Single Species Example from Gulf of Alaska with Catchability Covariates RACE Bottom Trawl Survey

Dr. Curry J. Cunningham

January 23, 2018

Contents

1 Purpose

The purpose of this document is to describe how to generate a model-based index of abundance unsing the spatio-temporal delta-GLMM in the VAST package, while incorporating **catchability covariates**.

1.1 Background

Catchability covariates attempt to explain residual variance in either the **encounter probability** or **positive catch rate** components of the delta model, given conditions influencing observation uncertainty at the time of sampling. In this way **catchability** covariates which are represented at the haul level, may be contrasted with **density** covariates which attempt to explain the underlying spatial distribution of the observed species and are represented at the knot level.

1.1.1 Example notation

Within the delta-model, the linear predictor for encounter probability can be written as:

$$p_1(i) = \beta_1(c_i, t_i) + \sum_{f=1}^{n_{\omega 1}} L_{\omega 1}(c_i, f)\omega 1(s_i, f) + \sum_{f=1}^{n_{\epsilon 1}} L_{\epsilon 1}(c_i, f)\epsilon_1(s_i, f, t_i)$$
$$+ \sum_{f=1}^{n_{\delta 1}} L_{\delta 1}(v_i, f)\delta_1(v_i, f) + \sum_{p=1}^{n_p} \gamma_1(c_i, t_i, p)X(x_i, t_i, p) + \sum_{k=1}^{n_k} \lambda(k)Q(i, k)$$

where $p_1(i)$ is the predicotr for observation i, $Q(i,k)$ are measured catchability covariates that explain variation in catchability and $\lambda_1(k)$ is the estimated impact of catchability covariates for this linear predictor, and $X(x_i, t_i, p)$ are measured **density** covariates that explain variation in density and $\gamma_1(c_i, t_i, p)$ is the estimated impact of density covariates.

1.2 Hypotheses

We will be testing two hypotheses: * Tow duration (measured in hours) influences the catchability of species in the survey, under the assumption that longer tows should increase encounter probability and/or the positive catch rate if individuals outswim the survey net but tire over time. + Note: Trawl distance (positively correlated with encounter probability) is already accounted for when effort is calculated as area swept (square km). * Measured gear temperature influences catchability, due to potential shifts in the vertical distribution of species.

Specifics of this Example:

- Uses RACE bottom trawl survey data.
  - Data are available from the /data folder
- Single species implementation.
- Gulf of Alaska survey data.
- Haul-level catchability covariates: (1) tow duration (hours), (2) gear temperature.

---

2 Setup

2.1 Install required packages

```
devtools::install_github("nwfsc-assess/geostatistical_delta-GLMM")
devtools::install_github("james-thorson/VAST")
devtools::install_github("james-thorson/utilities")
```

2.2 Load required packages

```
require(dplyr)
require(VAST)
require(TMB)
require(FishData)
# require(tidyverse)
```

2.3 Setup model

2.3.1 Define species of interest (based on species code) and survey name.

Species are selected by defining the vector `species.codes` in combination with the `combineSpecies` variable. While most species will have a single species code, there are some examples (i.e. GOA Dusky Rockfish) that require multiple species codes to be combined for a single species index. In this later case `combineSpecies = FALSE` would be specified.

Here are some examples to choose from:

