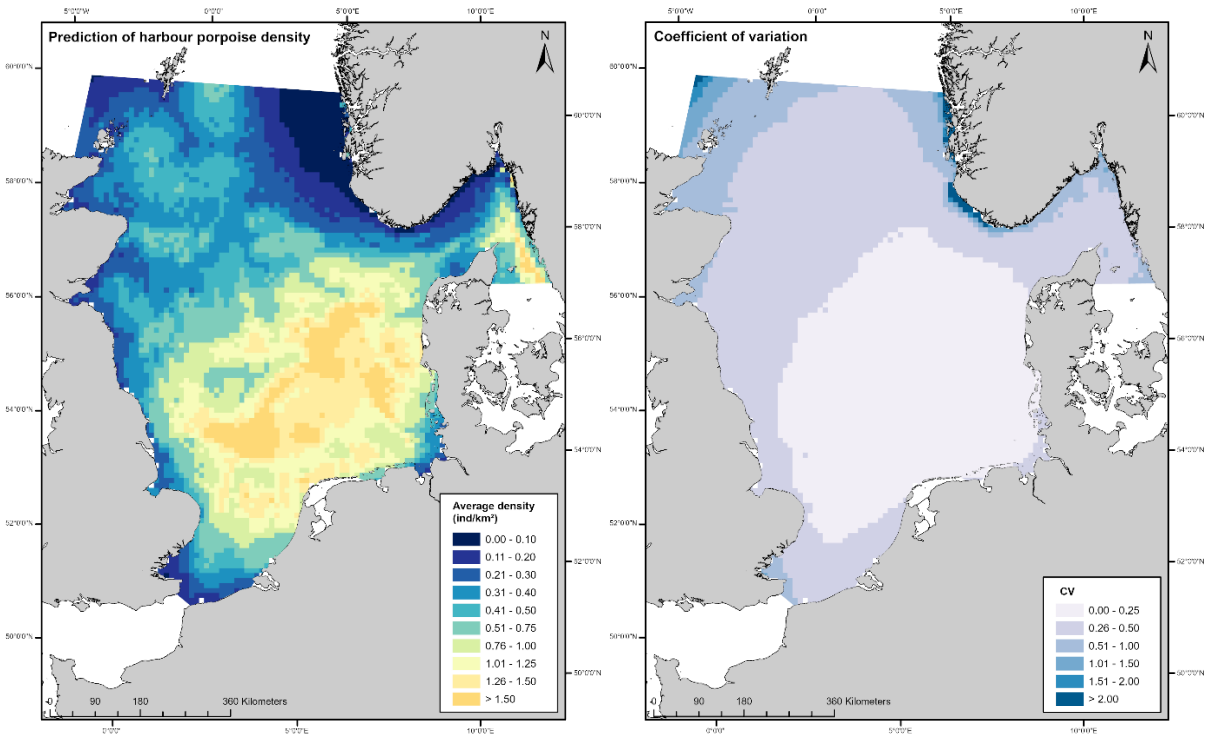

***Figure 2**. Predicted surface of estimated summer density [left] and associated coefficient of variation (CV) [right] for harbour porpoise, fitting the model to survey data collected between 2020-2024.*

Model diagnostic

The fitted smooth relationships between relative density and the spatial-autocorrelation smooths (main smooth and deviation by year), as well as each covariate selected in the final best model, are shown in **Figure 3**. The spatial autocorrelation te(X, Y) is explaining most of the variance in this final model, while other covariates contributed in an equivalent magnitude to the model. SST contributed the least, as shown by the smaller amplitude of its partial effect on the y-axis.

The Q-Q plot (**Figure A. 1**) shows that the deviance residuals between 0 and 2 fall below the confidence band while those above 2 falls above the confidence bands. This pattern suggests some model misfit, especially for large quantiles as residuals are far from the confidence bands.

