Appendix 1. Supporting Text, Tables, and Figures.

Text A1.1. Environmental data

During the breeding season, the distance and duration of Razorbills' foraging trips are constrained by the need to make round trips to the nest (i.e. central place) to incubate their egg or feed their chick. Hence, the distance to any foraging location from the colony becomes an important environmental feature that will affect the space use of birds during nesting. Since Razorbills avoid flying over large patches of land, we reported distance to the colony as the effective geographical distance (Michels et al., 2001), defined here as the Euclidean distance from each grid cell to the colony, avoiding travel through land. Using functions from the *gdistance* package in R (van Etten, 2017), we performed a least-cost path analysis (Adriaensen et al., 2003) using a colony-centered resistance layer with terrestrial grid cells having a high resistance cost (i.e., 999999) and water cells having a low resistance cost (i.e., 1). The parameter to determine which adjacent pixels are used to determine cumulative cost values was set to 8 as a tradeoff between a more circular aspect around colonies (when using a parameter of 16) and computing time. Moreover, a parameter of 8 is the most common way to connect grids in GIS software (van Etten, 2017). The resulting least-cost path raster was a continuous variable of cumulative distance from the colony, avoiding any land surface. To adjust each analysis grid to a biologically meaningful extent, we excluded from further analysis all grid cells with a distance to the colony value larger than the mean of maximum foraging distances recorded across all colonies (79 km). For each grid cell, we also calculated the minimum distance to the nearest shoreline using a coastline feature (1:250 000 scale) and Euclidean distance tool in ArcGIS 10.2.2 (ESRI).

Bathymetric data were extracted from a 30 arc-second global grid of elevation on the British Ocean Data Centre database (BODC; https://www.bodc.ac.uk). Depth was defined as the below-zero (sea-level) values of elevation. After removing mainland values, there were still some raster cells containing small islands with non-negative values, thus creating a bias of "available" depth for Razorbills. All remaining non-negative values were set to zero and considered as not available. Slope and rugosity of seabed floor were extracted from depth, calculated in a 3 x 3 km roving window so that the value of each cell was calculated from the values of its surrounding cells (Horn, 1981). Both metrics were calculated with the *terrain* function of the *raster* package (Hijmans and van Etten, 2012).

We used chlorophyll-*a* concentration (mg/m^3) as a proxy for primary productivity estimated from Visible Infrared Imaging Radiometer Suite (VIIRS). Chlorophyll monthly composite maps (0.0375 degree resolution) were obtained from the National Oceanic and Atmospheric Administration (NOAA) ERDDAP data server (https://coastwatch.pfeg.noaa.gov) for June and July from 2015 to 2018. Each monthly georeferenced data sets where transformed into a raster using *NetCDF Raster Layer* tool in ArcGIS 10.2.2.

Sea surface temperature (SST; °C) data for the St. Lawrence Gulf and Estuary was extracted as monthly composite maps for June and July from 2015 to 2018 from Galbraith et al. (2018) at a resolution of 0.01 degree of latitude and 0.015 degree of longitude.

Dynamic oceanographic data such as SST and chlorophyll-*a* concentration can vary temporally. Between-years Pearson's correlation coefficients for chlorophyll-*a* concentration monthly composites were 0.67 ± 0.10 (±SD) for June and 0.67 ± 0.10 for July and correlations for SST were 0.85 ± 0.04 for June and 0.87 ± 0.11 for July. Between-months correlation coefficients ranged from 0.65 to 0.81 for chlorophyll-*a* concentration monthly composites and from 0.66 to 0.83 for SST. For prediction purpose, we averaged years (2015–2018) and months (June and July) and used the resulting rasters for both covariates in the boosted regression tree (BRT) model. Climatological covariates are preferred since they represent the long-term state (i.e., 4 years in this case) of the environment at a given location during moment of interest (i.e., breeding season in this case (Mannocci et al., 2017)). This monthly scale was considered a good trade-off to capture coarse dynamic marine features, while reducing gaps in maps (Derville et al., 2018, Perez-Correa et al., 2020). The spatial coverage of the SST layer was incomplete in some areas with a complex shoreline. We interpolated data from an 5 x 5 km roving window to fill those gaps (i.e., NAs) using the function *focal* implemented for R package *raster* (Hijmans, 2019). Each grid cell with missing values at the center of a non-empty roving window was assigned a value equal to the average of the non-empty pixel values within the window as described in Jaafar and Ahmad (2020), which used Landsat 7 images: thermal, red, near-infrared, and surface reflectance shortwave bands.