number	name	species.code	include	survey	Region
1	Pacific ocean perch	30060	Y	GOA	Gulf_of_Alaska
2	Pacific ocean perch	30060	Y	AI	Aleutian_Islands
3	Walleye pollock	21740	Y	GOA	Gulf_of_Alaska
4	Walleye pollock	21740	Y	AI	Aleutian_Islands
5	Pacific cod	21720	Y	GOA	Gulf_of_Alaska
6	Pacific cod	21720	N	EBS_SHELF	Eastern_Bering_Sea
7	Pacific cod	21720	Y	AI	Aleutian_Islands
8	Northern rockfish	30420	Y	GOA	Gulf_of_Alaska
9	Northern rockfish	30420	Y	AI	Aleutian_Islands
10	Dover sole	10180	Y	GOA	Gulf_of_Alaska
11	Big skate	420	Y	GOA	Gulf_of_Alaska
12	Atka mackerel	21921	Y	AI	Aleutian_Islands
13	Harlequin rockfish	30535	Y	GOA	Gulf_of_Alaska
14	Arrowtooth flounder	10110	Y	GOA	Gulf_of_Alaska
15	Arrowtooth flounder	10110	Y	EBS_SHELF	Eastern_Bering_Sea
16	Spiny dogfish	310	Y	GOA	Gulf_of_Alaska

Example: Pacific Cod in the Gulf of Alaska

```
species.codes = c(21720)
survey = "GOA"
combineSpecies = FALSE
```

```
if(survey=="GOA") { Region = 'Gulf_of_Alaska' }
if(survey=="EBS_SHELF") { Region = "Eastern_Bering_Sea" }
if(survey=="AI") { Region = "Aleutian Islands" }
```

2.3.2 Observation reference location settings

```
lat_lon.def = "start"
```

2.3.3 Spatial settings

The following settings define the spatial resolution for the model (defined by number of knots `n_x`), and whether to use a grid or mesh approximation through the `Method` variable.

```
Method = c("Grid", "Mesh", "Spherical_mesh")[2]
grid_size_km = 25
n_x = c(100, 250, 500, 1000, 2000)[1]
Kmeans_Config = list( "randomseed"=1, "nstart"=100, "iter.max"=1e3 )

#Strata Limits
#Basic - Single Area
strata.limits = data.frame(STRATA = c("All_areas"))

#VAST Version - latest!
Version = "VAST_v4_0_0"
```

2.3.4 Model settings

```
bias.correct = FALSE

FieldConfig = c(Omega1 = 1, Epsilon1 = 1, Omega2 = 1, Epsilon2 = 1)
RhoConfig  = c(Beta1 = 0, Beta2 = 0, Epsilon1 = 0, Epsilon2 = 0)
OverdispersionConfig  = c(Delta1 = 0, Delta2 = 0)

#Observation Model
ObsModel = c(1,0)
```

2.3.5 Save settings

DateFile is the folder that will hold my model outputs.

```
DateFile = paste0(getwd(), "/VAST_output/")

#Create directory
dir.create(DateFile, recursive=TRUE)
```

2.4 Specify model outputs

The following settings define what types of output we want to calculate.

```
Options = c(SD_site_density = 0, SD_site_logdensity = 0,
            Calculate_Range = 1, Calculate_evenness = 0, Calculate_effective_area = 1,
            Calculate_Cov_SE = 0, Calculate_Synchrony = 0,
            Calculate_Coherence = 0)
```

---

3 Prepare the data

- Note: This section can be replace by function `create_VAST_input()`, from `R/create-VAST-input.r`

3.1 Load RACE data

To create the input data files for VAST model, first we must load RACE survey data. Two data files are necessary **(1)** catch data and **(2)** haul data.