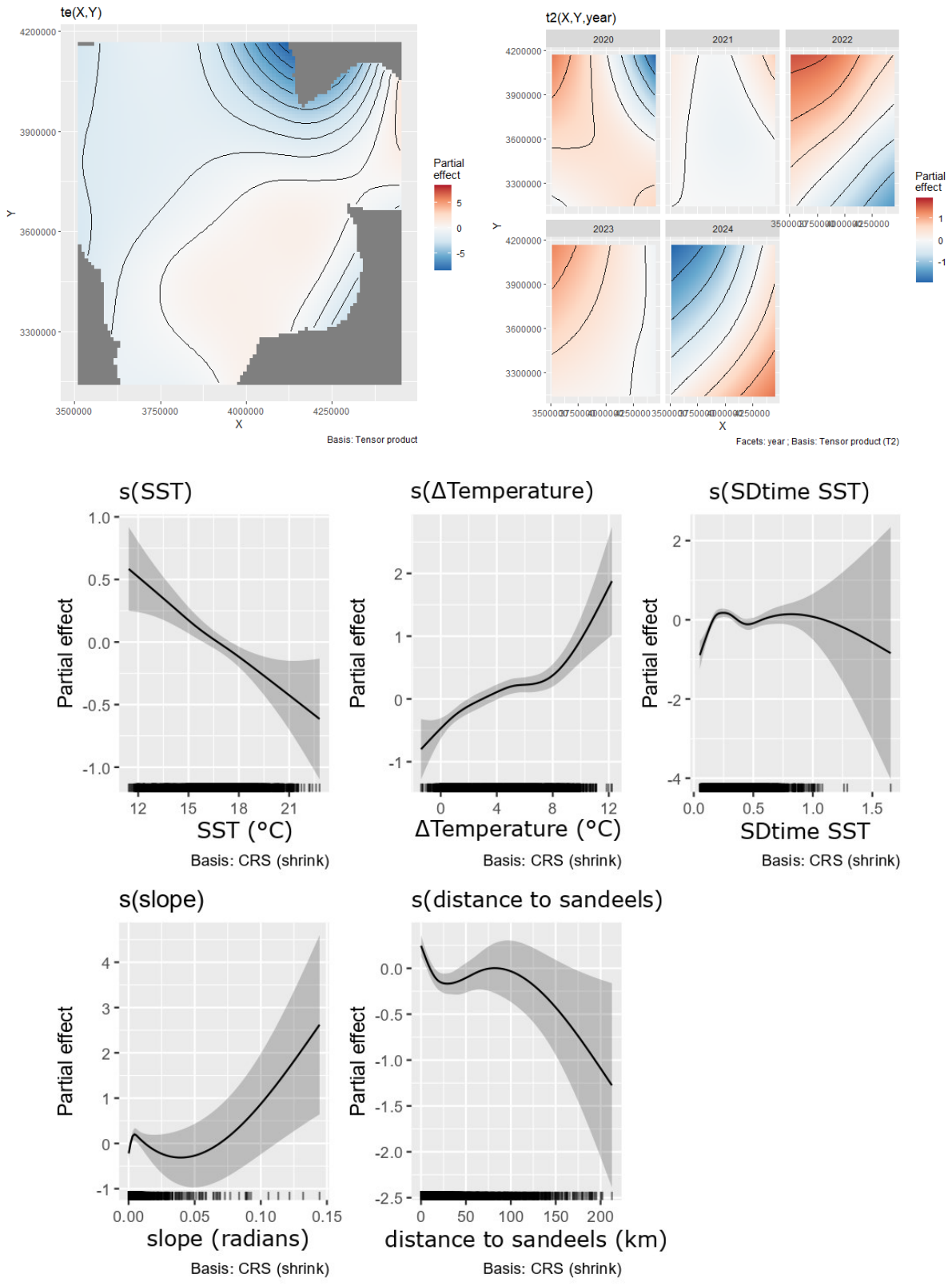

***Figure 3**. Functional plots of environmental variables relative to harbour porpoise density as indicated by the estimated smooth functions for the selected covariates in the best model (see **Table 3** for explanation on covariates). Y-axes represent function of the term (linear or spline). Zero on the y-axes corresponds to no effect of the predictor variable on the estimated response variable. Scaling of y-axis varies among predictor variables to emphasize model fit. The shading reflects 2× standard error bands (i.e., 95% confidence interval); tick marks ("rug plot") above the X-axis show data values.*

Gap analysis in the environmental space

A gap analysis in the environmental space was conducted, following the method of Pigeault et al. (2024b). Results shown in **Figure 4** and **Figure 5** were consistent with the CV displayed in **Figure 2**: predictions in shallow waters were particularly well supported by the calibration dataset, as their environments were similar to a large number of data collected in the North Sea and used in the model fitting. The predictions in the northern and most southern part (i.e. towards the Channel) of the study area were the least supported by the calibration dataset: their environments are more different from most of the data used during model fitting. However, no extrapolations were involved in the southern part of the study area. Few extrapolations were involved in the northern part of the study area, but it did not involve more than 5% to 10% of the days for most of the grid cells. Only the northern-most cells were extrapolated above 10% of the days and did not exceed 20% (except some special cells in the northern Norwegian coast slope, or northern Shetland Islands slope). SST was the covariate involving the larger number of extrapolations. These extrapolations were especially located in the north on the first predicted days of each year's survey period (i.e. beginning of June) when temperatures were colder than those captured by the calibration dataset but reversed in less than two weeks for most of the predictions.

