

Albacore Diet Synthesis C

Natasha Hardy

15/04/2021

Trait variation in albacore diet across geographies

About

This document contains code for data manipulation, refinement and conversion to presence absence, as well as conducting multivariate community analyses and traitglm on albacore diets across ocean basins, using habitat association, gregariousness traits and cluster # output from cluster analyses.

Blueprint for paragraphs:

1. (Results) Describe the dominant significant traits and non-sig traits in relation to albacore age and geography sampled (Fig. R3).
2. Relate to clusters (potential additional analysis/figure).
3. Discussion - Investigate how species traits may be used to explain variation in prey use by albacore tuna using a multi-matrix approach.

Current Issues / Limitations to discuss in paper

- We do currently use community data that includes single occurrence species, this means species that only occurred once in albacore diets across the world. As these are also associated with trait information and we aim to use traits to explain differences in communities, we believe this isn't problematic, but worth noting. In a traditional multivariate analysis, single occurrence species would inflate differences between communities that are analysed on a species-only basis, whilst our sampling (a.k.a. tuna foraging) does not adequately pick up these rarer species.
- The data are longer (225 obs) than they are wide (137 species), but only just. There are up to 30 species that only occurred once, and that would also cost us about 30 observations. Could be worth re-running models excluding single occurrence to check results.

Workspace

Data (L)

Load frequency of occurrence data.

```
# Upload + manipulate Species & Covar DF Doesn't like the here() function now
# for some reason
alb_adult_fo = read.csv(paste(here("./data/output_data/alb_global_wide_dietfo.csv"),
  sep = ""), check.names = FALSE, row.names = 1) %>%
  dplyr::select(StudyID:ncol(.), -grouped_id) #-c(, `Diplospinus multistriatus`, `Sardina pilchardus`

print(dim(alb_adult_fo)) #225 150
```

```
## [1] 225 150
```

Manipulate and subset data. Here we obtain 137 species as columns and 225 obs.

```
# Check for empty columns
which(colSums(alb_adult_fo[, 14:ncol(alb_adult_fo)]) == 0) #None! #Otherwise remove these
```

```
## named integer(0)
```

```
# Check for empty rows
which(rowSums(alb_adult_fo[, 14:ncol(alb_adult_fo)]) == 0) #Remove these - this becomes a thing once w
```

```
## named integer(0)
```

```
# None
```

```
# Subset species df
alb_adult_prej = alb_adult_fo[, 14:ncol(alb_adult_fo)]
names(alb_adult_prej) <- str_replace_all(names(alb_adult_prej), " ", "_")
```

```
# check df
dim(alb_adult_prej) #137 species & 225 observations
```

```
## [1] 225 137
```

```
# str(alb_adult_prej) #data are integers & numeric
```

```
# Round to no decimal places if needed alb_adult_prej[,1:ncol(alb_adult_prej)]
# = round(alb_adult_prej[,1:ncol(alb_adult_prej)], digits = 1) Convert to
# integer values alb_adult_prej[,1:ncol(alb_adult_prej)] =
# lapply(alb_adult_prej[,1:ncol(alb_adult_prej)], as.integer) Try numeric
alb_adult_prej[, 1:ncol(alb_adult_prej)] = lapply(alb_adult_prej[, 1:ncol(alb_adult_prej)],
  as.numeric)
```

Convert data to presence absence and we obtain the following observations and species:

```
# Convert to presence/absence
alb_adult_prejPA = alb_adult_prej
alb_adult_prejPA[alb_adult_prejPA > 0] = 1
alb_adult_prejPA_df = cbind(alb_adult_fo[, 1:13], alb_adult_prejPA)

dim(alb_adult_prejPA) #225 obs, 137 species
```

```
## [1] 225 137
```

Traits PROBABLE (Q)

