S1 - Estimating boat-based survey effort without GPS tracklines

S. Derville, L. G. Torres, R. Dodémont, V. Perard, C. Garrigue

20 juillet, 2018

Supporting information S1 to "From land and sea, long-term data reveal persistent humpback whale breeding habitat in New Caledonia", 2018. Contact information: solene.derville@ird.fr

Runs on R version 3.4.4 (2018-03-15) Platform: x86_64-pc-linux-gnu (64-bit)

In this appendix, we describe the method to estimate the spatial extent of survey effort during non-systematic cetacean surveys at sea where boat GPS tracklines were not recorded.

1- Estimating daily "pseudo-tracklines"

When GPS tracks are not available, survey effort extent may be estimated from various latitude-longitude positions recorded along the day. Here, we used the locations for acoustic samplings, and the start/end of focal follows. Locations were linked together through time to reconstruct daily "pseudo-tracklines" (e.g., Fig. S1.a)

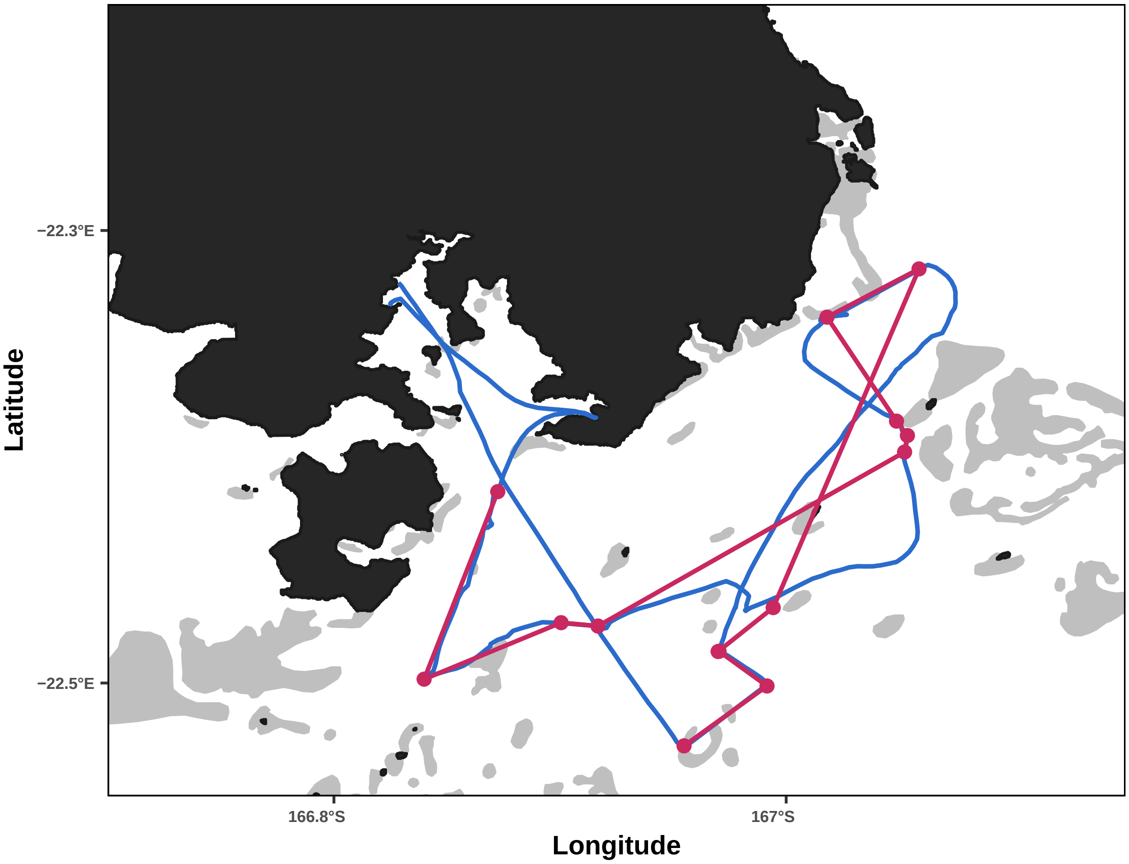

Figure S1.a: GPS trackline (blue) and pseudo-trackline (pink) reconstructed for a day of survey in the South Lagoon of New Caledonia (2012-08-20). Locations along the pseudo-trackline are represented with pink points.