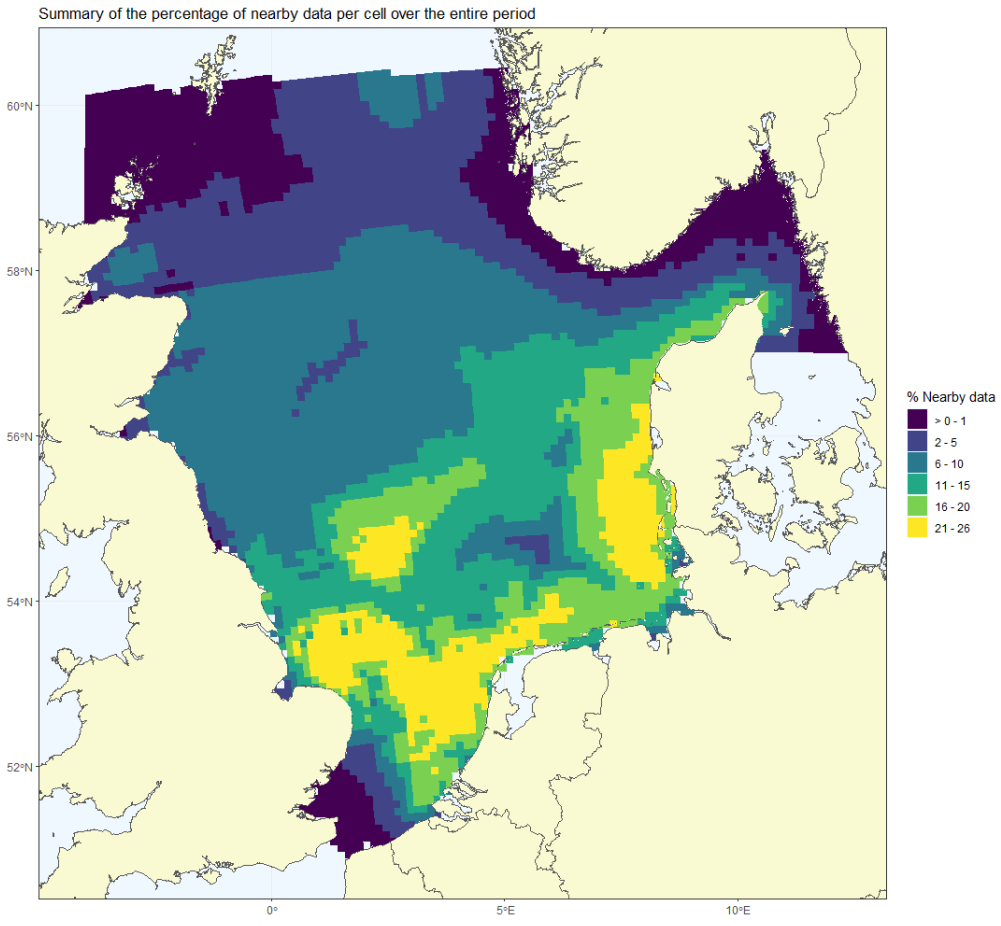

Figure 4. *Mean percentage of nearby data per prediction over the entire study period. Nearby data are calibration data that have similar environments to those of a prediction, thus allowing for a more reliable estimate of* ***harbour porpoise*** *density.*

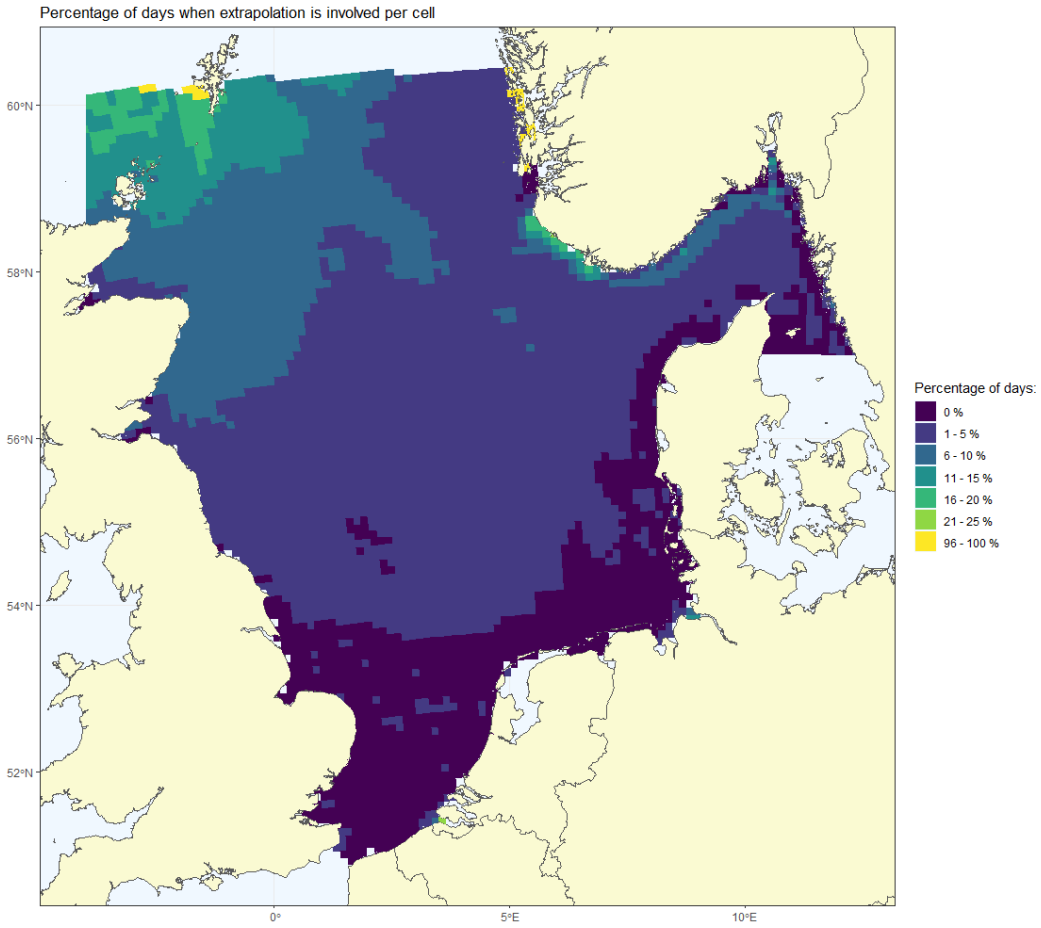

***Figure 5**. Percentage of days per prediction over the entire study period when harbour porpoise is environmentally extrapolated.*

Hot spot analysis

Figure 6 shows that the hotspots, as calculated with the top density percentages of the modelled output, are situated in the southern-eastern part of the North Sea as well as in the southern Skagerrak/northwestern Kattegat. The hotspots highlighted by the Gi* metric show an almost identical pattern to the top 15% hotspot when used on the modelled output (see **Figure 6** and **Figure 7A**), but a very different, fragmented pattern when used on the effort-corrected sightings (see **Figure 7B**).

Since hotspots have been shown to indicate critical habitats, such as areas used for reproduction, a further exploratory analysis was conducted. In this analysis, the locations of mother-calf pair sightings, as observed during the aerial surveys in the study period, were overlaid with the modelled density surface (**Figure A. 2**). In total, there were 444 calves sightings within the model area during the study period.

Table A. 1 summarises the numbers of sightings of calves inside each hotspot, as well as the percentage compared to the total of mother-calf pair sightings. Here, for this analysis on percentage contribution, we chose to only include the calf sightings inside the Danish or German EEZ due to the high difference in survey effort within the model area. As

indicated in **Figure 1**, due to the survey coverage from SCANS-IV, the southern and northern regions were adequately represented in 2022; however, the number of national surveys available in these regions was comparatively lower in other years.

