The following supplements accompany the article

Habitat use of culturally distinct Galápagos sperm whale *Physeter macrocephalus* clans

Ana Eguiguren*, Enrico Pirotta, Maurício Cantor, Luke Rendell, Hal Whitehead

*Corresponding author: anaeguibur@gmail.com

Marine Ecology Progress Series 609: 257–270 (2019)

Supplement 1

Summary results from multiscale modelling approach

Because the relationship between sperm whale occurrence and environmental variables is strongly scale-dependent (Jaquet 1996; Wong and Whitehead 2014), we ran a preliminary multi-scale analysis, through which environmental variables were included in candidate models at either fine (0.10°), intermediate (0.30°) and coarse spatial scales (0.50°), and at either weekly or monthly temporal resolutions. The fine spatial scale, which is presented in the final results, was chosen to reflect the distance over which sperm whales could be detected (ca. 10 km) and the coarse spatial scale was chosen to represent the distance that sperm whales can travel in a day (ca. 50 km) (Whitehead et al. 2008). Weekly and monthly temporal scales were chosen to reflect the dimension of oceanographic features (upwellings and fronts) that could influence the distribution of sperm whales. For this multiscale modelling approach, we fitted alternative initial models of uncorrelated variables at each spatiotemporal resolution. After backwards stepwise selection, used to determine which variables should be retained at each spatiotemporal scale, we chose the model with the spatiotemporal resolution that resulted in the lowest QIC.

This analysis showed that the overall habitat use patterns were very similar to those obtained by modelling the relationships between environmental variables and clan identity at a single (fine) scale. Specifically, final models selected through the multiscale approach were characterised by similar retained variables and performance measurements (Table S1), relationships between retained variables and clan identity (Fig. S1) and predicted distributions (Fig. S2) for both 1980's and 2010's data. For this reason, we decided to report only results from the finest spatiotemporal scale.

Table S1. Final models selected through single-scale and multi-scale model selection procedures.

Decade	Model selection process	Formula	Spatial scale	Predictive accuracy (%)[1]	Standard Error[2]
1980's	single	s(latitude) + s(longitude) + slope + weekly sdSST	0.10°	73.3	5
	multi	s(latitude) + s(longitude) + slope + monthly relSST + weekly sdSST	0.10°	73.0	6
2010's	single	s(longitude) + s(latitude) + weekly relSST + weekly sdSST	0.10°	22.3	12
	multi	s(longitude) + s(latitude) + weekly relSST + weekly sdSST	0.30°	38.7	13

1. Predictive accuracy measured through leave-one-out cross-validation
2. Standard error was calculated as the standard error of predictive accuracy across encounters over the square root of the number of encounters.

