**NWASC Segmentation Instructions**

07/25/2016

This document contains instructions for segmenting data in the Northwest Atlantic Seabird Catalog.

Latest R version: 3.3.1

Required packages and latest versions: **CTS and DTS**;CTS only

broom: 0.4.1

**dplyr: 0.5.0**

**geosphere: 1.5-5**

**lubridate: 1.5.6**

**maptools: 0.8-39**

**rgeos: 0.3-19**

tibble: 1.1

**tidyr: 0.5.1**

zoo: 1.7-13

Required files and locations:

1. CTS
2. Q:/Kyle\_Working\_Folder/Segmentation/Data/database\_extract\_cts\_obs.RData

* Contains observation table and transect table in form of R objects ‘cts.dat’ and ‘obs.dat’
* Extracted from NWASC on 06/21/16 with the following script: Q:/Kyle\_Working\_Folder/Segmentation/From\_Arliss/Export\_R\_workspace\_for\_Kyle.R
* Required variables:
  1. cts.dat: **start\_dt, end\_dt, survey\_method\_cd, survey\_type\_cd, dataset\_id, source\_dataset\_id, transect\_id, segmented\_transect\_id, start\_tm, end\_tm, transect\_width\_nb, time\_from\_midnight\_start, time\_from\_midnight\_stop**
  2. obs.dat: **obs\_dt, obs\_start\_tm, dataset\_id, transect\_id, st\_astext, spp\_cd obs\_count\_intrans\_nb, time\_from\_midnight**

Note: st\_astext column is not necessary in ‘cts.dat’. Spatial information is taken from file in step 2 below. Only columns necessary for segmentation are included in tables.

1. Q:/Kyle\_Working\_Folder/ Segmentation/Data/transect\_shp\_files/line

* Contains effort shapefile with all multipart spatial line geometries
* Extracted from NWASC on 03/07/16

Note: only variables necessary in data attribute are ‘**transect\_i**’ and ‘**dataset\_id**’

1. R scripts:
   1. Pre-segmentation: Q:/Kyle\_Working\_Folder/ Segmentation/pre\_seg\_new.R
   2. Segmentation: Q:/Kyle\_Working\_Folder/ Segmentation/seg\_new.R
2. DTS
3. Q:/Kyle\_Working\_Folder/Segmentation/Data/database\_extract\_dts\_obs.RData

* Contains observation table and transect table in form of R objects ‘dts.dat’ and ‘obs.dat’
* Extracted from NWASC on 06/21/16 with the following script: Q:/Kyle\_Working\_Folder/Segmentation/From\_Arliss/Export\_R\_workspace\_for\_Kyle.R
* Required variables:
  1. dts.dat: **start\_dt, end\_dt, survey\_method\_cd, survey\_type\_cd, dataset\_id, source\_dataset\_id, transect\_id, segmented\_transect\_id, start\_tm, end\_tm, transect\_time\_min\_nb, traversal\_speed\_nb, transect\_width\_nb, heading\_tx, st\_astext, time\_from\_midnight\_start, time\_from\_midnight\_stop**
  2. obs.dat: Same as in CTS above

1. R script: Q:/Kyle\_Working\_Folder/Segmentation/process\_dts.R

Instructions:

1. CTS
2. Open pre-segmentation script and change path names for items 1 and 2 above as necessary.
3. Source pre-segmentation script (approx. run time: 10 min)

e.g., source(“Q:/Kyle\_Working\_Folder/ Segmentation/pre\_seg\_new.R”)

* This creates objects ‘**obs.pre**’ (variable names: **long, lat, transect\_id, spp\_cd, count**) and ‘**shp.pre**’ (variable names: **long, lat, piece, order, dataset\_id, transect\_id**) in the global environment to be used by the segmentation function.

1. Open segmentation script and run function (approx. run time: 20 min)

e.g., seg.dat.cts = segmentCTS(obs.pre, shp.pre, cts.dat)

* See comments in script header for instructions on optional arguments

This produces a wide-form data frame with column names: **source\_dataset\_id, segmented\_transect\_id, transect\_id, seg\_num, start\_dt, seg\_dist, transect\_width\_nb, mid\_long, mid\_lat, survey\_type\_cd, survey\_method\_cd**

1. DTS
2. Open script and change path name for item 1 above as necessary.
3. Run function (approx. run time: under 1 min)

e.g., seg.dat.dts = segmentDTS(obs.dat, dts.dat)

* See comments in script header for instructions on optional arguments

This produces a wide-form data frame with column names: **source\_dataset\_id, segmented\_transect\_id, transect\_id, start\_dt, seg\_dist, transect\_width\_nb, mid\_long, mid\_lat, survey\_type\_cd, survey\_method\_cd**

1. Objects seg.dat.cts and seg.dat.dts can be combined into a single data frame with code at end of “process\_dts.R” script.