2- Estimating annual maps of "pseudo-effort"

Pseudo-tracklines were considered a subsample of the real tracklines followed by the research boat over the course of a day. Seasonal maps of pseudo-effort were produced using a method similar to that applied to real GPS tracklines: they were interpolated at one position/min, sections off-effort were removed, and the remaining positions were smoothed with a LOESS applied with a varying bandwidth (i.e. span ranging from 0.002 and 0.02). The following codes were used to derive pseudo-effort maps at various bandwidth from locations recorded along each day of survey.

```
# this function projects and overlays a set of points with a Raster Layer. It counts the
 # number of points in each cell
# INPUTS:
  # points = a dataframe with a "x" and a "y" column in WGS84. also include a column "id"
         #on which the dataframe is split
  # grid = RasterLayer object on which the count is performed
  # UTM = logical argument. TRUE if coordinates are already projected and FALSE if they
          #are in lat/long
  # x, y = name of the columns containing x and y coordinates

Fun_CountinRaster <- function(points, UTM, x, y, grid) {
    ef.list <- dlply(points, ~id, function(t) {
      if (UTM == F){
        #project to UTM
        utm <- as.matrix(cbind(t[, x], t[, y]))
        utm <- data.frame(project(utm,
                      "+proj=utm +zone=58S +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0"))
        #put UTM coordinates into a dataframe
        t <- data.frame(utmx = utm[, 1], utmy = utm[, 2])
      } else { t<- data.frame(t[,c(x, y)])
                names(t) <- c("utmx","utmy")}
      # convert to SpatialPoints object
      coordinates(t) = ~ utmx + utmy
      proj4string(t) <-CRS("+proj=utm +zone=58S +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0")
      # count number of sightings per grid cell in raster g
      ef <- rasterize(t, grid, fun = "count")
      # when count = NA
      ef <- calc(ef, function(x) { x[is.na(x)] <- 0 ; return(x) })
    return(ef)
    })
    names(ef.list) <- unique(points$id)
  return(ef.list)
}

# locs.df is a dataframe containing interpolated locations along the pseudo-tracklines
  # it includes the geographical position of locations and the associated date-time
# grid is an epmty raster covering the study area

grids_pseudoONeffort <- Fun_CountinRaster(
  points = locs.df,
  UTM = T, x = "x", y = "y", grid = grid)
```

```
# this function interpolates rasters containing effort values per grid cell
# using a loess method which smoothing parameter can be manipulated (span)
# It is applied to a list of rasterlayers and returns a list of raster layers

Fun_INTmap <- function(raster.list, span){
  result <- llply(raster.list, function(x){
    d <- as.data.frame(x[[1]])
    int <- data.frame(coordinates(x))
    d <- cbind(d, coordinates(x))
    loess <- loess(layer ~ x*y, data = d, degree = 2, span = span)
    p <- predict(loess, newdata = int)
    int$z <- as.numeric(p)
    int[which(int$z<0),]$z <- 0
    coordinates(int) <- ~x+y
    int <- SpatialPixelsDataFrame(int, data=int@data, tolerance=0.01)
    int <- raster(int)
    return(int)
  })
  #result is a list containing as many elements as years and each element is a raster
  max_per_layers <- llply(result, function (r) {
    #calculate maximum effort value per year
    return(r@data@max)
   })
  #rescale to 0-100 using the maximum of yearly maxima
  result <- llply(result, function (r) {
    r <- calc(r, function(x) (x * 100) / max(unlist(max_per_layers)))
    return(r)
  })
  return(result)
}

# Smooth pseudo-tracklines with a given span
pseudoONeffort.map <- Fun_INTmap(grids_pseudoONeffort, span = 0.01)
```

Seasonal maps of pseudo-effort were created with various spans using the LOESS approach. In order to assess the best parametrization for this smoothing, we calculated the surface area covered by the pseudo-effort density contour and the real effort density contour per breeding season. The difference between these two surface areas (at a given contour of the density surface) is expected to converge towards zero when the pseudo-effort map is most similar to the real effort map (Fig. S1.b). The 0.01 bandwidth was selected to provide the most realistic estimate of pseudo-effort distribution from 1995 to 2002.