- Using all the occurrence data in 'alb_adult_preypA' df.
- Could filter out low occurrences using 'alb_adult_preypA_clean' df, extra code to do this is provided in the 'Albacore_synthesis_c_extra.Rmd'.
- Note that we obtain 135 species and their traits, so we'll need to filter the species df by the traits.

```
alb_prob_clust = read.csv(paste(here("./data/output_data/prob.prey.clusternum_habgreg.divis.k7.csv"),
  sep = ""), header = T, row.names = 1) %>%
  dplyr::select(preyp_sp:prob.clust.num, -refuge_cat)
```

```
dim(alb_prob_clust) #we obtain 156 species, 17 traits or covariates
```

```
## [1] 156 17
```

```
# Relabel to add underscore in names
alb_prob_clust$preyp_sp = str_replace_all(alb_prob_clust$preyp_sp, " ", "_")
# Assign species as row names
rownames(alb_prob_clust) <- alb_prob_clust$preyp_sp
# Order species list
alb_prob_clust <- alb_prob_clust[order(rownames(alb_prob_clust)), ]

# Ensure traits are treated as factors for now
alb_prob_clust[, 2:8] <- lapply(alb_prob_clust[, 2:8], factor) #to convert to factors

# Need to filter by prey species we want to use.
alb_prob_traits = alb_prob_clust %>%
  filter(preyp_sp %in% names(alb_adult_preypA)) %>%
  filter(life_stage != "larva")

# we get 123 species and 17 trait values previously filtered from this df:
# alb_adult_preypA - this includes single occurrences filtered out low
# occurrences df here: alb_adult_preypA_clean

dim(alb_prob_traits) #123 17
```

```
## [1] 123 17
```

```
str(alb_prob_traits)
```

```
## 'data.frame': 123 obs. of 17 variables:
## $ preyp_sp : chr "Acanthephyra_pelagica" "Alpheus_glaber" "Anchylomera_blossevillei" "Ancis
## $ life_stage : Factor w/ 3 levels "adult","juvenile",...: 1 1 1 2 2 1 1 1 1 1 ...
## $ vert_habitat : Factor w/ 5 levels "bathypelagic",...: 5 2 4 3 4 5 4 4 2 4 ...
## $ horz_habitat : Factor w/ 5 levels "coastal","continental shelf",...: 4 1 2 2 2 4 4 4 2 4 ...
## $ diel_migrant_cat: Factor w/ 2 levels "diel (no)","diel (yes)": 2 2 2 2 1 2 1 1 2 2 ...
## $ season_cat : Factor w/ 2 levels "season (no)",...: 2 1 2 2 2 2 2 2 1 ...
## $ body_shape : Factor w/ 6 levels "compressiform",...: 3 3 4 4 4 3 1 1 2 1 ...
## $ phys_defense : Factor w/ 2 levels "0","1": 2 2 2 1 2 1 2 2 1 2 ...
## $ transparent : int 0 0 0 0 0 0 0 0 1 ...
## $ col_disrupt : int 0 0 0 1 1 0 1 1 1 0 ...
```

```
## $ silver      : int  0 0 0 0 0 1 0 0 0 0 ...
## $ countershade : int  0 0 0 0 1 0 0 0 0 0 ...
## $ gregarious  : chr  "schooling" "schooling" "schooling" "schooling" ...
## $ maxFO       : num  38.9 11.1 4.2 33 33.3 ...
## $ maxN        : num  0 0 1.65 4.9 0 4 2.37 0 0 0 ...
## $ maxM        : num  0 0 3.3 0.71 0 0 1.6 0.1 0 0 ...
## $ prob.clust.num : int  1 2 3 3 4 1 4 4 3 2 ...
```

