### *Model Diagnostics*

*Whale Sighting Occurrence Model Diagnostics*

Model diagnostic plots for the best sighting occurrence model are given in Fig. S1. QQ plot in Fig. S1a does not show any sign of over-dispersion. The other diagnostic plots are however not very informative due to the binary nature of the response variable. The best occurrence model was interpreted without incorporating any autoregressive structure (AR(1) or ARIMA) given autocorrelation was low in strength as it was only found in the 1$^{st}$, 5$^{th}$, 9$^{th}$ and 11$^{th}$ lags (i.e., $\varphi_1$, $\varphi_5$, $\varphi_9$ and $\varphi_{11}$ < 0.05, Fig. S2), and general additive mixed models (GAMMs, Chen, 2000) are reported to perform poorly with binary data (Wood, 2016). Nevertheless, it should be noted that standard error (SE), confidence interval (CI) and coefficient of variation (CV) quantifying the uncertainty in covariate effects were likely to have been somewhat underestimated with without incorporating any autoregressive structure.
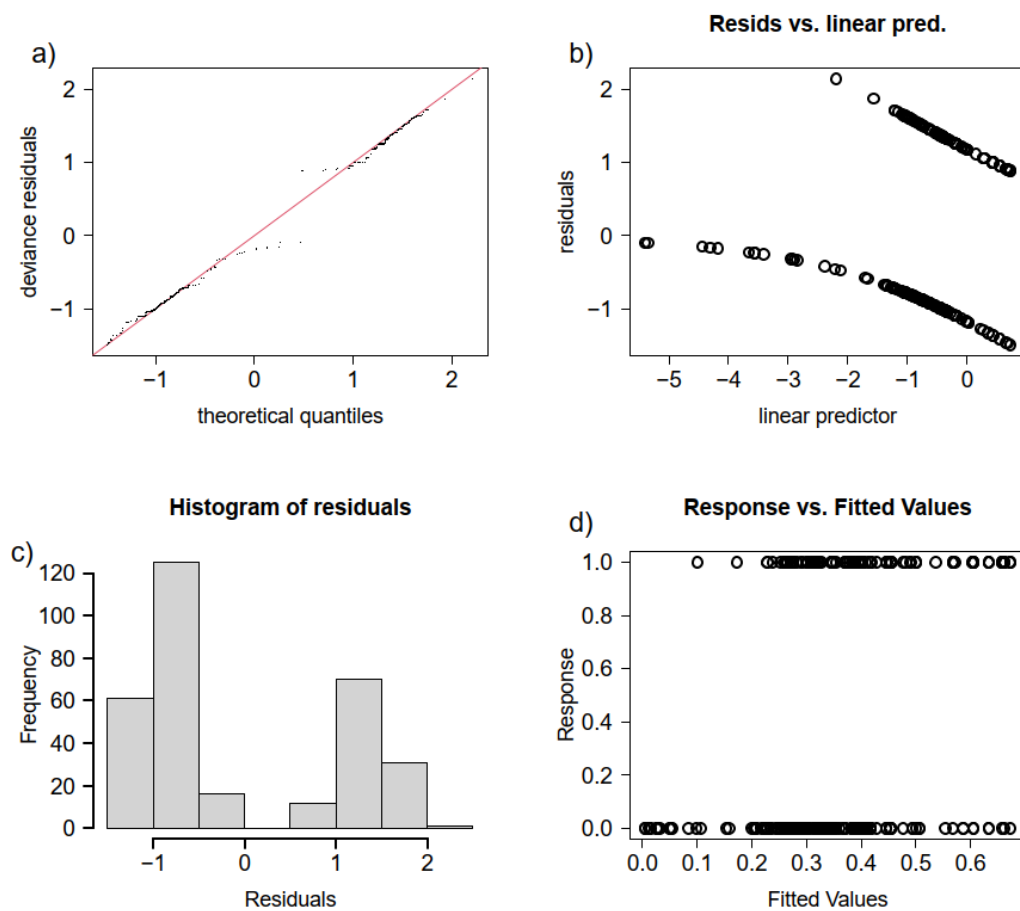