Text A1.2 Spatial autocorrelation

Spatial autocorrelation is an inherent problem in ecological sampling and must be accounted for in distribution models (Dormann, 2007). For each colony grid, we calculated Moran's *I* (a measure of global spatial autocorrelation (Moran, 1950)) and Geary's *C* (a measure of local spatial autocorrelation (Geary, 1954)). Values of Morans's *I* range between -1 and 1 while Geary's *C* range from 0 to 2. Values near 0 and 1 for Moran's *I* and Geary's *C*, respectively, indicate random spatial patterns and thus, no spatial autocorrelation. Both metrics were calculated for the 50 nearest neighbouring grid cells using functions *moran.test* and *geary.test* implemented in the package *spdep* (Bivand et al., 2013, Bivand and Wong, 2018) using the software R (R Core Team, 2019). Our data provided low evidence of spatial autocorrelation for both habitat suitability (Moran's I = 0.002 ± 0.006 SD; Geary's C: 0.997 ± 0.006 SD, n = 6) and density (Moran's I = 0.13 ± 0.10 SD; Geary's C: 0.91 ± 0.08 SD, n = 6) grids. Therefore, no additional measures were taken to account for spatial autocorrelation in the modelling process.

Table A1.1. List of environmental covariates used to predict the habitat suitability and density of Razorbill in Quebec using boosted regression trees.

Covariate (abbreviation)	Unit	Original resolution	Description
Distance from the colony (distcol)	m	1 km	Minimum distance by sea from the centroid of each grid cell to the colony location
Distance to the nearest shoreline (distshore)	m	1 km	Euclidean distance from the centroid of each grid cell to nearest shoreline
Depth	m	30 arc-second	Depth at centroid of each grid cell
Rugosity	m	30 arc-second	Standard deviation in depth in a 3 x 3 km roving window from the centroid of each grid cell
Slope	º	30 arc-second	Slope from a 3 x 3 km roving window from the centroid of each grid cell
Chlorophyll *a* concentration (chloro)	mg/m^3	0.0375º	Average of monthly composite in chlorophyll *a* concentration calculated for June and July 2015-2018
Sea surface temperature (sst)	ºC	0.01º lat * 0.015º long	Average of monthly composite in sea surface temperature calculated for June and July 2015-2018
Region	-	-	Region of the St. Lawrence river where the colony is present; either Gulf or Estuary

Table A1.2. Cross-validation results to find optimal parameters of boosted regression trees used to predict the habitat suitability and density of Razorbills around their colony in the St. Lawrence Gulf and Estuary, Québec, Canada. Multiple combinations of learning rate (*lr*) and tree depth (*td*) were tested to find the optimal number of trees (*nt*) for a given combination. A first BRT model is built using *nt* = 50, and average cross-validation deviance over 10 folds ± standard deviation (CV dev. ± SD) is recorded. The next BRT is built by adding 50 trees and so on, until the minimization of cross-validation deviance. Here we present the optimal number of trees (optimal *nt*) for each combination, for both models. The combination used to fit each model is in bold.

Tree paraneters		Habitat suitability [a]		Density [a]	
lr	*td*	optimal *nt*	CV dev. ± SD	optimal *nt*	CV dev. ± SD
0.001	3	10000	0.068 ± 0.001	6150	4.59 ± 0.379
0.001	4	10000	0.067 ± 0.001	5400	4.58 ± 0.432
0.001	5	10000	0.066 ± <0.001	5150	4.50 ± 0.298
0.001	6	NA	NA	5500	4.47 ± 0.459
0.001	7	NA	NA	5350	4.50 ± 0.359
0.001	8	NA	NA	4850	4.57 ± 0.285
0.005	3	7700	0.065 ± 0.001	2000	4.53 ± 0.317
0.005	4	6350	0.064 ± <0.001	1600	4.51 ± 0.283
0.005	5	5900	0.064 ± 0.001	1250	4.49 ± 0.227
0.005	**6**	5100	0.064 ± <0.001	**1400**	**4.43 ± 0.321**
0.005	7	4850	0.064 ± 0.001	1350	4.56 ± 0.345
0.005	8	5000	0.063 ± 0.001	1350	4.51 ± 0.415
0.01	3	5950	0.064 ± 0.001	850	4.52 ± 0.353
0.01	4	4650	0.064 ± 0.001	1250	4.44 ± 0.285
0.01	5	4450	0.064 ± 0.001	750	4.45 ± 0.323
0.01	6	4350	0.063 ± 0.001	750	4.45 ± 0.288
0.01	7	4000	0.063 ± <0.001	500	4.53 ± 0.203
0.01	8	3250	0.063 ± <0.001	500	4.50 ± 0.366
0.05	3	3050	0.064 ± 0.001	NA	NA
0.05	4	2600	0.063 ± <0.001	NA	NA
0.05	**5**	**2100**	**0.063 ± <0.001**	200	4.45 ± 0.377
0.05	6	1850	0.063 ± 0.001	NA	NA
0.05	7	1600	0.063 ± 0.001	NA	NA
0.05	8	1550	0.063 ± 0.001	NA	NA

