Supplementary File B:
Detailed methods on data processing

Biologically Important Areas II for cetaceans within U.S. and adjacent waters – Hawai'i Region

Michaela A. Kratofil[1], Annette E. Harnish[1], Sabre D. Mahaffy[1], E. Elizabeth Henderson[2], Amanda L. Bradford[3], Stephen W. Martin[4], Barbara A. Lagerquist[5,6], Daniel M. Palacios[5,6], Erin M. Oleson[3], Robin W. Baird[1*]

[1]Cascadia Research Collective, Olympia, WA, USA
[2]Naval Information Warfare Center Pacific, San Diego, CA, USA
[3]Pacific Islands Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Honolulu, HI, USA
[4]National Marine Mammal Foundation, San Diego, CA, USA
[5]Marine Mammal Institute, Oregon State University, Newport, OR, USA
[6]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR, USA

*Correspondence: Robin W. Baird (rwbaird@cascadiaresearch.org)

This document provides details on data processing methods common among all Hawai'i BIAs, particularly those that incorporated satellite tag data (general methods summarized in Kratofil et al., this issue). BIA-specific methods (e.g., residence time/occupancy pattern for MHI humpback whales) are described in the respective BIA's full description in Supplementary File A.

Satellite tag data: Pre-processing

Cascadia Research Collective, Pacific Islands Fisheries Science Center, and Naval Information Warfare Center Pacific data:

Argos location data were first processed through the Distance-Angle-Rate filter of the Douglas-Argos Filter (DAF, Douglas et al. 2012) accessed through Movebank (Kranstauber et al., 2011) to remove erroneous locations based on unrealistic traveling speeds and turning angles. User-defined settings for each species are detailed in Table S1 below. For Fastloc-GPS tags, GPS locations were restricted to those with a residual error less than 35 (Dujon et al., 2014) and/or time error less than 10 seconds long. Restricted GPS locations were additionally filtered through a general-purpose speed filter on Movebank (Kranstauber et al., 2011) with user-defined error set to 1,000 m and speed corresponding to the same rate as used in the DAF. GPS locations derived from messages from only 4 satellites often have poorer location accuracy than those calculated by at least 5 satellites, and thus, some studies have restricted GPS locations to those with at least 5 or 6 satellite acquisitions for analyses (Lowether et al., 2015; Witt et al., 2010). However, this

additional quality control criterion excludes a substantial proportion of the GPS locations in these datasets. Therefore, we retained GPS locations calculated with 4 satellites (and after filtering as detailed above) and account for their greater location uncertainty further in the analyses. Douglas-filtered Argos locations and restricted Fastloc GPS locations were merged (for applicable tags) for subsequently analyses.

Oregon State University data:
Humpback whale telemetry data from Oregon State University were pre-processed following their custom track editing protocol (see Palacios et al., 2019, 2020). First, locations on land and Argos locations with a location quality class (LC) of Z (poorest) were removed. The remaining locations were filtered as follows: lower-quality LCs (0, A, or B) were excluded if they were received within 20 minutes of higher-quality locations (1, 2, or 3). Travel speeds between the remaining locations were calculated, and where the speed between two locations exceeded 14 kilometers per hour, the location resulting in the shortest track was retained and the other discarded.

Satellite tag data: Movement modeling

After pre-processing, resultant locations were fit to a continuous-time correlated random walk model via *crawl* v2.2.3 (Johnson et al., 2008; Johnson and London, 2018) in the R environment for statistical computing (R Core Team, 2021). This movement modeling approach allowed us to estimate locations at regular intervals (from irregular Argos/GPS data) while also accounting for positional uncertainty that typifies Argos data. To formulate the error component of the *crawl* model, Argos error ellipse measures were used if tracks were derived from deployments after 2007 (i.e., data were Kalman-processed, Collecte Localisation Satellites, 2015), and estimated accuracies for Argos location classes (LC3-1 as 250m, 500m, and 1500m, respectively, and LC0-B unbounded) were used for tags deployed 2007 and earlier when the Kalman processing algorithm was not available. GPS locations (deployment locations and/or from Fastloc GPS tags) do not include LC classes nor error ellipse measures. Therefore, where applicable, GPS locations were assigned an error radius of 50 meters if they were deployment locations or derived from at least 5 satellites and 1,000 meters if they were derived from 4 satellites (Bryant 2007; Dujon et al., 2014; Hazel, 2009). Fitted *crawl* models were used to predict locations at a time step either (1) appropriate for the data considering the median time between filtered locations (e.g., 1 hour) or (2) appropriate for subsequent analyses (e.g., residence time calculations, kernel density analyses). Details on specific *crawl* time-steps used for each species/BIA are provided in Table S1.