**Fig. S1 | Model diagnostic plots for the best sighting occurrence model. No sign of data over-dispersion was identified.**
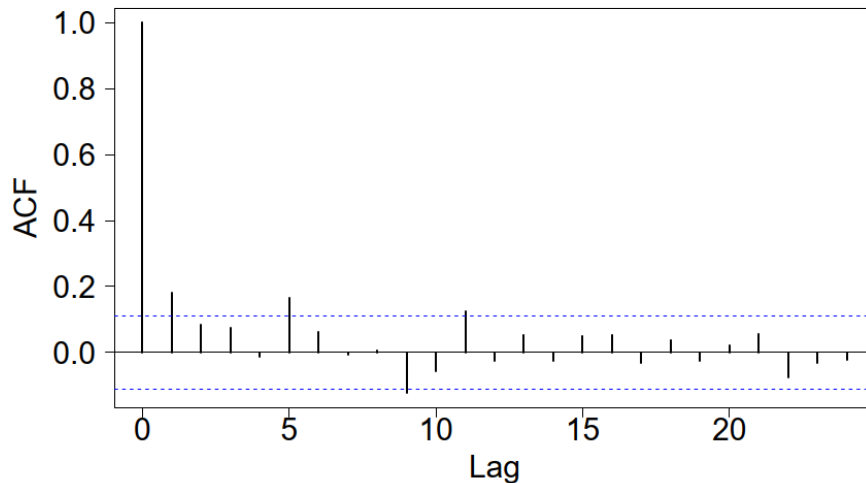
**Fig. S2| ACF plot for the best sighting occurrence model. Autocorrelated errors were found in the 1st, 5th, 9th and 11th lags, which are illustrated by the 2nd, 6th, 10th and 12th bars exceeding the ACF threshold values (horizontal dot lines).**

*Spatial Predictions of Sighting Occurrence*

Spatial estimates of whale sighting occurrence were based on model-averaged predictions, plotted in Fig. S3 together with the associated values of coefficient of variation (CV). Within the surveyed area, locations with high predicted occurrence estimates matched the spatial distribution of whale sightings, given sightings were mainly recorded along the submarine canyon to the north and southeast of Jan Mayen Island (see Fig. 1 of main text). The models estimated a high occurrence probability band running from 71.8° N and 4°W, to 68.5°N and 8°W, and bottlenose whales were also more likely to be found in northern waters off Jan Mayen Island and northwest waters (Greenland Sea) of the prediction area. Prediction uncertainty of these areas indicated by smaller CV values was lower. On the contrary, areas to the south of Jan Mayen as well as the southeast waters (Norwegian Sea) of the prediction area were estimated to have lower bottlenose whale occurrence probability, where the probability in general was lower than 0.2.
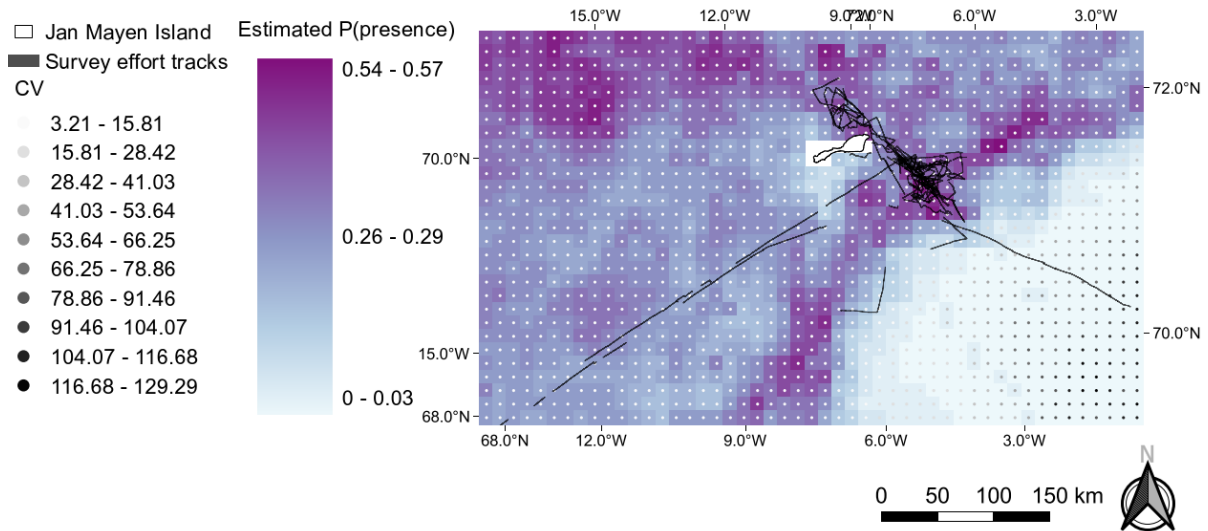
**Fig. S3 | Spatial estimates of whale sighting occurrence illustrated by the colour of grid cells. The associated coefficient of variation (CV) is represented by centroid point with higher CV values indicated by darker dot colours.**