[a] The habitat suitability model and the density model were fitted using a Bernoulli and a Poisson distribution, respectively (*n* = 9064 grid cells).

Table A1.3 Distance from the colony (min, max, mean ± sd; km), sample size (positions and individuals) and date range (min, max).

Colony	min	max	mean	sd	*n* positions	*n* individuals	min date	max date
Betchouane Island	0.05	43.8	7.8	7.9	1757	11	June 7	July 14
Bicquette Island	0.05	37.4	7.4	8.0	2350	11	June 6	July 9
Corossol Island	0.08	156	22.3	20.1	1326	5	June 6	July 21
Gros Pèlerin Island	0.04	79.5	17.7	18.6	2804	17	May 26	July 4
Gros Pot Island	0.04	50.3	6.0	9.3	421	4	June 10	June 29
Iles Sainte-Marie	0.05	106	31.4	27.9	1145	10	July 4	July 25
All	0.04	156	15.2	18.5	9803	58	May 26	July 25

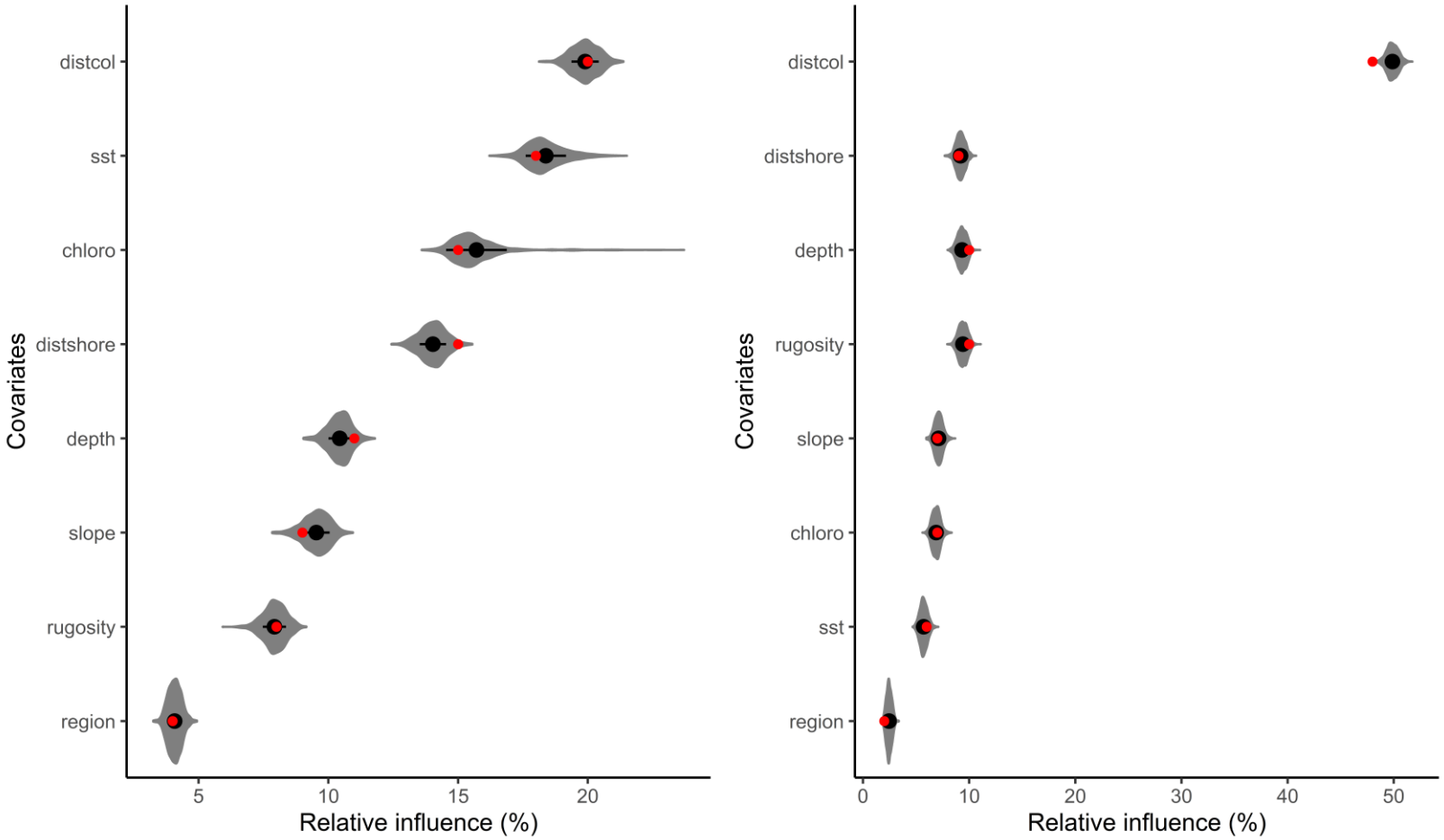

Figure A1.1 Sensitivity analysis of the effect of pseudo-absence sampling on the relative influence of covariates on models. Habitat suitability (left) and density (right) models were run 1,000 times with a random set of pseudo-absences for each iteration. Violin plots (gray) show the distribution of the relative influence of each covariate with means (black points) and standard deviations (black lines). Red points refer to the values computed by the model reported in the paper.