Trait value checks

Here we check the trait values for each trait, and check the distribution of values. Note we merge certain values with low occurrences of species within them, and if we have a logical reason to merge.

```
summary(alb_prob_traits$life_stage) #we will remove larva as a value from the df, larval taxa didn't make it
```

```
##      adult juvenile      larva
##      60         63         0
```

```
summary(alb_prob_traits$vert_habitat) #merge bathypelagic and mesopelagic, these prey could not have been seen
```

```
## bathypelagic      benthic      demersal      epipelagic      mesopelagic
##           2           10           17           53           41
```

```
summary(alb_prob_traits$horz_habitat) #merge of continental shelf/slope, and coastal/reef assoc, we do not have enough data to keep them separate
```

```
##           coastal continental shelf continental slope           oceanic
##           16           41           6           57
## reef-associated
##           3
```

```
summary(alb_prob_traits$diel_migrant_cat) #a lot more diel migrants than not
```

```
## diel (no) diel (yes)
##      27      96
```

```
summary(alb_prob_traits$season_cat) #more seasonal taxa than not
```

```
## season (no) season (yes)
##      33      90
```

```
summary(as.factor(alb_prob_traits$gregarious)) #still more gregarious than not, just an artefact of the data
```

```
## schooling solitary
##      109      14
```

```
summary(as.factor(alb_prob_traits$prob.clust.num)) #good distribution of species within clusters
```

```
## 1  2  3  4  5  6  7
## 27 24 24 18  8 13  9
```

Cleaning and preparing trait values for analyses

```

#levels(alb_prob_traits$life_stage) #check
#alb_prob_traits$life_stage <- factor(alb_prob_traits$life_stage,
#                                     levels = c("adult", "juvenile", "larva"),
#                                     labels = c("adult", "juvenile", "juvenile"))

#Levels vertical habitat
#levels(alb_prob_traits$vert_habitat) #check
alb_prob_traits$vert_habitat <- factor(alb_prob_traits$vert_habitat,
                                     levels = c("benthic", "demersal", "epipelagic",
                                                "mesopelagic", "bathypelagic"),
                                     labels = c("benthic", "demersal", "epipelagic",
                                                "mesopelagic", "mesopelagic"))

#Too low sample size for bathypelagic, merged with mesopelagic here

#Levels horizontal habitat
#levels(alb_prob_traits$horz_habitat) #check
alb_prob_traits$horz_habitat <- factor(alb_prob_traits$horz_habitat,
                                     levels = c("reef-associated", "coastal",
                                                "continental shelf", "continental slope",
                                                "oceanic"),
                                     labels = c("coastal", "coastal",
                                                "continental shelf", "continental shelf",
                                                "oceanic"))

#Levels gregarious
#levels(alb_adult_traits$gregarious) #check
alb_prob_traits$gregarious <- factor(alb_prob_traits$gregarious,
                                     levels = c("solitary", "schooling"))

write.csv(alb_prob_traits, here("../data/output_data/alb_prob_traits_k7.csv"))

```

UPDATE DIET DATA (L) & SELECT COVARS (R)

Here we filter our prey species (presence/absence data) by the species that were also used for cluster analysis, because we have complete trait information for these species, as well as frequency of occurrence data.

```

alb_prob_preypA <- alb_adult_preypA[colnames(alb_adult_preypA) %in% rownames(alb_prob_traits)]

dim(alb_prob_preypA)  #225 123

```

```
## [1] 225 123
```

```
# alternatively filter by stricter dataset
```

Now we need to check for and remove any zero-sum columns or rows of data, and adjust the co-variate data.

```

# Need to check for empty columns and rows again Check for empty columns
which(colSums(alb_prob_preypA) == 0)  #None! #Otherwise remove these