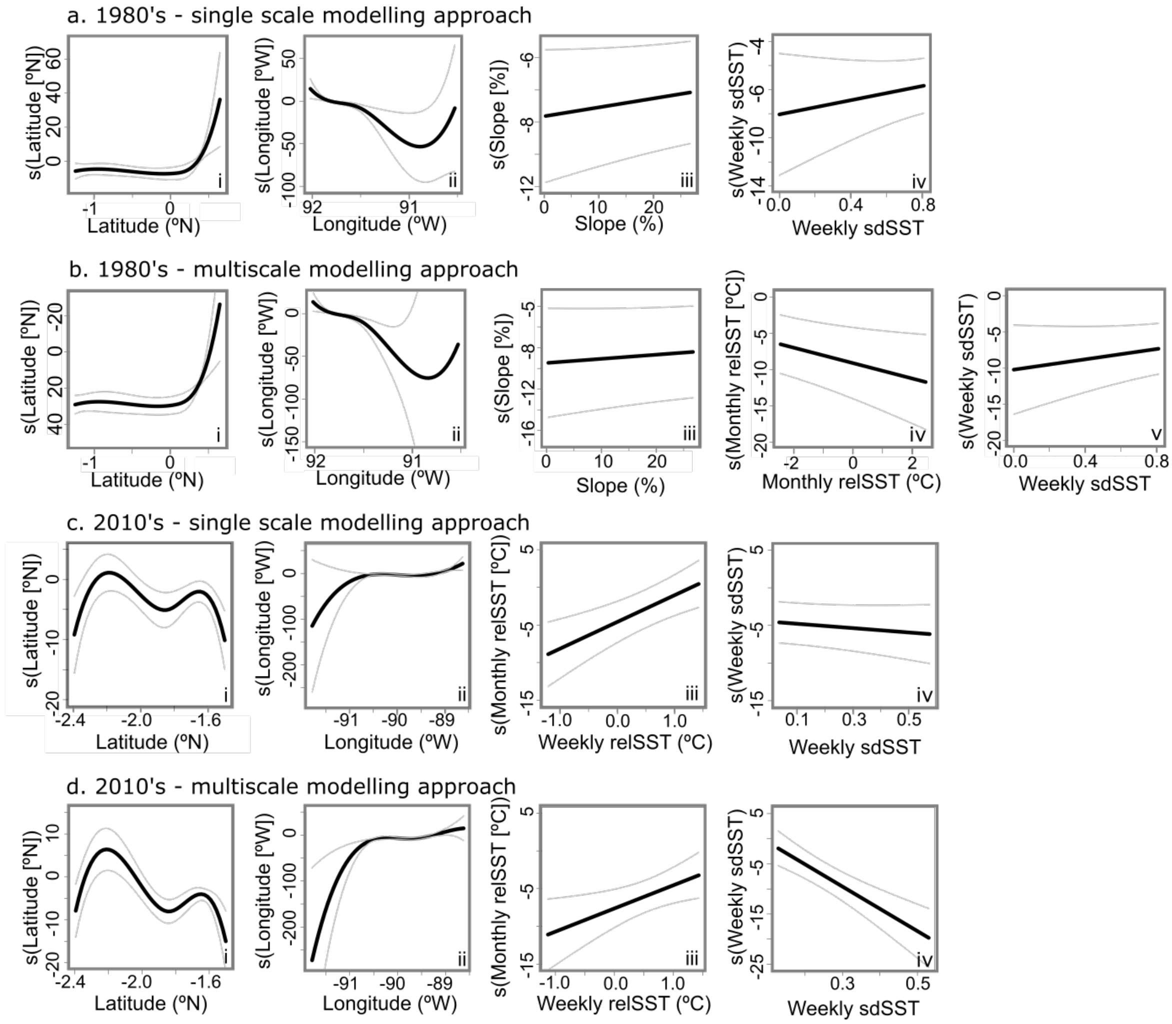

Figure S1. Partial plots of $\log_e$(odds) of female and juvenile sperm whales found off the Galápagos Islands belonging to (a & b) the *Plus-One* clan in the 1980's study period and (c & d) the *Four-Plus* clan in the 2010's study period. Models selected through (a & c) a single scale and (b & d) a multiscale modelling approach are shown. In (a), clan identity = *Plus-One* is modelled as a function of (a-i) latitude, (a-ii) longitude, (a-iii) slope, and (a-iv) weekly sdSST through a single scale modelling approach, in which environmental variables were only included at fine spatial and temporal resolutions (0.10° and weekly, respectively). In (b), clan identity = *Plus-One* is modelled as a function of (b-i) latitude, (b-ii) longitude, (b-iii) slope, and (b-iv) monthly relSST, and (b-v) weekly sdSST through a multiscale modelling approach, in which model selection favored variables included at a fine spatial scale (0.10°). In (c), clan identity = *Four-Plus* is modelled as a function of (c-i) latitude, (c-ii) longitude, (c-iii) weekly relSST, and (c-iv) weekly sdSST through a fine multiscale modelling approach. In (d), clan identity = *Four-Plus* is modelled as a function of (d-i) latitude, (d-ii) longitude, (d-iii) weekly relSST, and (d-iv) weekly sdSST relSST through a multiscale modelling approach, in which model selection favored variables included at an intermediate spatial scale (0.30°).