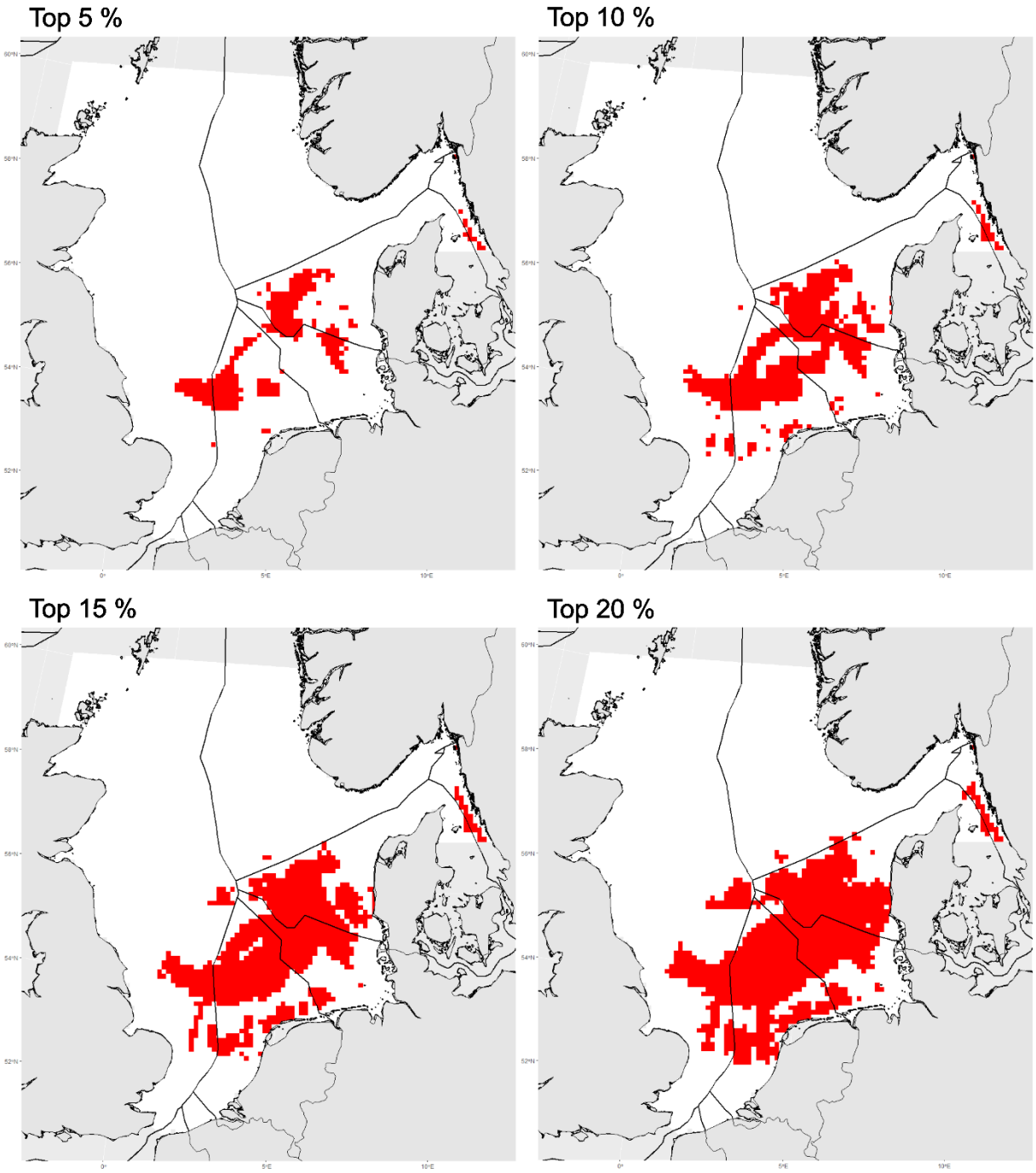

***Figure 6**. Top 5, 10, 15 and 20% of densities (red) as calculated from the output from the modelled density surface (see **Figure 2**), overlaid with the EEZ boundaries (black)*

Selecting the top percentages of modelled densities is a simple and straight-forward method to highlight hotspots of highest porpoise abundance. The Gi* metric in principle offers additional information as it also gives a measure of the statistical significance of the hotspots, through a measure of p-value.

P-values associated with Gi* can be calculated, but because the null hypothesis of spatial homogeneity has been rejected already by virtue of using a DSM (unless the DSM only includes an intercept which is highly unlikely), the associated p-values are not interpretable. In such a case, the Gi* simply reflects the highest densities predicted by the model and is, thus, no different than a hotspot based on highest percentages. Our results corroborated this, as the hotspots highlighted by the Gi* calculated on the modelled output is virtually the same as the top 15% of densities (see **Figure 6** and **Figure 7A**). The Gi* metric identifies hotspots based on strong local spatial autocorrelations. These local autocorrelations are predicted from the effects of several spatially explicit covariates and a residual smooth effect of geographical space (te(X, Y) term; see model structure). Although a DSM does not explicitly model local spatial autocorrelation, predicted densities from a DSM are expected to display a smooth pattern. It is thus unsurprising that hotpots and coldspots identified with the Gi* metric largely aligned with the contours of high and low predicted density area.

The Gi* can however be calculated directly on the sightings data, and the associated p-value may be interpreted in the usual way (i.e. assuming a null hypothesis of spatial homogeneity in sightings). **Figure 7B** shows that we obtained hotspots with a more fragmented pattern, that might be harder to interpret and use in policy making but contains nonetheless useful information. The fragmented pattern is also expected compared to the same analysis performed on the output of a DSM since the latter smooths out the data to a surface. It is also important to note that by using the sightings directly, even if correcting for effort, we lose a lot of the benefits of the modelling process (linking to environmental covariates, extrapolating outside the sampled areas, etc.).