*Additional Whale Sighting Model Diagnostics*

Residual distribution for the best Poisson model appeared to be slightly right-skewed (Fig. S4), yet the model was assumed to be robust to slight residual deviations from normality. As serial correlation was only found in the 18$^{th}$ lag (i.e., $\varphi_{18} < 0.05$, Fig. S5), while singular convergence error was identified in Poisson GAMM with autoregressive structure, the best Poisson model was thus interpreted without any autoregressive structure. Standard errors, confidence intervals and coefficients of variation were also likely to have been slightly underestimated under serial correlation.
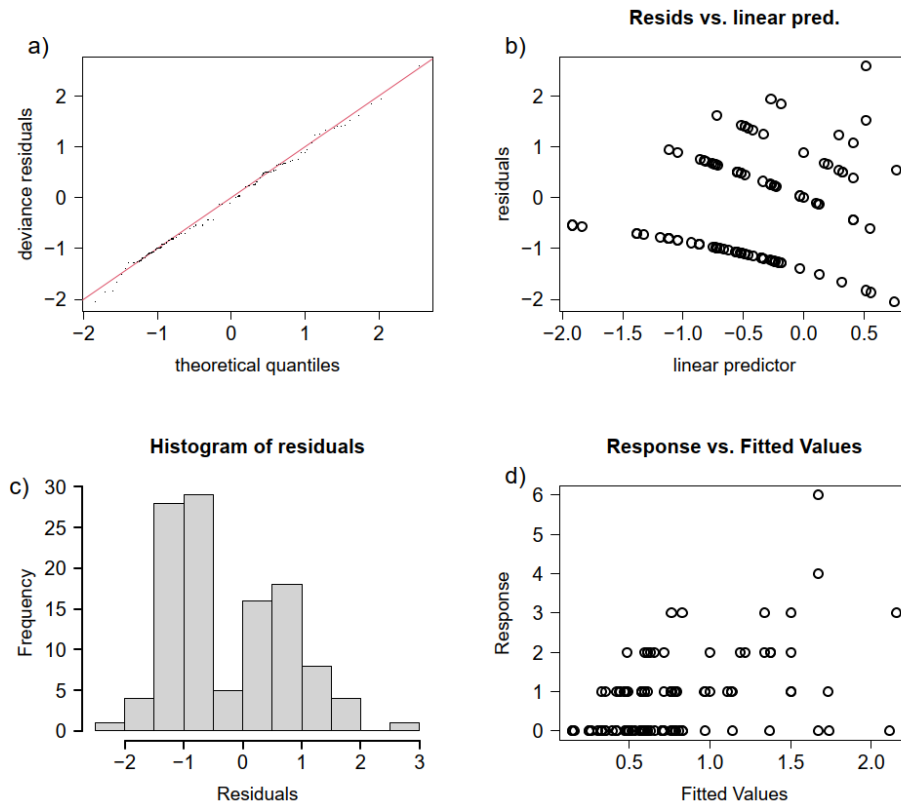
**Fig. S4 | Model diagnostic plots for the best Poisson model. Residual distribution appeared to be slightly right-skewed, yet the model was assumed to be robust to slight residual deviations from normality.**
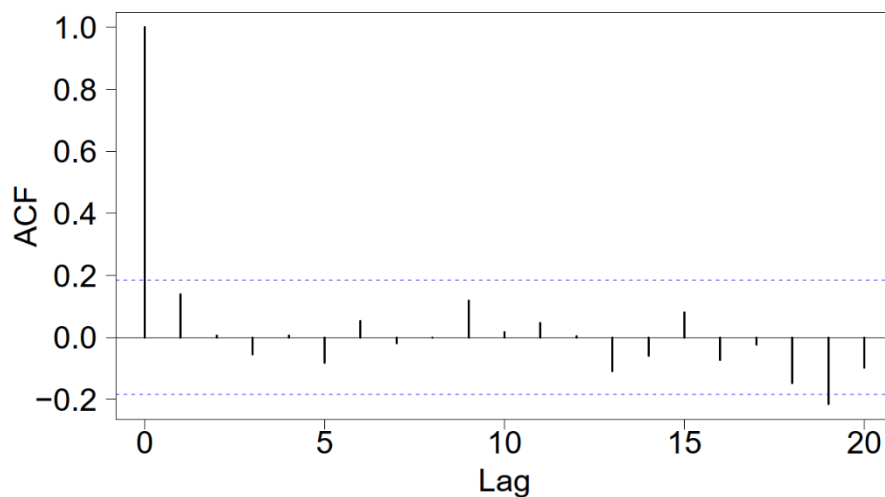