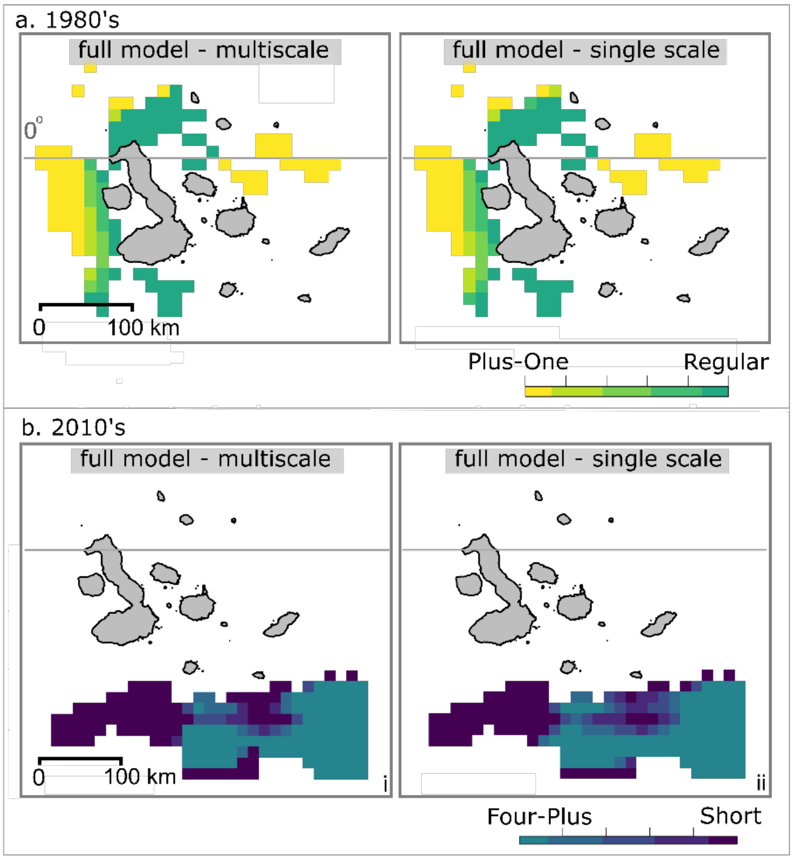

Figure S2. Predicted distribution of sperm whales of different clans off the Galápagos Islands mapped at 0.12° resolution in (a) the 1980's and (b) the 2010's. Maps were generated through final models including geographic and environmental variables selected through a single scale modelling approach (a-ii & b-ii) or a multiscale modelling approach (a-i & b-i).

Supplement 2

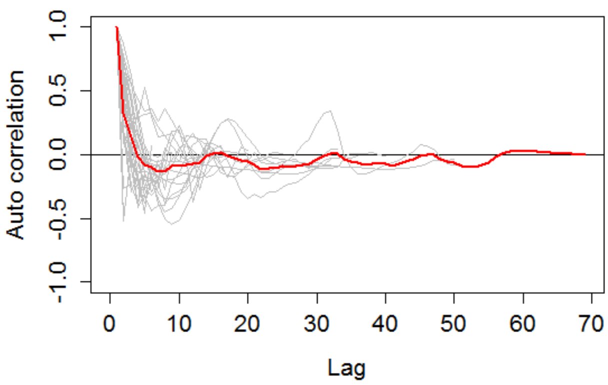

Fig. S3. Autocorrelation function (ACF) plot of the residuals of the final model of clan identity of sperm whales off the Galápagos Islands in the 1980's study period. Data points were collected approximately every hour, so a lag of 1 represents ~1 hr. Grey lines show the autocorrelation within individual encounters and the red line represents the mean auto-correlation across encounters.