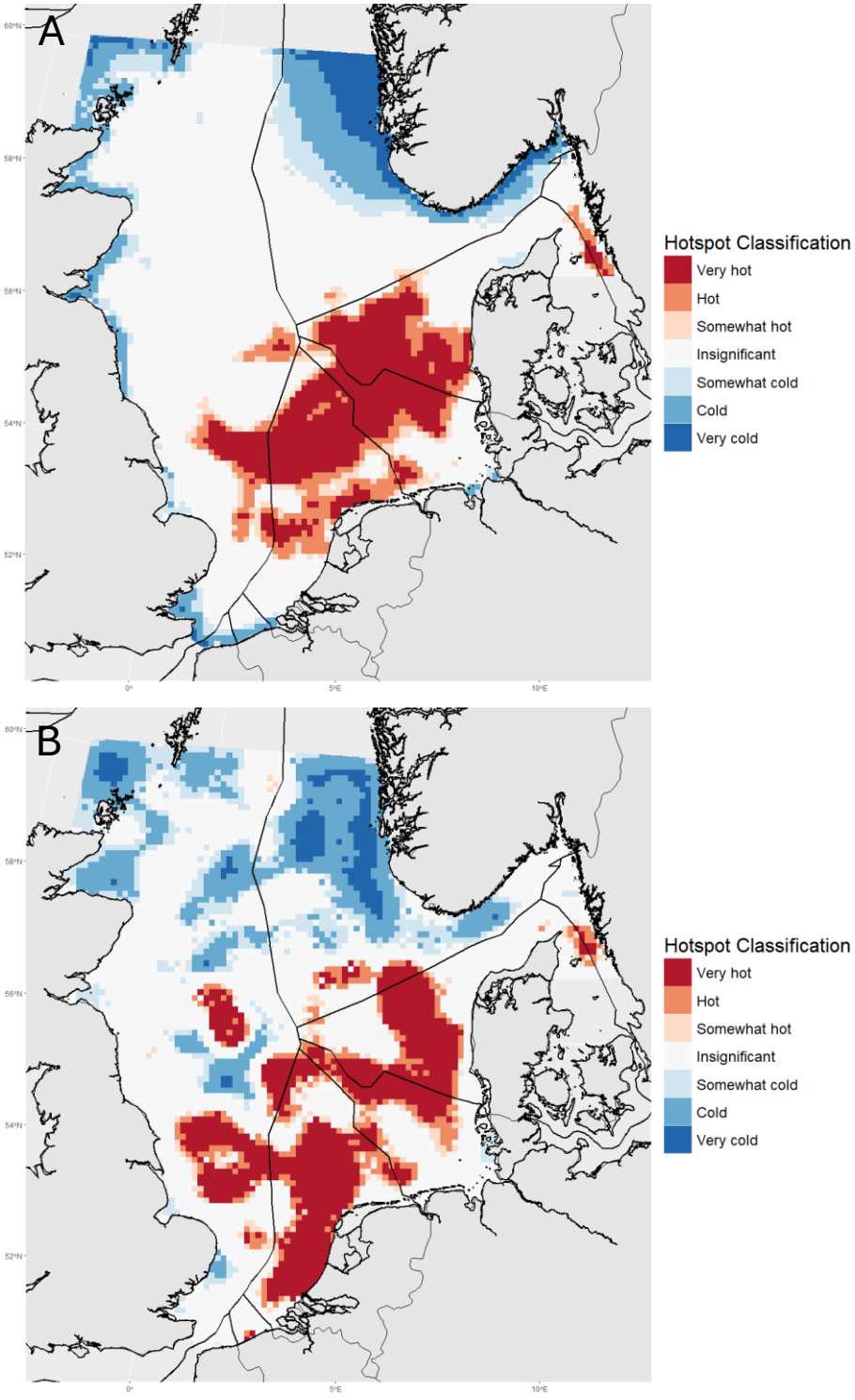

***Figure 7**. Hotspots and Coldspots identified with the Gi* metric calculated on A) the output from the modelled density surface (see **Figure 2**) and B) the observed sightings of harbour porpoise corrected by survey effort, overlaid with the EEZ boundaries (black)*

In conclusion, we can postulate that the predicted density surface already captured the hotspots very well, and a decision which quantile/percentile to use depends on the specific application for management and conservation.

REFERENCES

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Modell. 200: 1–19. doi:10.1016/j.ecolmodel.2006.07.005

Becker, E.A., Forney, K.A., Fiedler, P..C, Barlow, J., Chivers, S.J., Edwards, C.A., Moore, A.M., Redfern, J.V. (2016). Moving towards dynamic ocean management: how well do modeled ocean products predict species distributions? Remote Sens. 8: 149. doi: 10.3390/rs8020149

Becker, E.A., Forney, K.A, Thayre, B.J., Debich, A.J., Campbell, G.S., Whitaker, K., Douglas, A.B., Gilles, A., Hoopes, R., Hildebrand, J.A. (2017). Habitat-based density models for three cetacean species off Southern California illustrate pronounced seasonal differences. Front. Mar. Sci. 4: 121. doi: 10.3389/fmars.2017.00121

Becker, E.A., Forney, K.A., Redfern, J. V, Barlow, J., Jacox, M.G., Roberts, J.J., Palacios, D.M. (2018). Predicting cetacean abundance and distribution in a changing climate. Divers. Distrib. 25: 626–643. doi:10.1111/ddi.12867

Becker, E. A., Forney, K. A., Miller, D. L., Fiedler, P. C., Barlow, J., & Moore, J. E. (2020). Habitat-based density estimates for cetaceans in the California Current Ecosystem based on 1991-2018 survey data. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-638.

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., & Blanchet, G. (2017). Package 'spdep'. Spatial dependence: Weighting schemes, statistics, R package version, 1(5).

Bivand, R.S., Wong, D.W.S. (2018). Comparing implementations of global and local indicators of spatial association. TEST 27, 716–748. https://doi.org/10.1007/s11749-018-0599-x

Bouchet, P.J., Miller, D.L., Roberts, J.J., Mannocci, L., Harris, C.M., Thomas, L. (2020). dsmextra: Extrapolation assessment tools for density surface models. Methods Ecol. Evol. 11, 1464-1469, 10.1111/2041-210X.13469

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L. (Eds.), (2004). Advanced Distance Sampling: Estimating Abundance of Biological Populations. OUP Oxford.

