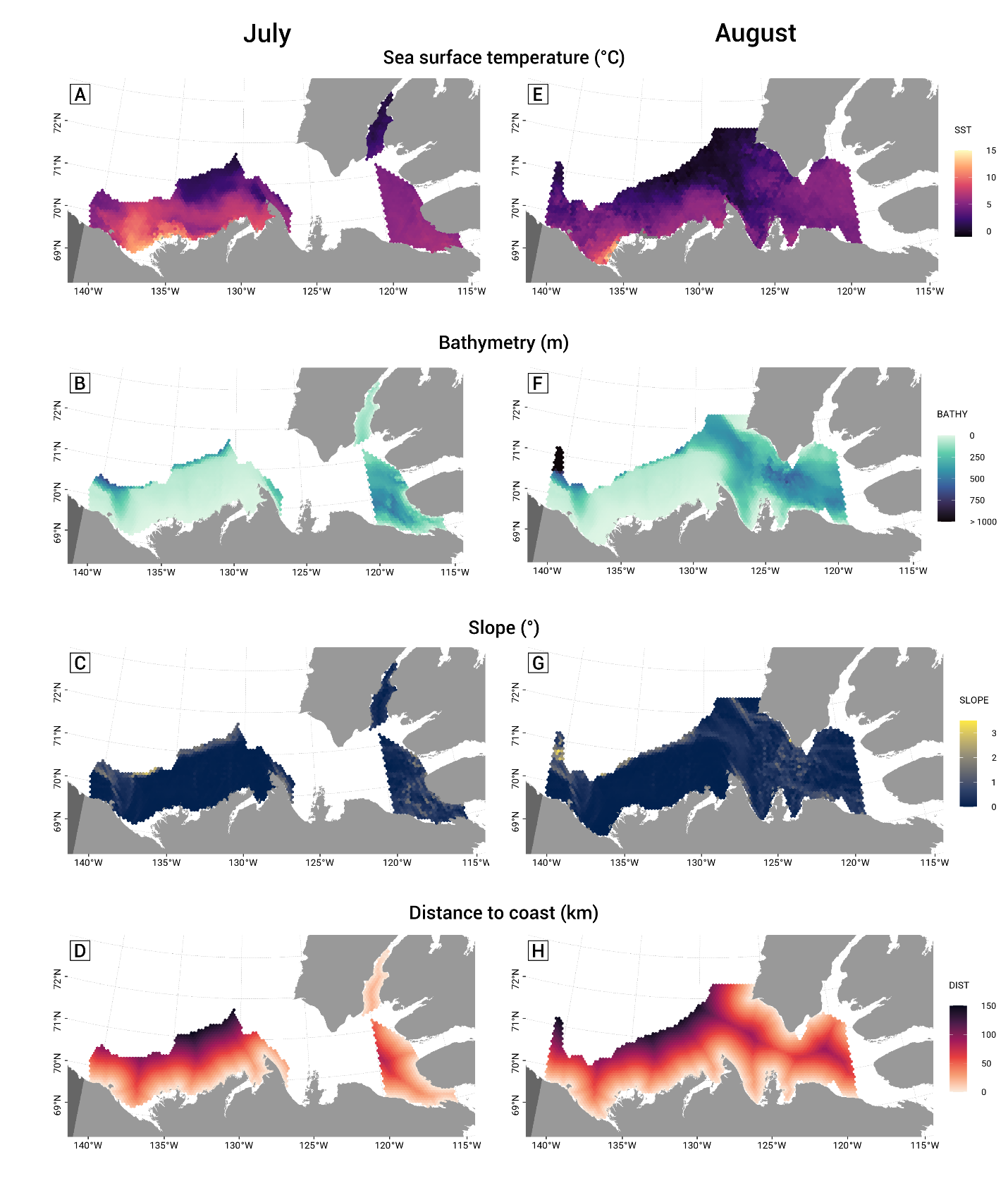
## Supplementary Material

### Environmental data of July and August 2019

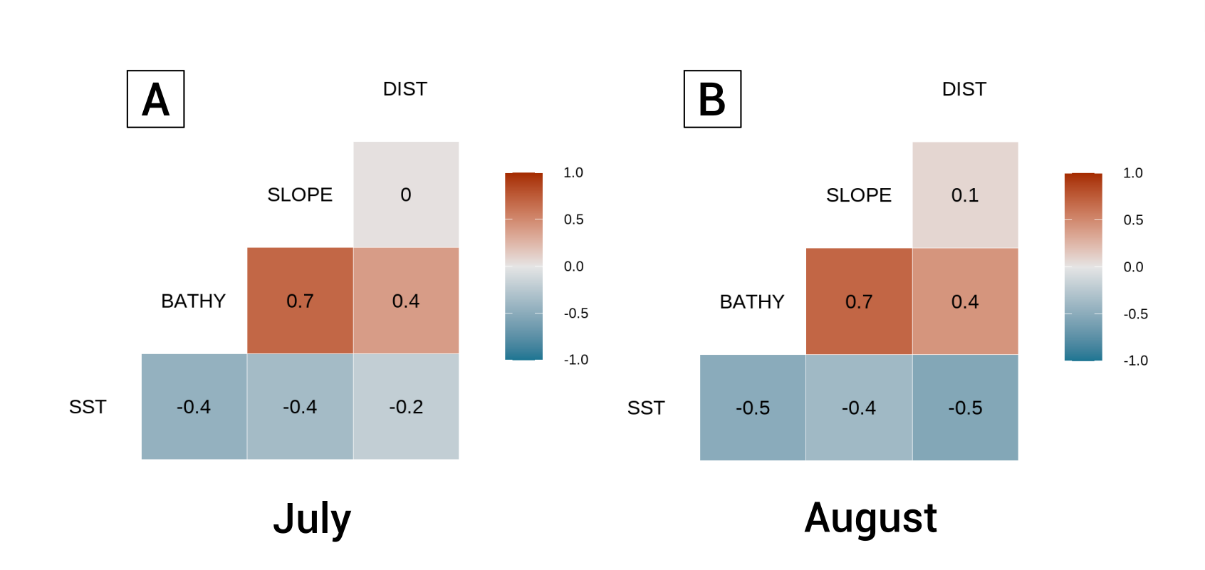


**Figure S1** Environmental covariates used in the models. Left: July surveyed area /Right: August surveyed area. A) and E) The sea surface temperature is the average temperature of the two 8-day periods of the July survey (20 July to 4 August 2019). and the three 8-day periods of the August survey (5 to 28 August 2019); B) and F) Bathymetry (m); C) and E) Slope (degrees); D) and H) Distances to the coast (km)

### Projection details

Equidistant conic projection used: "+proj=eqdc +lat\_1=69.42d +lat\_2=71.48d +lat\_0=69.75d +lon\_0=-129.2d +x\_0=0 +y\_0=0"

### Correlation matrix



**Figure S2** Spearman correlation matrix with coefficients between the four environmental variables for July and August. Matrix created with the “GGally” R package v.2.1.2.