**Fig. S5 | ACF plot for the best Poisson model. Autocorrelated error was found in the 18th lag, which is illustrated by the 19th bar exceeding the ACF threshold values (horizontal dot lines).**

*Spatial Estimates of Number of Additional Whale Sightings Given First Encounter*

Spatial estimates of the number of additional whale sightings given first encounter (total number of whale sightings -1) were based on predictions averaged over the confidence set of models, plotted in Fig. S6 together with the associated CV. The number of additional whale sightings given first encounter was estimated to be higher on the northeast of study area (from 72° N and 1.5° W, to 70.8° N and 6° W). More additional whale sightings were also predicted in the area between 9.5°W and 11°W. On the contrary, fewer additional whale sightings were predicted on the west (Greenlandic Sea) of the prediction area, as well as waters between 7° W and 8° W. Prediction uncertainty was higher on the northwest and southeast of the wider prediction area.
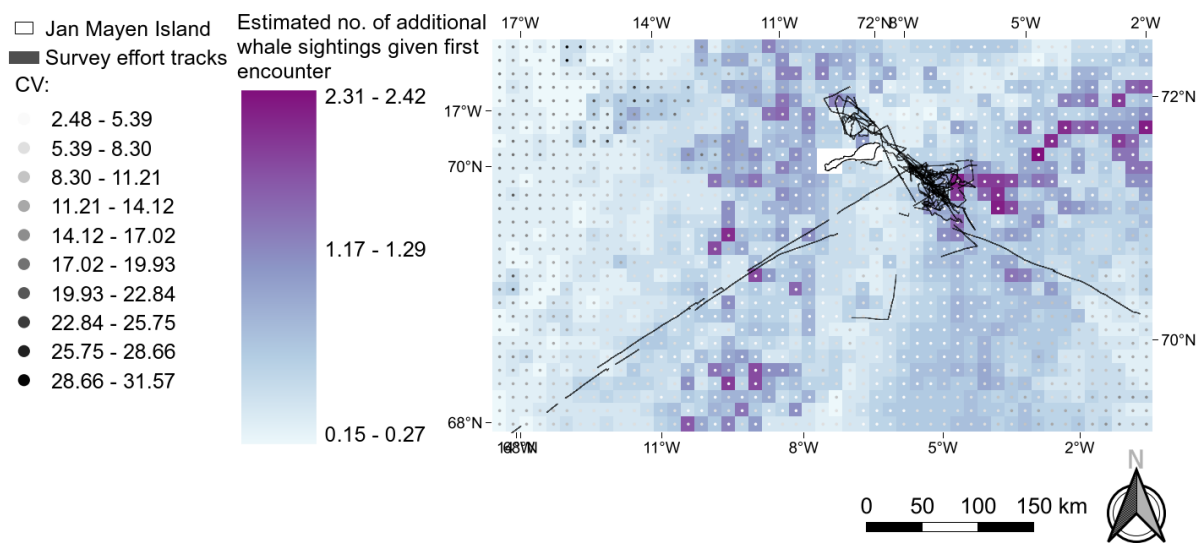


**Fig. S6 | Spatial estimates of the number of additional whale sightings given first encounter illustrated by the colour of grid cells. The associated coefficient of variation (CV) is represented by centroid point with higher CV values indicated by darker dot colours.**

## Literature Cited

Chen C (2000) Generalized additive mixed models. Communications in Statistics - Theory and Methods 29(5–6):1257–1271

Wood S (2016) Package 'mgcv.' cran.r-project.org/web/packages/mgcv/mgcv.pdf (accessed 25 Jan 2017)