3.1.1 Load and join data

Catch data

```
catch = readRDS("data/race_base_catch.rds")

haul = readRDS("data/race_base_haul.rds")
haul = haul[haul$ABUNDANCE_HAUL == "Y", ]

# Join datasets
catchhaul = right_join(x = catch, y = haul, by = c("HAULJOIN"))

# Add in zero observations for catch weight, for no
# catches.
catchhaul.2 = FishData::add_missing_zeros(data_frame = catchhaul,
    unique_sample_ID_colname = "HAULJOIN", sample_colname = "WEIGHT",
    species_colname = "SPECIES_CODE", species_subset = species.codes,
    if_multiple_records = "First", Method = "Fast")
```

```
## Species to include: 21720

## Number of samples to include for each species: 31306

## Finished processing for 21720
```

```
# Load and attach cruise info
cruise.info = read.csv("data/race_cruise_info.csv",
    header = TRUE, stringsAsFactors = FALSE)

catchhaul.3 = inner_join(x = catchhaul.2, y = cruise.info[,
    c("Cruise.Join.ID", "Year", "Survey")], by = c(CRUISEJOIN.x = "Cruise.Join.ID"))

# Limit to survey of interest
catchhaul.3 = catchhaul.3[catchhaul.3$Survey == survey,
    ]

# Aggregate multiple `species.codes`, if we are
# combining into a single index.
if (combineSpecies == TRUE) {
    catchhaul.4 = data.frame(catchhaul.3 %>% group_by(HAULJOIN) %>%
        mutate(WEIGHT = sum(WEIGHT, na.rm = TRUE)))
    # Since we have aggregated, only retain rows for
    # 1st listed species code
    catchhaul.5 = catchhaul.4[catchhaul.4$SPECIES_CODE ==
        species.codes[1], ]
```

```
} else {
    catchhaul.5 = catchhaul.3
}
```

3.2 Standardize data

In order to standardize the survey catch data, we must calculate effort as area swept per tow.

3.2.1 Calculate effort

Input effort is in $kilometers^2$

```
catchhaul.5$effort = catchhaul.5$NET_WIDTH*catchhaul.5$DISTANCE_FISHED/1000
```

3.3 Add species names

First, we load the list describing both species names and `species.codes`

```
species.code.data = read.csv("data/race_species_codes.csv",
                             header=TRUE, stringsAsFactors=FALSE)

#Next, join this to our overall survey data list
load.data = merge(x=catchhaul.5, y=species.code.data[,c("Species.Code","Common.Name")],
                        by.x="SPECIES_CODE", by.y="Species.Code")
```

3.4 Build `Data_Geostat`

Now, we will create the list `Data_Geostat` which is the input for the VAST model.

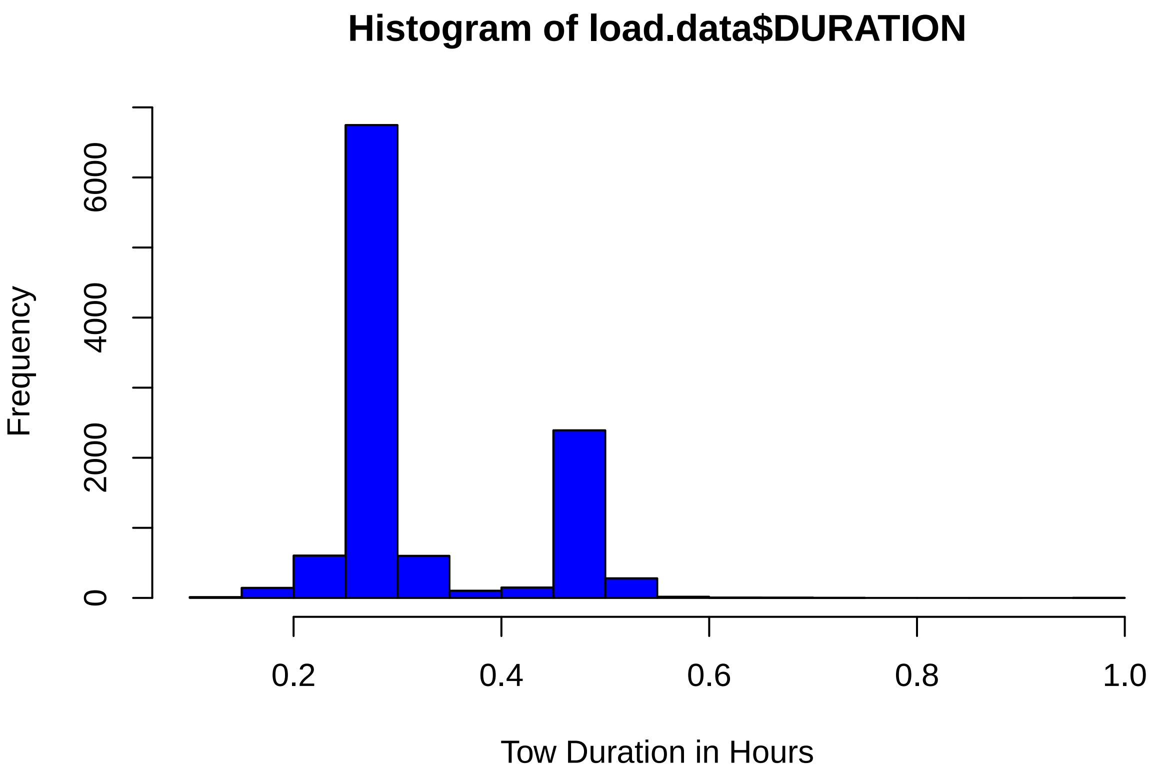
```
Data_Geostat = NULL

if(length(species.codes) > 1) {
  Data_Geostat$spp = load.data$Common.Name
}
Data_Geostat$Catch_KG = as.numeric(load.data$WEIGHT)
Data_Geostat$Year = as.integer(load.data$Year)
Data_Geostat$Vessel = "missing"
Data_Geostat$AreaSwept_km2 = as.numeric(load.data$effort)
Data_Geostat$Pass = 0
```