### Details of the models

**Table S1** Hierarchical Generalized Additive Model parameters used with “mgcv” package (Wood, 2021).

|  |  |
| --- | --- |
| Argument | Explanation |
| s(), te(), t2() | Identify the terms to be smoothed. s() is used for one-dimensional term while te() is used for interaction between covariates (multi-dimensional) that are on different scales (Pedersen et al., 2019; McCabe et al., 2021)\*. Pedersen et al. (2019) recommend t2(full = TRUE) for multi-dimensional global smoother with random effect that have a shared penalty (model GS in their paper which our model was based on). |
| bs | Three types of smoothers are used in this model (Wood, 2017; Pedersen et al., 2019). The factor-smoother interaction fs is a type of smoother that allow separate set of basis functions for each group but estimate one smoothing parameter for all of them together (Pedersen et al., 2019). The thin plate regression splines (TPRS or tp) type is used for general continuous covariates and penalize functions that are too wiggly. The random effects (re) type penalizes the functions that are too far from the average. |
| k | k is the maximum effective degree of freedom (EDF, which represent the complexity of a penalized smooth term) or maximum number of basis functions. Our choice of the value k=10 is a trade-off between k being large enough to capture the variation in the response curve but not overfit (Pedersen et al., 2019; McCabe et al., 2021)\*. For the random effect smoother, k is equal to the number of levels of the factor variable (Pedersen et al., 2019), so 3 in our case. |
| m | The argument m is a penalty order. m=2 penalize on the squared derivative which means the more wiggly the functions are, the more penalized (models that overfit have large derivatives) (Pedersen et al., 2019). m=1 is recommended on the group-level smoothers to reduce issues of collinearity (concurvity) between the global smoother and group-level smoother (Pedersen et al., 2019; McCabe et al., 2021)\*. |
| family | We used binomial as our explanatory variable (BELUGA.P) is binary (absence – 0 or presence – 1) and to estimate the relative probability. |
| link | The complementary log-log or cloglog link is used when the number of 0s and 1s are unequal (Zuur et al., 2009). |
| method | The models were fit using the restricted maximum likelihood, REML smoothing parameter estimation. The use of REML is often preferred over the generalized cross-validation (GCV) (Wood, 2011; Wood, 2017). The score is the negative of the restricted log likelihood, and so the lower the value, the best (Woods, 2017). |

\*See McCabe et al. 2021 – Supplementary material

**Equation 1 – univariate model**

mod1 <- gam(y ~

s(X1, bs = "tp", k = 10, m = 2) +

s(X1, FACTOR, bs = "fs", k = 10, m = 2) +

s(X2, bs = "tp", k = 10, m = 2) +

s(X2, FACTOR, bs = "fs", k = 10, m = 2) + …,

data = df, family = binomial(link = "cloglog"), method = "REML")

Global smoother for a covariate (function response of beluga)

s(X1, bs = "tp")

Smooth term for random effect with a shared penalty (penalty for being too far from the average of X1)

s(X1, FACTOR, bs = "fs")

**Equation 2 – bivariate model**

mod2 <- gam(y ~

te(X1, X2, bs = "tp", k = 10, m = 2) +

t2(X1, X2, FACTOR, bs = c("tp", "tp", "re"), k = c(10, 10, 3), m = 2, full = TRUE),

data = df, family = binomial(link = "cloglog"), method = "REML")

Global smoother for interaction between covariates

te(X1, X2, bs = "tp")

Smoother for group-level with shared penalty and interaction between covariates

t2(X1, X2, FACTOR, bs = c("tp", "tp", "re"), full = TRUE)

### Model GS with environmental variables

#### July

B

A

Chart, line chart

Description automatically generatedChart

Description automatically generated

**Figure S3** Response curve of the best model for July with sea surface temperature and slope. Values are scaled and centred.

Chart, line chart

Description automatically generatedChart

Description automatically generated

B

A

**Figure S4** Response curve of the second best model for July with sea surface temperature and bathymetry. Values are scaled and centred.