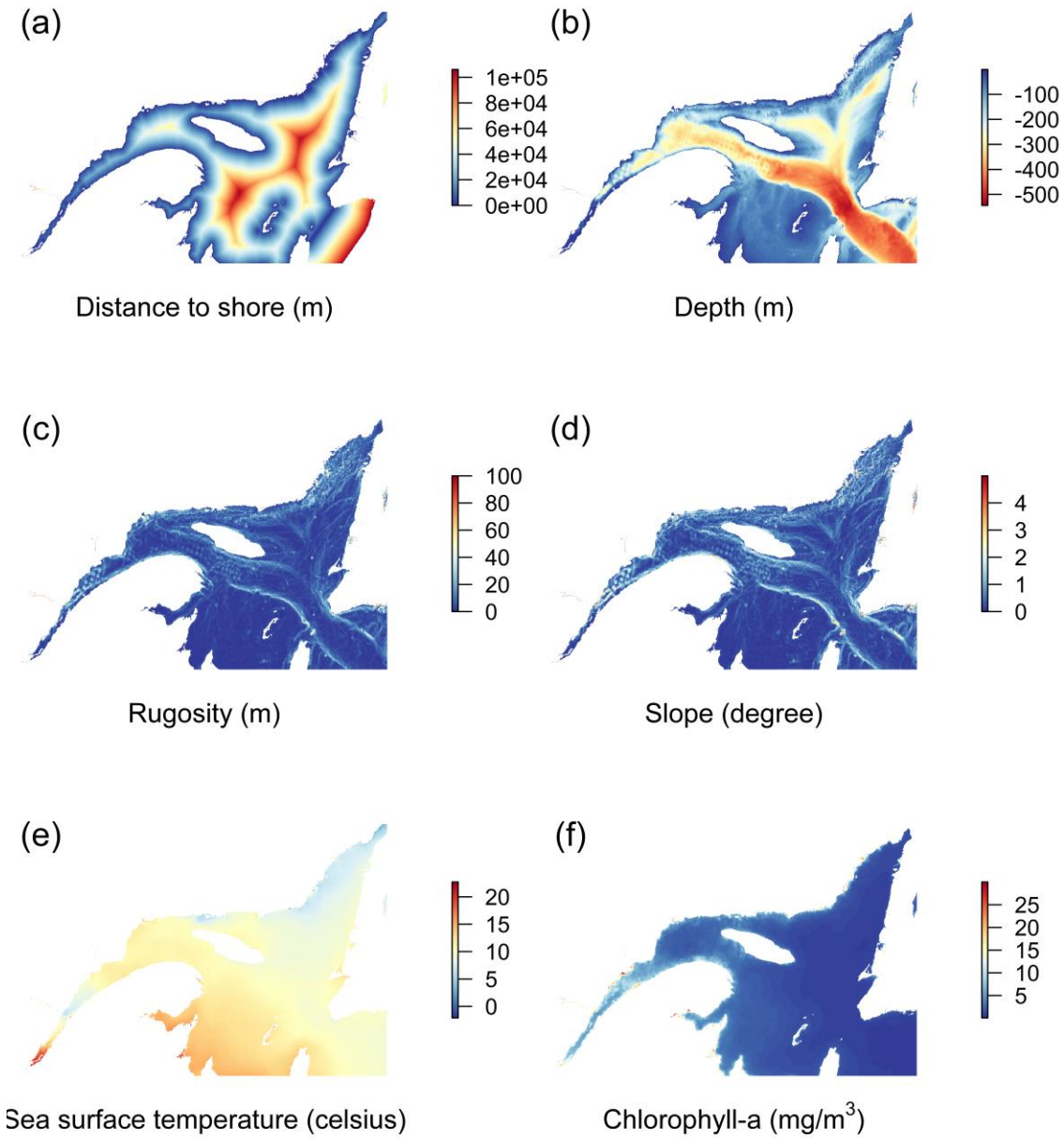

Figure A1.2. Spatial variation maps of six out of eight covariates used to predict Razorbill distribution and density in the St. Lawrence Gulf and Estuary. Here we present the distribution of (a) distance to the nearest shoreline (m), (b) depth (m), (c) rugosity (m), (d) slope (degree), (e) sea surface temperature (ºC) averaged for June and July 2015–2018, and (f) chlorophyll-a concentration (mg/m^3) averaged for June and July 2015–2018. See Table A1.1 for detailed descriptions and abbreviations of covariates.