3.4.1 Add catchability covariates to `Data_Geostat`

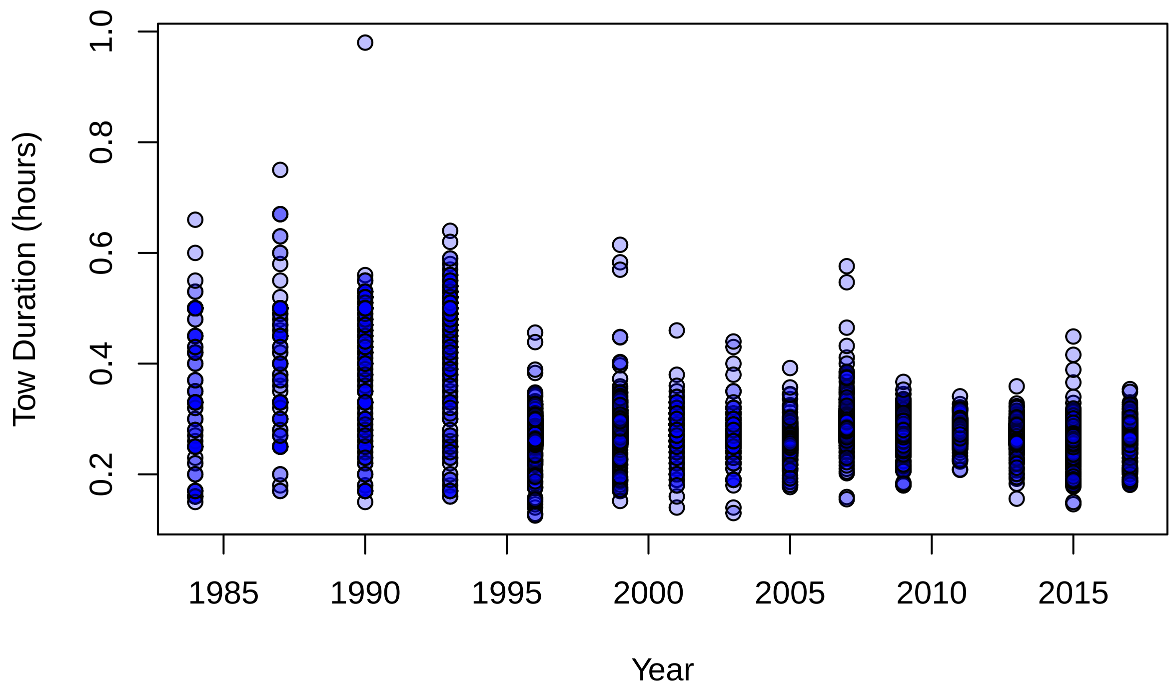
Here we can attach our two catchability covariates to our list of input data.