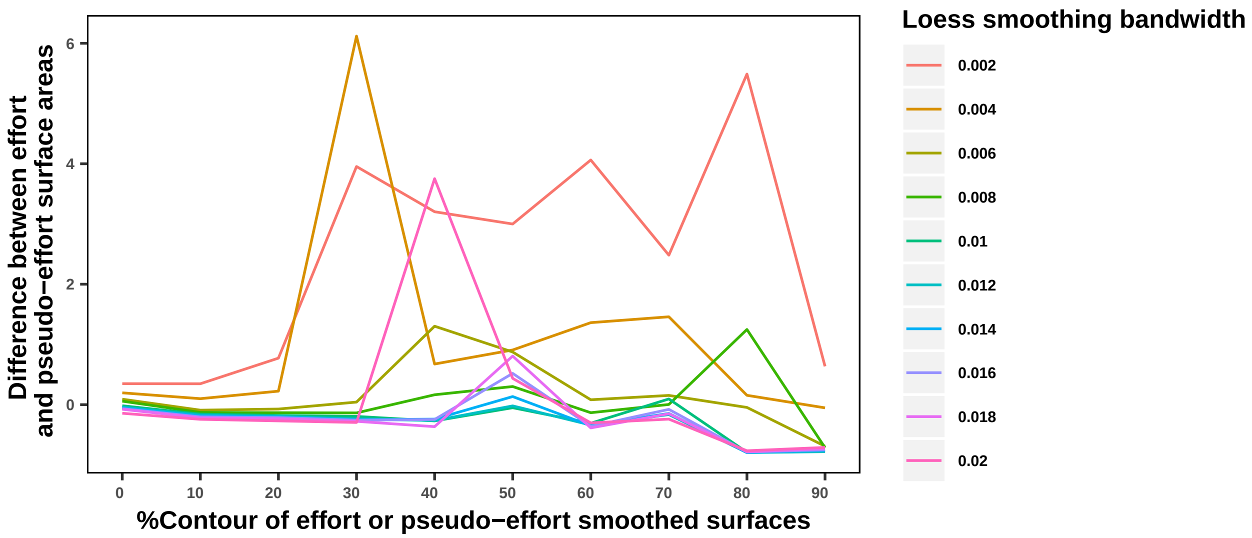

Figure S1.b: Difference between real effort and pseudo-effort surface areas for different bandwidth and at different % contours of the density surface.

The pseudo-tracklines and pseudo-effort maps were comparable in shape and extent to the real effort maps (e.g., Fig. S1.c).

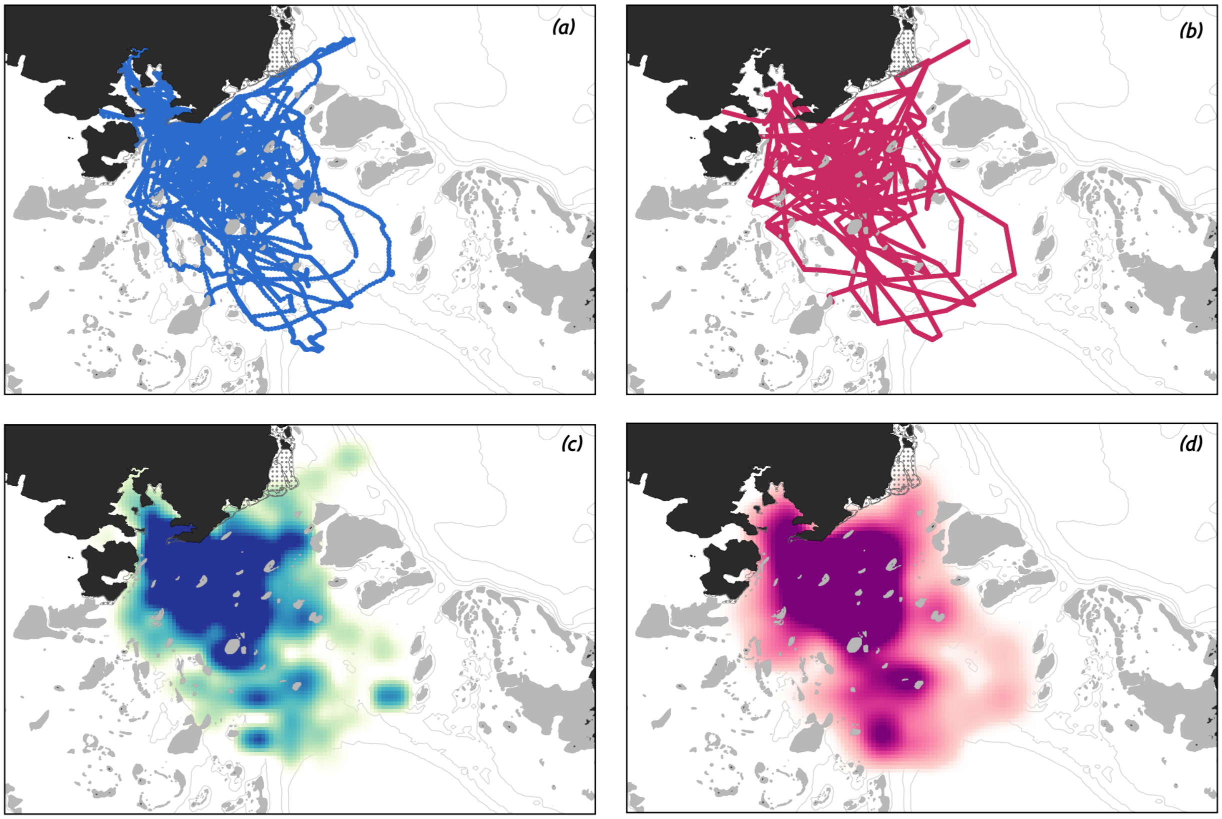

Figure S1.c: Comparative maps of GPS tracklines (a) with reconstructed pseudo-tracklines (b) for the humpback whale breeding season 2012 in the South Lagoon of New Caledonia. Maps of effort (c) and pseudo-effort (d) show the density surface of search effort intensity over a colour scale (darker shades indicate more effort).