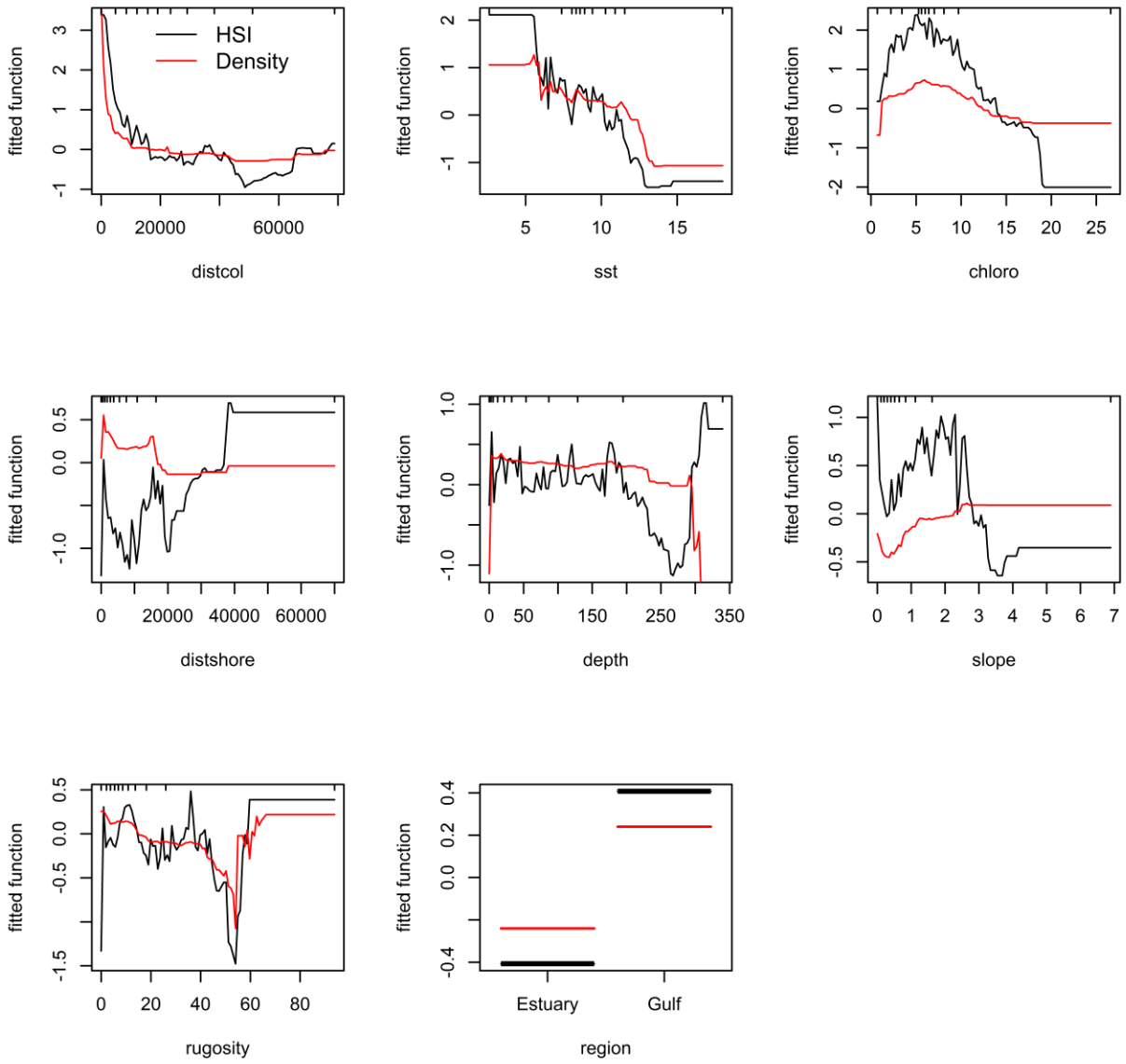

Figure A1.3. Partial dependence plots showing the relationship between environmental covariates and predicted habitat suitability index (HIS; black line) and density (red line) of Razorbills around their colony in the St. Lawrence Gulf and Estuary, Québec, Canada. Partial dependence fitted functions (lines) show the effect of a variable on the response after accounting for the average effects of all other variables in the model. Rug plots at inside top of plots show distributions across that variable, in deciles. See Table A1.1 for detailed descriptions of covariates.

LITERATURE CITED

Adriaensen, F., J. P. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulinck, and E. Matthysen. 2003. The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning* **64**:233-247.

Bivand, R. S., E. Pebesma, and V. Gomez-Rubio. 2013. *Applied spatial data analysis with R, Second edition. Springer, NY.* http://www.asdar-book.org/.

Bivand, R. S., and D. W. S. Wong. 2018. Comparing implementations of global and local indicators of spatial association. *TEST*:716-748.

Derville, S., L. G. Torres, C. Iovan, and C. Garrigue. 2018. Finding the right fit: comparative cetacean distribution models using multiple data sources and statistical approaches. **24**:1657-1673.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* **16**:129-138.

Galbraith, P. S., J. Chassé, C. Caverhill, P. Nicot, D. Gilbert, D. Lefaivre, and C. Lafleur. 2018. Conditions océanographiques physiques dans le golfe du Saint-Laurent en 2017. Page v + 82. Canadian Science Advisory Secretariat, Fisheries and Oceans Canada - Research Document 2018/050.

Geary, R. C. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* **5**:115-146.