Hypothesis #1: Catchability is is influenced by tow duration in hours.

First, lets see what the distribution of tow durations look like…



So there seems to be a break down between 0.5 hour and 0.25 hour tows, but when did they occur?

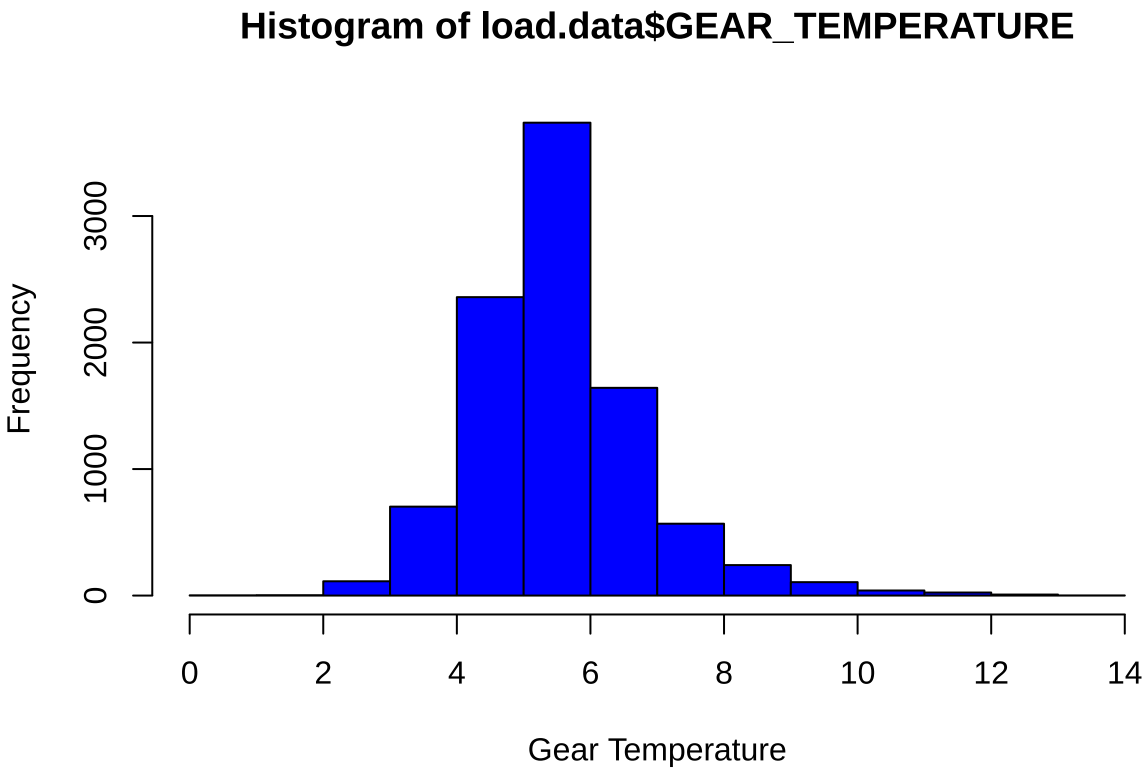


Ok lets attach these data to `Data_Geostat`

```
Data_Geostat$Duration <- as.numeric(load.data$DURATION)
```

Hypothesis #2: Catchability is is influenced by gear temperature.

Again, lets see what the distribution of gear temperatures look like. . .



And lets add this to `Data_Geostat`

```
Data_Geostat$Gear_Temperature <- as.numeric(load.data$GEAR_TEMPERATURE)
```

3.4.2 Define location of samples

Depending on `lat_lon.def` specification we will either use the **start**, **end**, or **mean** location recorded for a survey haul.