3- Evaluating the quality of pseudo-effort maps

From 2003 to 2017, the pseudo-effort maps could be compared to the real effort maps obtained from GPS tracklines recorded in the field. On average, the 95% contour of pseudo-effort density surfaces overlaped with 59 % of the 95% contour of real effort density surfaces (Fig. S1.d).

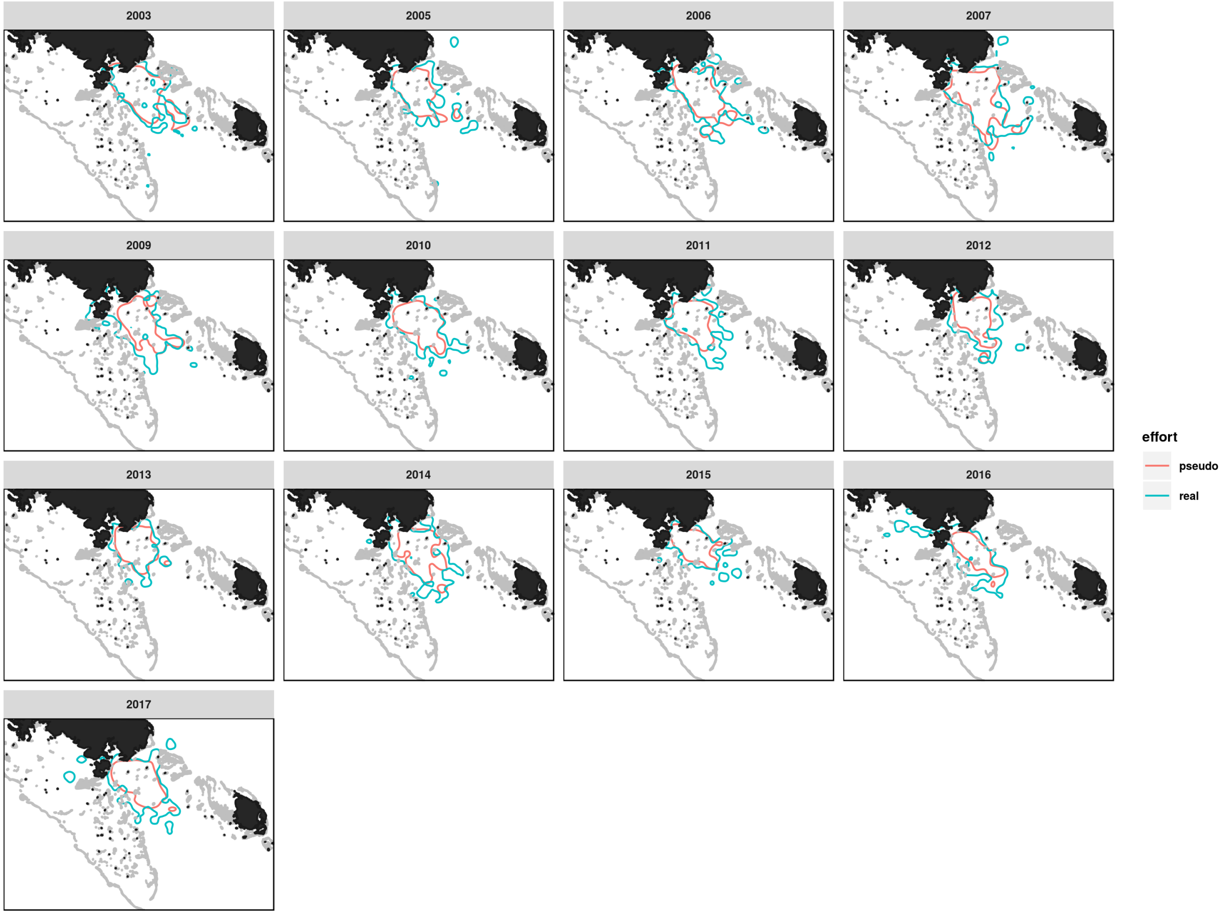

Figure S1.d: Contours of the 95% density surface calculated from real effort (GPS tracklines, blue) and pseudo-effort (acoustic sampling/focal follow positions, pink) from 2003 to 2017 in the South Lagoon of New Caledonia.