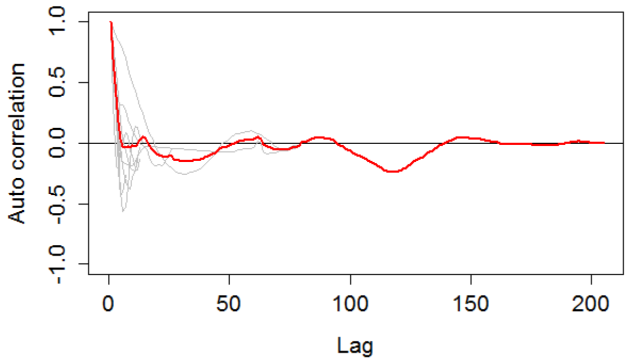

Fig S4. Autocorrelation function (ACF) plot of the residuals of the final model of clan identity of sperm whales off the Galápagos Islands in the 2010's study period. Data points were collected approximately every hour, so a lag of 1 represents ~1 hr. Grey lines show the autocorrelation within individual encounters and the red line represents the mean autocorrelation across encounters.

Supplement 3

Interpolation and data filtering methods

Since the geographic positions were collected at irregular intervals in the 1980's period using SATNAV, we linearly interpolated latitude and longitude so that positions were available approximately every hour. This aimed to minimize biases that result when the intervals at which data are collected depend on unknown factors that may affect the response (i.e., clan identity) (Bůžková et al. 2010). We chose a linear interpolation method for simplicity and because the resulting error that arises from linear interpolation (Tremblay et al. 2006) is expected to be smaller than the *ca.* 0.5-3.0 kilometre spread of a sperm whale group (Whitehead 2003), and much smaller than the scale at which we analyzed environmental variables (*ca.* 10 km). The 2010's study period data were collected more often and regularly. To make them consistent with previous years, we subsampled geographic positions so that they were also available approximately every hour.

We restricted our analysis to regions that were consistently surveyed throughout study years (see Milligan 2013). To do this, we calculated the total number of geographic positions recorded each decade during and between encounters in 0.10º by 0.10º cells and selected only geographic positions in cells where the count exceeded the 70th percentile for that decade. We also included only the geographic positions that occurred over waters >1,000 m deep because shallow waters were not consistently monitored acoustically, as well as points that occurred more than five kilometres away from shore because the quality of Pathfinder data around coastlines is low (NDOC/SOG 2009). For the 2010's study period, we only included geographic positions south of the archipelago (Fig. 1b), because we were interested in evaluating if clans had differential preferences when in sympatry. Thus, including locations to the west of the archipelago (where only the *Plus-Four* clan was found) would confound the analysis.

Supplement 4

Correlation coefficients for environmental and geographic variables used to predict clan identity in the 1980's and 2010's

Table S2. Correlation matrix for variables used to model differences in the niche of sperm whale clans off the Galápagos Islands, extracted at fine spatial resolution (0.10°) for the 1980's study period. Coefficients above the 0.4 threshold are bolded.

	latitude	depth	slope	weekly relSST	weekly sdSST
longitude	0.39	0.11	-0.16	-0.16	-0.06
latitude		0.18	0.02	0.09	0.09
depth			**0.81**	-0.07	0.20
slope				-0.04	0.25
weekly relSST					0.14

Table S3. Correlation matrix for variables used to model differences in the niche of sperm whale clans off the Galápagos Islands extracted at fine spatial resolution (0.10°) for the 2010s's study period. Coefficients above the 0.4 threshold are bolded.

	latitude	depth	slope	weekly relSST	weekly sdSST	Chla
longitude	0.05	**0.41**	0.02	0.23	-0.10	0.10
latitude		**0.59**	**0.45**	-0.35	0.00	0.05
depth			**0.68**	-0.17	-0.14	0.12
slope				-0.01	-0.12	-0.02
weekly relSST					-0.22	**0.67**
weekly sdSST						-0.24

Supplement 5

Prediction maps— methods for limiting mapped regions and rasterizing probability grid