- Note: Using the starting location of each haul is probably best, as: `lat_lon.def="start"`.

```
if(lat_lon.def=="start") {
  Data_Geostat$Lat = load.data$START_LATITUDE
  Data_Geostat$Lon = load.data$START_LONGITUDE
}
if(lat_lon.def=="end") {
  Data_Geostat$Lat = load.data$END_LATITUDE
  Data_Geostat$Lon = load.data$END_LONGITUDE
}
if(lat_lon.def=="mean") {
  Data_Geostat$Lat = rowMeans(cbind(load.data$START_LATITUDE,
                                   load.data$END_LATITUDE), na.rm=TRUE)

  Data_Geostat$Lon = rowMeans(cbind(load.data$START_LONGITUDE,
                                   load.data$END_LONGITUDE), na.rm=TRUE)
}
```

Next, we ensure this `Data_Geostat` is a proper data frame.

```
Data_Geostat = data.frame(Data_Geostat)
```

To double check lets see how `Data_Geostat` looks…

```
kable(head(Data_Geostat))
```

Catch_KG	Year	Vessel	AreaSwept_km2	Pass	Duration	Gear_Temperature	Lat	Lon
190.09	2005	missing	0.027696	0	0.281	NA	52.5579	-169.783
30.89	2005	missing	0.020099	0	0.252	4.9	52.6384	-169.781
242.18	2005	missing	0.024110	0	0.255	4.7	52.6713	-169.428
39.92	2005	missing	0.020601	0	0.266	5.2	53.2410	-168.072
16.18	2005	missing	0.019489	0	0.255	5.1	53.1677	-167.981
61.22	2005	missing	0.021170	0	0.252	4.9	53.0684	-167.671

3.5 Limit `Data_Geostat` to only tows with recorded `Gear_Temperature`

Upon closer inspection you will notice that some tows did not have a recorded `Gear_Temperature`, appearing as an `NA`. It appears to be ~ 13.7%

```
nrow(Data_Geostat[is.na(Data_Geostat$Gear_Temperature),])/nrow(Data_Geostat)*100
```

```
## [1] 13.6865
```

Given we want to compare models fit to the same data, lets remove these tows without `Gear_Temperature` observations.

```
Data_Geostat = Data_Geostat[!is.na(Data_Geostat$Gear_Temperature),]
```

3.6 Create the extrapolation grid

We also generate the extrapolation grid appropriate for a given region. For new regions, we use Region="Other".

- Note: We are not defining strata limits, but could do so based on latitude and longitude definitions.

```
Extrapolation_List = SpatialDeltaGLMM::Prepare_Extrapolation_Data_Fn(Region = Region,
    strata.limits = strata.limits)
```

3.7 Create spatial list

Next, generate the information used for conducting spatio-temporal parameter estimation, bundled in list `Spatial_List`.

```
Spatial_List = SpatialDeltaGLMM::Spatial_Information_Fn(grid_size_km = grid_size_km,
    n_x = n_x, Method = Method, Lon = Data_Geostat[,
        "Lon"], Lat = Data_Geostat[, "Lat"], Extrapolation_List = Extrapolation_List,
    randomseed = Kmeans_Config[["randomseed"]], nstart = Kmeans_Config[["nstart"]],
    iter.max = Kmeans_Config[["iter.max"]], DirPath = DateFile,
    Save_Results = TRUE)
```

3.8 Update `Data_Geostat` with knot references

We then associate each of our haul observations with its appropriate knot.

```
Data_Geostat = cbind(Data_Geostat, knot_i = Spatial_List$knot_i)
```

---

4 Build and run models

Here we are going to build **3** models.

- **Null Model** which does **not** estimate catchability covariates.
- **Model 1** testing Hypothesis #1 with **tow duration** as a catchability covariate.
- **Model 2** testing Hypothesis #2 with **temperature** as a catchability covariate.