Satellite tag data: Re-routing around land

Crawl locations were re-routed around land as needed using the *pathroutr* package (London, 2021), either using a polygon representing the island with an added distance band (for nearshore dwelling species) or a polygon representing an isobath, to ensure re-routed locations were within reasonable proximity to shore or depths for the species. For all cases, the polygon serving as the basis of the BIA had an inner boundary defined as the barrier polygon used in this re-routing step. *Crawl* positions during periods of large transmission gaps (with a 1-day gap threshold) were removed from each individual's track (where applicable) to limit locations characterized by large positional uncertainty resulting from interpolation over long periods without any original Argos data.

Satellite tag data: Kernel density methods

Some BIAs involved kernel density estimation (KDE) of satellite tag data to generate a utility distribution (UD) of the sample population (Worton, 1989) from which an isopleth of the UD was drawn to serve as the basis for the BIA (e.g., 50% for core areas). The general KDE process follows: to mitigate spatial autocorrelation, a coarse timestep of *crawl* locations was used (e.g., 4-hour); and one of each pair of tagged individuals moving in concert was removed (to reduce pseudoreplication). Following above, *crawl* positions during periods of large transmission gaps (with a 1-day gap threshold) were removed from each individual's track (where applicable) to avoid generation of artificially "dense" areas resulting from interpolation over long periods without any underlying Argos/GPS data. All tag locations were pooled together, and the contribution of each tag's location was weighted to the overall kernel density based on deployment length, and the KDE was re-scaled so it integrated to 1 (Hauser et al. 2014; Hill et al. 2019), such that locations from shorter deployments would have less weight than those with longer deployments. This weighting method also mediates bias associated with tag deployment location, as animals with longer transmission durations will have had a greater opportunity to move outside of the deployment area. Kernel densities were estimated using the bivariate plug-in bandwidth (or smoothing parameter) matrix (Duong & Hazelton, 2003, 2005; Duong, 2007) accessed through the *ks* package for R (Duong, 2021). The location weighting was completed using the weights argument within the *ks* package (Duong, 2021).

Table S1. Information on satellite tag data processing methods by each species/population with a BIA.

BIA-Species	DAF^: minrate* (km/hr)	Parent BIA crawl timestep (hr)	Child BIA crawl timestep (hr)	*Pathroutr*: barrier polygon
MHI Short-finned pilot whale	15	4	4	300-m isobath
MHI Insular False killer whale	20	4	4	Land + 50-m distance band
NWHI False killer whale	20	4	NA	Land + 50-m distance band
KNO Rough-toothed dolphin	20	1	4	200-m isobath
MNHI Rough-toothed dolphin	20	1	NA	300-m isobath
Kohala Melon-headed whale	15	1	NA	200-m isobath
KNOMN Bottlenose dolphin	20	1	1	Land + 50-m distance band
HI Bottlenose dolphin	20	1	NA	Land + 50-m distance band

OMNHI Pantropical spotted dolphin	20	1	1	Land + 50-m distance band
OMNHI Blainville's beaked whale	10	4	4	300-m isobath
HI Cuvier's beaked whale	10	4	4	800-m isobath
HI Pygmy killer whale	15	1	NA	100-m isobath
OMN Pygmy killer whale	15	1	NA	400-m isobath
MHI Humpback whale	15	10 min[+]	10 min[+]	Land + 50-m distance band

*minrate = maximum realistic sustainable rate of movement (based on field observations)
^Douglas-Argos Filter
[+]Humpback whale telemetry data transmitted at a much coarser frequency than this time-step; a fine time-step of 10 minutes was used for this R-BIA for residence time and occupancy pattern calculations. These methods are detailed in Supplementary File B (full MHI humpback whale R-BIA description).