```

```
## named integer(0)
```

```
# Check for empty rows
which(rowSums(alb_prob_preypA) == 0) #Remove these - this becomes a thing once we start removing outliers
```

```
## 31 32 58 64 69 72 79 81 86 90 102 119 140 141 151 156
## 31 32 58 64 69 72 79 81 86 90 102 119 140 141 151 156
```

```
# Need to remove these from the alb_prob_preypA data and the covars data
```

```
alb_prob_preypA = alb_prob_preypA[-c(31, 32, 58, 64, 69, 72, 79, 81, 86, 90, 102,
119, 140, 141, 151, 156), ]
# 209 observations for 123 prey species
dim(alb_prob_preypA) #209 123
```

```
## [1] 209 123
```

```
alb_prob_covars = alb_adult_fo[-c(31, 32, 58, 64, 69, 72, 79, 81, 86, 90, 102, 119,
140, 141, 151, 156), 1:13]
# 209 observations for 13 environmental variables
```

```
dim(alb_prob_covars) #209 13
```

```
## [1] 209 13
```

Covars (R) - MANIPS

Checking co-variables for analyses.

```
#Summary stats on covars
unique(alb_prob_covars$StudyID) #22 studies
unique(alb_prob_covars$OceanBasin) #6 Ocean basins --> "N Atlantic" "Mediterranean" "N Pacific"
unique(alb_adult_fo$OceanBasinQ) #10 Ocean basin quarters, but data gets thin for many
unique(alb_prob_covars$YearEnd) #~27 years (or groups of years) are represented in this study
unique(alb_prob_covars$code) #10 Longhurst provinces represented, similar to Ocean Basin Q but also get
unique(alb_prob_covars$pred_life) #Need to clean these up to --> adult, juvenile and mixed cohorts.
```

Cleaning and manipulating co-variables for analyses. Any study for which we weren't able to estimate albacore life stage or size, either based on location or fork length data, we will consider to be of 'mixed' cohort.

```
#Clean covars
alb_prob_covars = alb_prob_covars %>%
  mutate(pred_life=case_when(pred_life == "juvenile" ~ "juvenile",
                             pred_life == "juvenile, adult" ~ "mixed",
                             pred_life == "adult" ~ "adult",
                             pred_life == "none" ~ "mixed")) #"none" ~ "none"

#As factors
alb_prob_covars[,1:ncol(alb_prob_covars)] = lapply(alb_prob_covars[,1:ncol(alb_prob_covars)]), as.factor)

dim(alb_prob_covars) #209 13
```

```
## [1] 209 13
```

Checking the level replication for each covariate. Note that our best covars ==> pred_life, OceanBasin, lat_cat, and maybe code (Longhurst) - but would need to exclude low numbered levels.

```
#Check representation
#summary(alb_adult_covars)

summary(alb_prob_covars$lat_cat)
#temperate tropical
# 178      31

summary(alb_prob_covars$pred_life)
# adult juvenile mixed
# 54      62      93

summary(alb_prob_covars$OceanBasin)
# Indian Mediterranean N Atlantic N Pacific S Atlantic S Pacific
# 15      29      78      62      8      17

summary(alb_prob_covars$OceanBasinQ)
#Not going to use

summary(alb_prob_covars$code)
#Not going to use

dim(alb_prob_covars) #209 observations for 13 covars
```

DF checks

```
## Make sure species names match abund matrix
#rownames(alb_prob_traits) == colnames(alb_prob_preypa) #awesome!
# And the mvabund matrix
alb_prob_preypa_mv = as.mvabund(alb_prob_preypa)

dim(alb_prob_preypa_mv) #209 123

#rownames(alb_prob_traits) == colnames(alb_prob_preypa_mv) #awesome!

#Check all matrix dimensions
dim(alb_prob_traits) #123 species and 17 traits
dim(alb_prob_preypa_mv) #209 obs, 123 species
dim(alb_prob_covars) #209 obs, 13 variables
```

SAVE DFS FOR ANALYSES

```
#L - species df
write.csv(alb_prob_preypa, here("./data/output_data/alb_prob_preypa_use.csv"))

#R - covars df
write.csv(alb_prob_covars, here("./data/output_data/alb_prob_covars_use.csv"))
```

```
#Q - trait df  
write.csv(alb_prob_traits, here("./data/output_data/alb_prob_traits_use.csv"))
```