EMODnet Digital Bathymetry (DTM 2020). EMODnet Bathymetry Consortium https://doi.org/10.12770/bb6a87dd-e579-4036-abe1-e649cea9881a

EMODnet Bathymetry World Coastline. EMODnet Bathymetry Consortium https://doi.org/10.12770/36821cff-56db-4e96-8426-ddfe68240c4b

Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J. V, Barlow, J., Vilchis, I.L., Ballance, L.T. (2012). Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. Endang. Species Res. 16: 113-133. doi:10.3354/esr00393

Geelhoed, S.C.V., Authier, M., Pigeault, R., Gilles, A. (2022). Abundance and distribution of cetaceans. In: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London. Available at: https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/indicator-assessments/abundance-distribution-cetaceans/

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Maechler, M., Hothorn, T. (2021). Package 'mvtnorm'. J. Comput. Graph. Stat. 11(950-971), 155.

Gilles, A., Scheidat, M., Siebert, U. (2009). Seasonal distribution of harbour porpoises and possible interference of offshore wind farms in the German North Sea. Mar. Ecol. Prog. Ser. 383, 295-307.

Gilles, A., Adler, S., Kaschner, K., Scheidat, M., Siebert, U. (2011). Modelling harbour porpoise seasonal density as a function of the German Bight environment: implications for management. Endang. Species Res. 14: 157–169

Gilles, A., Viquerat, S., Becker, E.A., Forney, K.A., Geelhoed, S.C.V., Haelters, J., Nabe- Nielsen, J., Scheidat, M., Siebert, U., Sveegaard, S., van Beest, F.M., van Bemmelen, R., Aarts, G. (2016). Seasonal habitat-based density models for a marine top predator, the harbor porpoise, in a dynamic environment. Ecosphere 7 (6), e01367. https://doi.org/10.1002/ecs2.1367

Gilles, A., Authier, M., Ramirez-Martinez, N.C., Araújo, H., Blanchard, A., Carlström, J., Eira, C., Dorémus, G., Fernández-Maldonado, C., Geelhoed, S.C.V., Kyhn, L., Laran, S., Nachtsheim, D., Panigada, S., Pigeault, R., Sequeira, M., Sveegaard, S., Taylor, N.L., Owen, K., Saavedra, C., Vázquez-Bonales, J.A., Unger, B., Hammond, P.S. (2023). Estimates of cetacean abundance in European Atlantic waters in summer 2022 from the SCANS-IV aerial and shipboard surveys. Final report published 29 September 2023. 64 pp. https://www.tiho-hannover.de/itaw/scans-iv-survey

Gilles, A., Authier, M., Pigeault, R., Ramirez-Martinez, N.C., Benoit, V., Carlström, J., Eira, C., Geelhoed, S.C.V., Laran, S., Sequeira, M., Sveegaard, S., Taylor, N.L., Saavedra, C., Vázquez-Bonales, J.A., Hammond, P.S. (2025). Spatial models of cetacean density in European Atlantic waters based on SCANS-IV summer 2022 survey data. Final report published 14 May 2025. 31 pp plus Appendix. https://tinyurl.com/3rv246v5

Global Ocean Biogeochemistry Hindcast. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). DOI: 10.48670/moi-00019 (Accessed on 26-05-2025)

Global Ocean Physics Reanalysis. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). DOI: 10.48670/moi-00021 (Accessed on 31-03-2025)

Hammond, P.S., Macleod, K., Berggren, P., Borchers, D.L., Burt, L., Cañadas, A., Desportes, G., Donovan, G.P., Gilles, A., Gillespie, D., Gordon, J., Hiby, L., Kuklik, I., Leaper, R., Lehnert, K., Leopold, M., Lovell, P., Øien, N., Paxton, C.G.M., Vázquez, J. A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. Biol. Conserv. 164, 107-122. https://doi.org/10.1016/j.biocon.2013.04.010

Hammond, P.S., Lacey, C., Gilles, A., Viquerat, S., Börjesson, P., Herr, H., Macleod, K., Ridoux, V., Santos, M.B., Scheidat, M., Teilmann, J., Vingada, J., Øien, N. (2021). Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. SCANS-III project report 1, 39 pp. https://scans3.wp.st-andrews.ac.uk/files/2021/06/SCANS-III_design-based_estimates_final_report_revised_June_2021.pdf

Hastie, T.J., Tibshirani, R.J., (1990). Generalized additive models. Chapman & Hall, London.

Hijmans, R. J., Van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E.B., Shortridge, A. (2013). Raster package in R. Version https://mirro rs sjtug sjtu edu cn/cran/web/packa ges/rast er/raster pdf

Jensen, H., Rindorf, A., Wright, P. J., Mosegaard, H. (2011). Inferring the location and scale of mixing between habitat areas of lesser sandeel through information from the fishery. ICES J. Mar. Sci. 68(1), 43-51.

King, G., Zeng, L. (2007). When can history be our guide? The pitfalls of counterfactual inference Int. Stud. Q 51, 183-210, 10.1111/j.1468-2478.2007.00445.x

King, G, Tomz, M, Wittenberg, J (2000). Making the Most of Statistical Analyses: Improving Interpretation and Presentation. Am. J. Pol. Sci. 44(2), 347–361. https://doi.org/10.2307/2669316

Lau-Medrano, W (2024). grec: Gradient-Based Recognition of Spatial Patterns in Environmental Data. R package version 1.6.0, https://CRAN.R-project.org/package=grec.