Other DAF user-defined settings common across all species/populations:
KEEP_LC = 2
Argos location class level to be exempt from removal by the filter.

Maxredun = 3km
Radius in kilometers within which two points are self-confirming (location redundancy).

Ratecoef = 25 degrees
Angle created by three consecutive points to identify acute angles characteristic of typical Argos location error.

Sightings data
As noted in Kratofil et al. (this issue), sighting and effort data used in BIA boundary determinations primarily came from Cascadia Research Collective small boat surveys and National Marine Fisheries Service large-vessel ship-based line-transect surveys around the main Hawaiian Islands (the latter source also included the Northwestern Hawaiian Islands). Cetacean research has been ongoing almost continuously in Hawaiian waters since the late 1960s (e.g., Norris and Dohl, 1980), and dozens of graduate degrees have been focused on Hawaiian cetaceans, primarily spinner dolphins or humpback whales. While these efforts have contributed substantially to understanding stock structure (e.g., Andrews et al., 2010 for spinner dolphins) as well as importance and spatiotemporal variability, they have primarily been undertaken in shallow, near-shore areas, and have typically been focused on a single species (mostly spinner dolphins and humpback whales). Attempting to compile and analyze these sources of effort and sighting data would be prohibitively time-consuming with limited benefit to boundary determinations. While data from aerial surveys for humpback whales have been used to inform

stock boundaries, aerial survey data for other species were not included in analyses given the survey mode (i.e., passing mode for most odontocetes) and higher altitude used for humpback surveys, potentially leading to higher species misidentification for smaller odontocetes.

BIA boundary creation

For BIA boundaries based off of minimum convex polygons (MCPs) around sighting and/or *crawl* locations, MCPs were generated using the *sf* package in R (Pebesma, 2018). For many BIA boundaries based off MCPs, a "band" was added to the MCP to account for positional uncertainty estimated by *crawl*. The *crawl* model provides estimates of standard error for predicted positions in the x and y planes (i.e., latitude and longitude); these estimates are informed by the movement model and the error component of the model (i.e., Argos error ellipse metrics or LC class; Johnson et al., 2008). Although statistical summaries of *crawl*-estimated standard errors varied across species and individual tagged whales, mean standard errors (in x and y) were generally between 2 and 3 km. Therefore, we applied a 3-km distance band to all MCPs across all applicable BIAs.

The area of each BIA was calculated using the *st_area* function within the *sf* package in R (Pebesma, 2018). This function calculates the geodetic area of an *sf* spatial polygon in the units of the coordinate reference system (CRS) specified for the spatial polygon. For all Hawaiʻi BIAs, the Hawaiʻi Albers Equal Area CRS (ESRI:102007) was specified for the polygons; the measurement units for this CRS are in meters. The calculated area of the BIA polygons (in meters) were converted to kilometers for reporting purposes. For BIAs that had more than one spatial polygon, the total area of the BIA was represented as the sum of the areas of all spatial polygons comprising the BIA.

References

Bryant, E. 2007. 2D location accuracy statistics for Fastloc RCores Running Firmware Versions 2.2 & 2.3. Otley, West Yorkshire. United Kingdom: Wildtrack Telemetry Systems Ltd. 1-5pp.

Collecte Localisation Satellites. 2015. Argos User's Manual. Available at http://www.argos-system.org/files/pmedia/public/r363_9_argos_users_manual-v1.6.4.pdf

Douglas, D.C., Weinzierl, R., Davidson, S.C., Kays, R., Wikelski, M., and G. Bohrer. 2012. Moderating Argos location errors in animal tracking data. Methods in Ecology and Evolution 6:999-1007.

Dujon, A.M., Lindstrom, R.T., and G.C. Hays. 2014. The accuracy of Fastloc-GPS locations and implications for animal tracking. Methods in Ecology and Evolution 5:1162-1169.

Duong, T., and M.L. Hazelton. 2003. Plug-in bandwidth matrices for bivariate kernel density estimation. Journal of Nonparametric Statistics 15:17-30.

Duong, T., and M.L. Hazelton. 2005. Cross-validation bandwidth matrices for multivariate kernel density estimation. Scandinavian Journal of Statistics 32:485-506.