4.1 Null Model

First, create a subdirectory for the **Null Model**

```
DateFile_null = paste0(DateFile,"Null/")
dir.create(DateFile_null)
```

Building and compiling:

- Note: in `Build_TMB_Fn()` whether to estimate **catchability** covariates is specified by the `Q_Config` argument, and whether to estimate **density** covariates by `CovConfig`.

```
# Build
if (length(species.codes) > 1 & combineSpecies == FALSE) {
    # MULTISPECIES
    TmbData_null = VAST::Data_Fn(Version = Version,
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = as.numeric(Data_Geostat[, "spp"]) - 1,
        b_i = Data_Geostat[, "Catch_KG"], a_i = Data_Geostat[,
            "AreaSwept_km2"], v_i = as.numeric(Data_Geostat[,
            "Vessel"]) - 1, s_i = Data_Geostat[, "knot_i"] -
            1, t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options)
} else {
    # SINGLE SPECIES
    TmbData_null = VAST::Data_Fn(Version = Version,
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = rep(0, nrow(Data_Geostat)), b_i = Data_Geostat[,
            "Catch_KG"], a_i = Data_Geostat[, "AreaSwept_km2"],
        v_i = as.numeric(Data_Geostat[, "Vessel"]) -
            1, s_i = Data_Geostat[, "knot_i"] - 1,
        t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options)
}

# Compile TMB object
TmbList_null = VAST::Build_TMB_Fn(TmbData = TmbData_null,
    RunDir = DateFile_null, Version = Version, RhoConfig = RhoConfig,
    loc_x = Spatial_List$loc_x, Method = Method, Q_Config = FALSE,
    CovConfig = FALSE)
Obj_null = TmbList_null[["Obj"]]
```

Fit VAST model to the data by optimizing the TMB function.

```r
Opt_null = TMBhelper::Optimize(obj = Obj_null, lower = TmbList_null[["Lower"]],
                          upper = TmbList_null[["Upper"]], getsd = TRUE,
                          savedir = DateFile_null,
                          bias.correct = bias.correct, newtonsteps=2)
```

Save outputs from estimation

```r
Report_null = Obj_null$report()

Save_null = list("Opt"=Opt_null, "Report"=Report_null,
                 "ParHat"=Obj_null$env$parList(Opt_null$par),
                 "TmbData"=TmbData_null)

save(Save_null, file=paste0(DateFile_null,"Save.RData"))
```

4.2 Model 1

First, create a subdirectory for the **Model 1**

```r
DateFile_Mod1 = paste0(DateFile,"Model 1/")
dir.create(DateFile_Mod1)
```

Building and compiling:

`Q_ik` is the argument to Data_Fn for catchability covariates.

- Note: `Q_ik` expects a matrix so we specify `Q_ik=as.matrix(Data_Geostat[,'Duration'])`

```r
# Build
if (length(species.codes) > 1 & combineSpecies == FALSE) {
    # MULTISPECIES
    TmbData_Mod1 = VAST::Data_Fn(Version = Version,
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = as.numeric(Data_Geostat[, "spp"]) - 1,
        b_i = Data_Geostat[, "Catch_KG"], a_i = Data_Geostat[,
            "AreaSwept_km2"], v_i = as.numeric(Data_Geostat[,
            "Vessel"]) - 1, s_i = Data_Geostat[, "knot_i"] -
            1, t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options,
        Q_ik = as.matrix(Data_Geostat[, "Duration"]))
} else {
    # SINGLE SPECIES
    TmbData_Mod1 = VAST::Data_Fn(Version = Version,
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = rep(0, nrow(Data_Geostat)), b_i = Data_Geostat[,
            "Catch_KG"], a_i = Data_Geostat[, "AreaSwept_km2"],
        v_i = as.numeric(Data_Geostat[, "Vessel"]) -
            1, s_i = Data_Geostat[, "knot_i"] - 1,
        t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options,
        Q_ik = as.matrix(Data_Geostat[, "Duration"]))
```

```r
}

# Compile TMB object
TmbList_Mod1 = VAST::Build_TMB_Fn(TmbData = TmbData_Mod1,
    RunDir = DateFile_Mod1, Version = Version, RhoConfig = RhoConfig,
    loc_x = Spatial_List$loc_x, Method = Method, Q_Config = TRUE,
    CovConfig = FALSE)
Obj_Mod1 = TmbList_Mod1[["Obj"]]
```