Lacey, C., Gilles, A., Börjesson, P., Herr, H., Macleod, K., Ridoux, V., Santos, M.B., Scheidat, M., Teilmann, J., Sveegaard, S., Vingada, J., Viquerat, S., Øien, N., Hammond, P.S. (2022). Modelled density surfaces of cetaceans in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. 55 pp. Available from: https://scans3.wp.st-andrews.ac.uk/files/2022/08/SCANS-III_density_surface_modelling_report_final_20220815.pdf

Marra, G, Wood, SN (2011). Practical variable selection for generalized additive models. Comput. Stat. Data Anal. 55, 2372-2387. https://doi.org/10.1016/j.csda.2011.02.004

Parvianen, M., Marmion, M., Luoto, M., Thuiller, W., Heikkinen, R.L. (2009). Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. Biol. Conserv., 142. https://doi.org/10.1016/j.biocon.2009.05.030

Pedersen, E.J., Miller, D.L., Simpson, G.L., Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 7:e6876 https://doi.org/10.7717/peerj.6876

Pigeault, R., Ruser, A., Ramírez-Martínez, N. C., Geelhoed, S. C., Haelters, J., Nachtsheim, D. A., Schaffeld, T., Sveegaard, S., Siebert, U., Gilles, A. (2024a). Maritime traffic alters distribution of the harbour porpoise in the North Sea. Mar. Pollu. Bull. 208, 116925. https://doi.org/10.1016/j.marpolbul.2024.116925

Pigeault, R., Authier, M., Ramirez-Martinez, N.C., Virgili, A., Geelhoed, S.C.V., Haelters, J., Louzao, M., Saaveedra, C., Gilles, A. (2024b). Assessing the reliability of species distribution models under changing environments: A case study on cetaceans in the North-East Atlantic. Glob. Ecol. Conserv. 56, e03299. doi:10.1016/j.gecco.2024.e03299

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Found. Stat. Comput., Vienna, Austria. https://www.R-project.org/

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L., Werner, F. (2006). Techniques for cetacean-habitat modeling. Mar. Ecol. Prog. Ser. 310, 271–295. https://doi.org/10.3354/meps310271

SAMMOA (2022) Software dedicated for aerial survey of marine megafauna. Observatoire Pelagis LRUniv-CNRS - Code Lutin, version 2.1.2. and revised version 2.1.3 (https://www.observatoire-pelagis.cnrs.fr/tools/sammoa)

Scheidat, M., Gilles, A., Kock, K.H., Siebert, U. (2008). Harbour porpoise Phocoena phocoena abundance in the southwestern Baltic Sea. Endang. Species Res. 5:215-223 https://doi.org/10.3354/esr00161

Tolimieri, N., Shelton, A.O., Feist, B.E., Simon, V. (2015). Can we increase our confidence about the locations of biodiversity 'hotspots' by using multiple diversity indices? Ecosphere 6:1–13. https://doi.org/10.1890/ES14-00363.1

Vehtari, A, Gelman, A, Gabry, J (2017). Practical Bayesian model evaluation using leave-one-out crossvalidation and WAIC. Stat. Comput. 27, 1413-1432. https://doi.org/10.1007/s11222-016- 9696-4

Virgili, A., Authier, M., Boisseau, O., Cañadas, A., Claridge, D., Cole, T., Corkeron, P., Dorémus, G., David, L., Di-Mégli, N., Dunn, C., Dunn, T.E., García-Barón, I., Laran, S., Lauriano, G., Lewis, M., Louzao, M., Manocci, L., Martínez-Cedeira, J., Palka, D., Panigada, S., Pettex, E., Roberts, J.J., Ruiz, L., Saavedra, C., Santos, M.B., Van Canneyt, O., Vázquez Bonales, J.A., Monestiez, P., Ridoux, V. (2019). Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale: Large-scale modelling of deep-diving cetacean habitats. Global Ecol. Biogeogr. 28(3), 300-314.

Wood, S.N. (2017). Generalized additive models - an introduction with R. Chapman and Hall/CRC, 496 pp.

Yao, Y, Vehtari, A, Simpson, D, Gelman, A (2017). Using stacking to average Bayesian predictive distributions. Bayesian Anal. 13(3), 917-1007. https://doi.org/10.1214/17-BA1091