Hijmans, R. J. 2019. Introduction to the 'raster' package (version 2.8-19). [online] URL: https://cran.r-project.org/web/packages/raster/vignettes/Raster.pdf.

Hijmans, R. J., and J. van Etten. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. [online] URL: http://CRAN.R-project.org/package=raster.

Horn, B. K. P. 1981. Hill shading and the reflectance map. *Proceedings of the IEEE* **69**:14-47.

Jaafar, H. H., and F. A. Ahmad. 2020. Time series trends of Landsat-based ET using automated calibration in METRIC and SEBAL: The Bekaa Valley, Lebanon. *Remote Sensing of Environment* **238**:111034.

Mannocci, L., A. M. Boustany, J. J. Roberts, D. M. Palacios, D. C. Dunn, P. N. Halpin, S. Viehman, J. Moxley, J. Cleary, H. Bailey, S. J. Bograd, E. A. Becker, B. Gardner, J. R. Hartog, E. L. Hazen, M. C. Ferguson, K. A. Forney, B. P. Kinlan, M. J. Oliver, C. T. Perretti, V. Ridoux, S. L. H. Teo, and A. J. Winship. 2017. Temporal resolutions in species distribution models of highly mobile marine animals: Recommendations for ecologists and managers. **23**:1098-1109.

Michels, E., K. Cottenie, L. Neys, K. De Gelas, P. Coppin, and L. De Meester. 2001. Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: a plea for using GIS modelling of the effective geographical distance. **10**:1929-1938.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* **37**:17-23.

Perez-Correa, J., P. Carr, J. J. Meeuwig, H. J. Koldewey, and T. B. Letessier. 2020. Climate oscillation and the invasion of alien species influence the oceanic distribution of seabirds. **10**:9339-9357.

R Core Team. 2019. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [online] URL: https://www.r-project.org/.

van Etten, J. 2017. R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software* **76**:1-21.