#### August

B

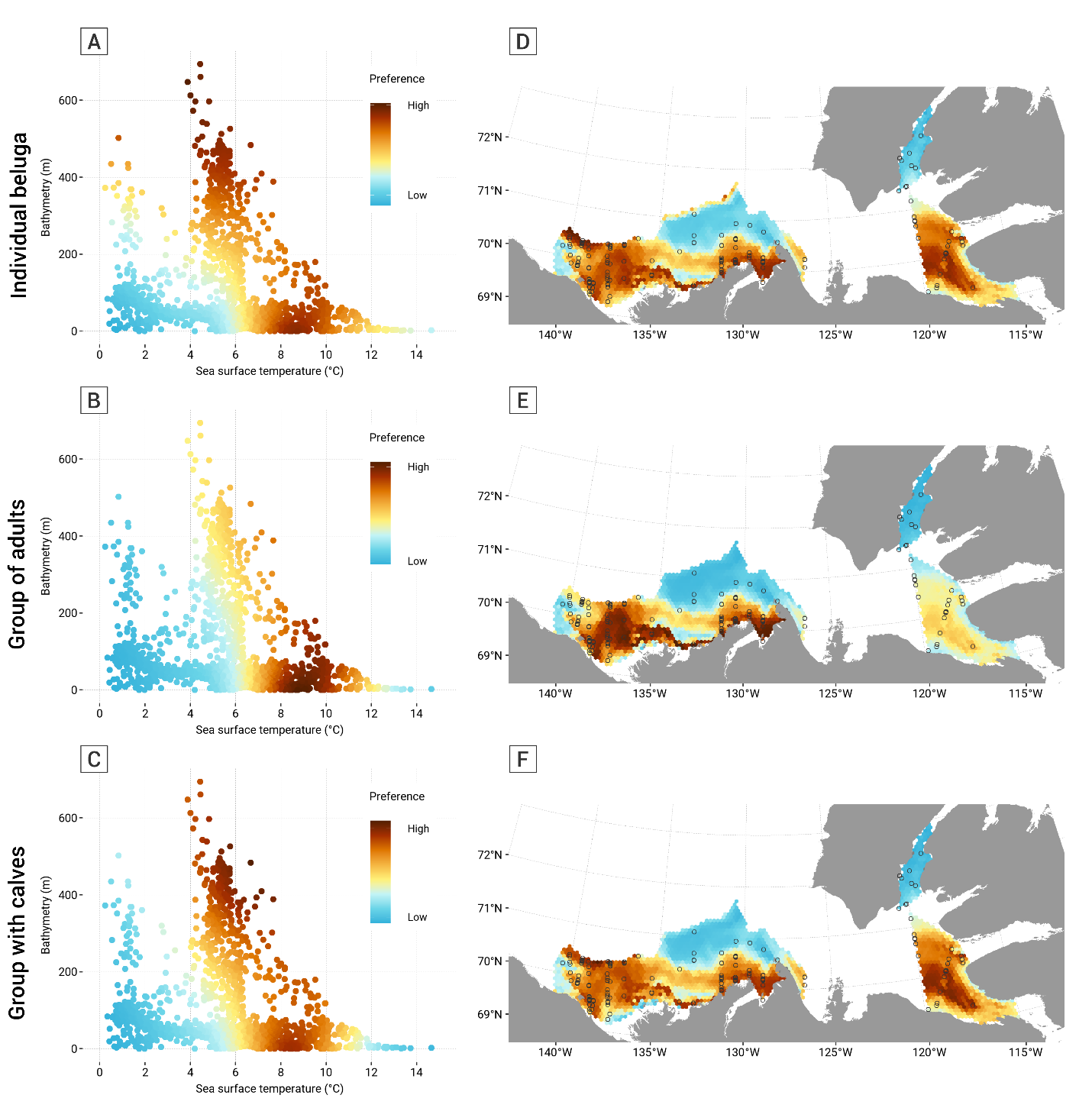
A

Chart

Description automatically generatedChart, line chart

Description automatically generated

**Figure S5** Response curve of the best model for August with sea surface temperature and bathymetry. Values are scaled and centred.



**Figure S6** Comparison of the relative preference (Left) and map of the preference (Right) of beluga group types for the models including sea surface temperature and bathymetry in July. A) and D) show individual belugas; B) and E) groups of adult belugas and C) and F) groups of belugas with at least one calf.

### Literature Cited

McCabe, J. D., Clare, J. D., Miller, T. A., Katzner, T. E., Cooper, J., Somershoe, S., Hanni, D., Kelly, C. A., Sargent, R., Soehren, E. C., Threadgill, C., Maddox, M., Stober, J., Martell, M., Salo, T., Berry, A., Lanzone, M. J., Braham, M. A. & McClure, C. J. W. (2021). Resource selection functions based on hierarchical generalized additive models provide new insights into individual animal variation and species distributions. *Ecography*, doi: https://doi.org/10.1111/ecog.06058

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

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, *73*,3-36. doi: https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R.* (2e ed.). Boca Raton, FL: CRC Press.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. (2009) GLM and GAM for Absence–Presence and Proportional Data. In (Eds.), *Mixed effects models and extensions in ecology with R* (pp. 245-259). New York: Springer New York.