Duong, T. 2007. Ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software 21:1-16.

Duong, T. 2021. Ks: Kernel smoothing. R package version 1.12.0 https://CRAN.R-project.org/package=ks

Hauser, D.D.W., K.L. Laidre, R.S. Suydam, and P.R. Richard. 2014. Population-specific home ranges and migration timing of Pacific Arctic beluga whales (Delphinapterus leucas). Polar Biology 37:1171-1183

Hazel, J. 2009. Evaluation of fast-acquisition GPS in stationary tests and fine-scale tracking of green turtles. Journal of Experimental Marine Biology and Ecology 374:54-68.

Hill, M.C., A.R. Bendlin, A.M. Van Cise, A. Milette-Winfree, A.D., Ligon, A.C. Ü, M.H. Deakos, and E.M. Oleson. 2019. Short-finned pilot whales (Globicephala macrorhynchus) of the Mariana Archipelago: individual affiliations, movements, and spatial use. Marine Mammal Science 35(3):797-824 doi:10.1111/mms.12567

Johnson, D.S., J.M. London, M.-A. Lea, and J.W. Durban. 2008. Continuous-time correlated random walk model for animal telemetry data. Ecology 89:1208-1215 doi:10.1890/07-1032.1

Johnson, D.S., and J.M. London. 2018. Crawl: an R package for fitting continuous-time correlated random walk models to animal movement data. Zenodo https://doi.org/10.5281/zenodo.596464

Kranstauber, B., A. Cameron, R. Weinzierl, T. Fountain, S. Tilak, M. Wikelski, and R. Kays. 2011. The Movebank data model for animal tracking. Environmental Modelling & Software 26(6):834-835 doi:10.1016/j.envsoft.2010.12.005

London, J.M. 2020. Pathroutr: an R package for (re-)routing paths around barriers (version v0.1.1-beta). Zenodo http://doi.org/10.5281/zenodo.4321827

Lowether, A.D., C. Lydersen, M.A. Fedak, P. Lovell, and K.M. Kovacs. 2015. The Argos-CLS Kalman Filter: Error structures and state-space modelling relative to Fastloc GPS data. PLoS ONE 10(4):e0124754 doi:10.1371/journal.pone.0124754

Palacios, D.M., B.R. Mate, C.S. Baker, C.E. Hayslip, T.M. Follett, D. Steel, B.A. Lagerquist, L.M. Irvine, and M.H. Winsor. 2019. Tracking North Pacific humpback whales to unravel their basin-wide movements. Final technical report prepared for Pacific Life Foundation. Marine Mammal Institute, Oregon State University, Newport, Oregon, USA. 30 June 2019. 58 pp. doi:10.5399/osu/1117 https://ir.library.oregonstate.edu/concern/technical_reports/z890s0924

Palacios, D.M., B.R. Mate, C.S. Baker, B.A. Lagerquist, L.M. Irvine, T.M. Follett, D. Steel, and C.E. Hayslip. 2020. Humpback whale tagging in support of marine mammal monitoring across multiple navy training areas in the Pacific Ocean: Final report for the Hawaiian breeding area in Spring 2019, including historical data from previous tagging efforts.

Prepared for Commander, US Pacific Fleet, and Commander, Naval Sea Systems
Command. Submitted to Naval Facilities Engineering Command Southwest, San Diego,
California, under Cooperative Ecosystem Studies Unit, Department of the Navy
Cooperative Agreement No. N62473-19-2-0002. 15 May 2020. 122 pp.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. The R
Journal 10(1):439-446. Doi: 10.32614/RJ-2018-009

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for
Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Witt, M.J., S. Åkesson, A.C. Broderick, M.S. Coyne, J. Ellick, A. Formia, G.C. Hays, P. Luschi,
S. Stroud, and B.J. Godley. 2010. Assessing accuracy and utility of satellite-tracking data
using Argos-linked Fastloc-GPS. Animal Behavior 80:571-581
doi:10.1016/j.anbehav.2010.05.022

Worton, B.J. 1989. Kernel methods for estimating the utilization distribution in home-range
studies. Ecology 70:164-168