APPENDIX

Model diagnostic

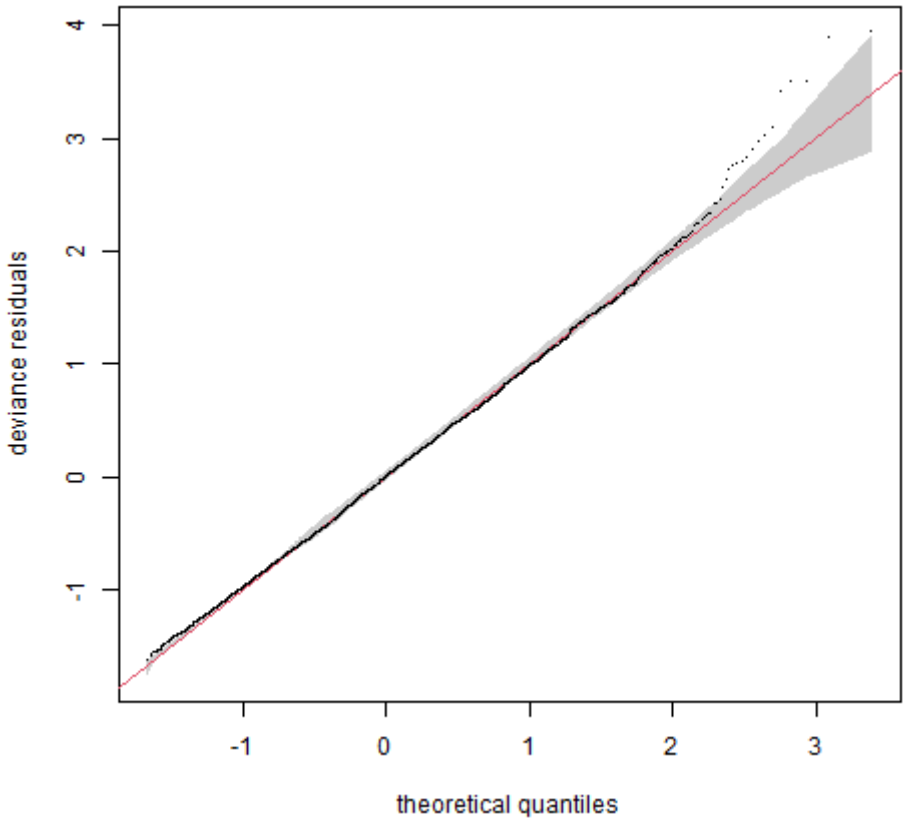

Figure A. 1. *Q-Q plot for the selected final best model*

Hotspot analysis

Table A. 1. *Number and percentage of* ***harbour porpoise*** *calves as overlayed with each hotspot extracted from modelled density surface (see* ***Figure A. 2****). Note: only hotspot extent and mother-calf pair sightings within the Danish or German EEZ (n=333) in the period 2020-2024 were included.*

Hotspot	Number of calves sightings inside hotspot	Percentage of all calves sightings
top 5%	63	19%
top 10%	112	34%
top 15%	163	49%
top 20%	212	64%

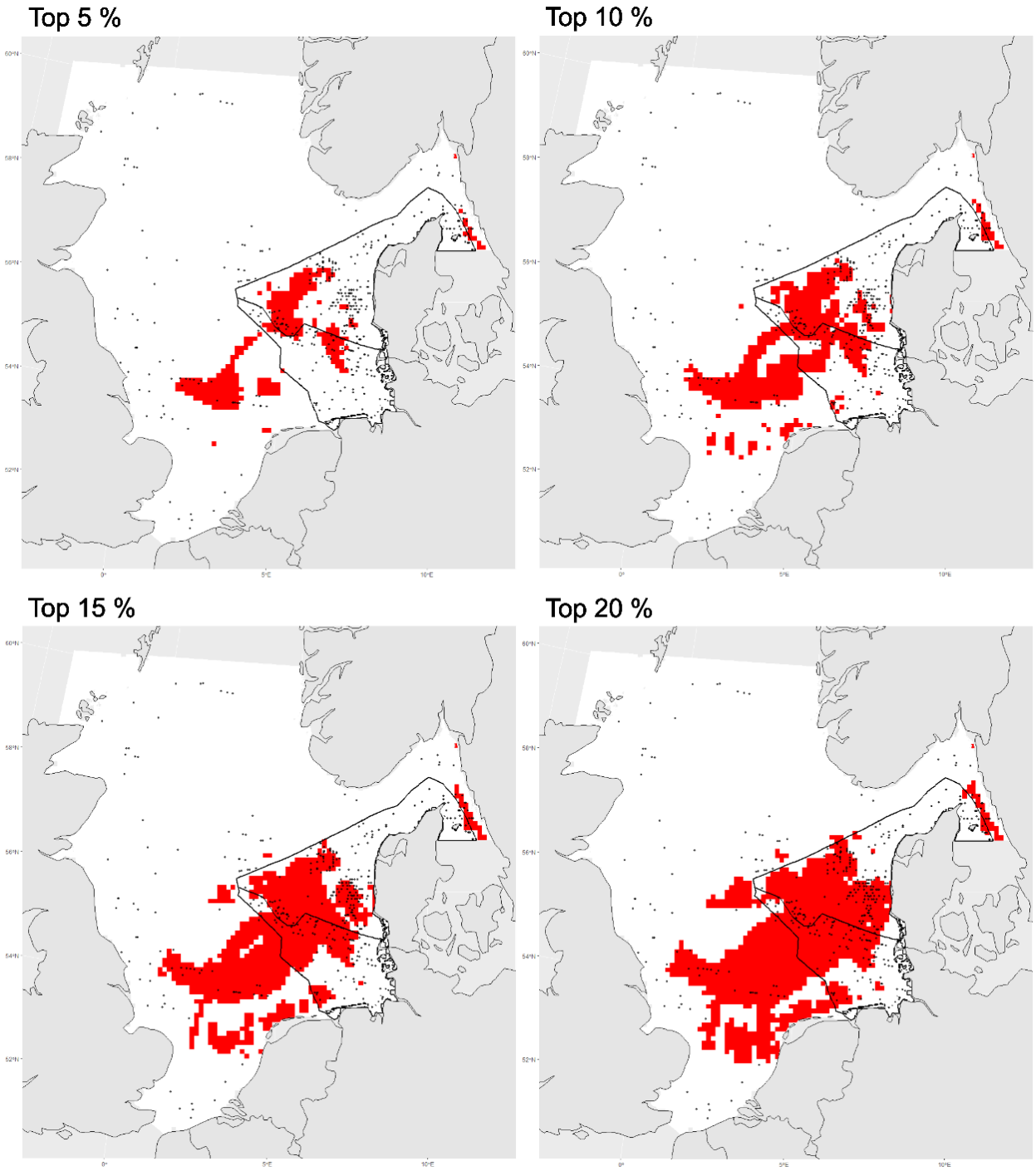

Figure A. 2. *Top 5, 10, 15 and 20% of densities (red) as calculated from the output from the modelled density surface (see* ***Figure 2****), overlayed with observed sightings of* ***harbour porpoise*** *calves (black point) and the EEZ of Germany and Denmark respectively (black line).*

Only sightings of mother-calf pairs inside the Danish or German EEZ (333 total) were used for **Table A. 1**. Both **Figure A. 2** and **Table A. 1** show that the predicted hotspots are overall consistent with the sightings of calves. The percentages of sightings inside the hotspots were always high relative to the size of the hotspot, highlighting that the hotspots are not only areas of high densities, but also of ecological importance during the reproduction period in summer.