To produce maps of expected annual clan distribution within areas where whales were found as predicted by the final models, we generated a regular grid of sample points separated by 0.10º latitude and 0.10º longitude for which we extracted topographic and oceanographic variables at the appropriate spatial and temporal scales in the R environment, using the packages rerddapXtracto, sp and raster (Pebesma & Bivand 2005, Bivand et al., 2013, 2016, Hijmans 2016, Mendelsson 2016). For oceanographic variables, which fluctuate over time, we extracted data at the appropriate temporal scale for all months surveyed during study years. We calculated averages of these variables for each point in the grid in each study period and used these as input to calculate the probability of whales belonging to a given clan using final models.

To map predictions only in the regions where whales were found, we generated a raster map of the geographic positions recorded during encounters with identified *Plus-One* and *Regular* clan whales in 0.10º by 0.10º cells between 1985 and 1989, and with the *Short and Plus-One* clans between 2013 and 2014. We included only sample points within 20 km of a cell where whales were found. We also excluded sample points in waters shallower than 1000 m. We standardized variables of remaining sample points with respect to the mean and standard deviation of the data used to fit the final models. We calculated predicted probabilities for each sample point using the final models and converted the grid of sample points to a raster image.

The Galápagos coastline shapefile is available through StatSilk (www.statsilk.com/maps/download-free-shapefile-maps).

Supplement 6

Summary tables of initial candidate models and final models

Table S4. Initial candidate models of clan identity of sperm whales found off the Galápagos Islands in the 1980's study period at fine spatial resolution (0.10°).

Model	Formula
A	Latitude + longitude + slope + weekly relSST + weekly sdSST
B	Latitude + longitude + depth + weekly relSST + weekly sdSST

Table S5. QIC and ΔQIC of final models of sperm whale clan identity for the Galápagos selected through backwards stepwise selection in the 1980's study period at fine spatial resolution (0.10°).

Model	Formula	QIC	ΔQIC
A	**s(latitude) + s(longitude) + slope + weekly sdSST**	**451.07**	**0**
E	s(latitude) + s(longitude) + weekly sdSST	465.03	13.96

Table S6. Initial candidate models of clan identity of sperm whales found off the Galápagos Islands in the 2010's study period at fine spatial resolution (0.10°).

Model	Formula
A	longitude + slope + weekly sdSST + chla
B	depth + weekly sdSST + chla
C	longitude + latitude + weekly sdSST + chla
D	longitude + slope + weekly relSST + weekly sdSST
E	depth + weekly relSST + weekly sdSST
F	longitude + latitude + weekly relSST + weekly sdSST

Table S7. QIC and ΔQIC of final models of sperm whale clan identity for the Galápagos selected through backwards stepwise selection in the 2010's study period at fine spatial resolution (0.10°).

Model	Formulas	QIC	ΔQIC
A	s(longitude) + chla	473.87	73.89
B	depth + chla	675.49	275.51
C	s(longitude) + s(latitude) + weekly sdSST + chla	426.96	26.98
D	s(longitude) + weekly relSST	487.04	87.06
E	depth + weekly relSST	632.71	232.74
F	**s(longitude) + s(latitude) + weekly relSST + weekly sdSST**	**399.98**	**0.00**

Reference List

Bůžková P, Brown ER, John-Stewart GC (2010) Longitudinal data analysis for generalized linear models under participant-driven informative follow-up: an application in maternal health epidemiology. Am J Epidemiol 171:189–197

Milligan M (2013) Fine-scale distribution, habitat use, and movements of sperm whales. MSc dissertation, Dalhousie University, Halifax, NS

NODC/SOG (2009) 4 km Pathfinder version 5 user guide. Satellite Oceanography Group, National Oceanographic Data Center (NDOC), NOAA Satellite and Information Service NEDSIS). Available at: www.nodc.noaa.gov/sog/pathfinder4km/userguide.html (accessed 6 Jun 2017)

Whitehead H (2003) Sperm Whales, Social Evolution in the Ocean. The University of Chicago Press, London