Fit VAST model to the data by optimizing the TMB function.

```r
Opt_Mod1 = TMBhelper::Optimize(obj = Obj_Mod1, lower = TmbList_Mod1[["Lower"]],
                          upper = TmbList_Mod1[["Upper"]], getsd = TRUE,
                          savedir = DateFile_Mod1,
                          bias.correct = bias.correct, newtonsteps=2)
```

Save outputs from estimation

```r
Report_Mod1 = Obj_Mod1$report()

Save_Mod1 = list("Opt"=Opt_Mod1, "Report"=Report_Mod1,
                 "ParHat"=Obj_Mod1$env$parList(Opt_Mod1$par),
                 "TmbData"=TmbData_Mod1)

save(Save_Mod1, file=paste0(DateFile_Mod1,"Save.RData"))
```

4.3 Model 2

First, create a subdirectory for the **Model 2**

```r
DateFile_Mod2 = paste0(DateFile,"Model 2/")
dir.create(DateFile_Mod2)
```

Building and compiling:

`Q_ik` is the argument to Data_Fn for catchability covariates.

- Note: `Q_ik` expects a matrix so we specify `Q_ik=as.matrix(Data_Geostat[,'Gear_Temperature'])`

```r
# Build
if (length(species.codes) > 1 & combineSpecies == FALSE) {
    # MULTISPECIES
    TmbData_Mod2 = VAST::Data_Fn(Version = Version,
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = as.numeric(Data_Geostat[, "spp"]) - 1,
        b_i = Data_Geostat[, "Catch_KG"], a_i = Data_Geostat[,
            "AreaSwept_km2"], v_i = as.numeric(Data_Geostat[,
            "Vessel"]) - 1, s_i = Data_Geostat[, "knot_i"] -
            1, t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options,
        Q_ik = as.matrix(Data_Geostat[, "Gear_Temperature"]))
} else {
    # SINGLE SPECIES
    TmbData_Mod2 = VAST::Data_Fn(Version = Version,
```

```
        FieldConfig = FieldConfig, OverdispersionConfig = OverdispersionConfig,
        RhoConfig = RhoConfig, ObsModel = ObsModel,
        c_i = rep(0, nrow(Data_Geostat)), b_i = Data_Geostat[,
            "Catch_KG"], a_i = Data_Geostat[, "AreaSwept_km2"],
        v_i = as.numeric(Data_Geostat[, "Vessel"]) -
            1, s_i = Data_Geostat[, "knot_i"] - 1,
        t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
        MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
        Method = Spatial_List$Method, Options = Options,
        Q_ik = as.matrix(Data_Geostat[, "Gear_Temperature"]))
}

# Compile TMB object
TmbList_Mod2 = VAST::Build_TMB_Fn(TmbData = TmbData_Mod2,
    RunDir = DateFile_Mod2, Version = Version, RhoConfig = RhoConfig,
    loc_x = Spatial_List$loc_x, Method = Method, Q_Config = TRUE,
    CovConfig = FALSE)
Obj_Mod2 = TmbList_Mod2[["Obj"]]
```

Fit VAST model to the data by optimizing the TMB function.

```
Opt_Mod2 = TMBhelper::Optimize(obj = Obj_Mod2, lower = TmbList_Mod2[["Lower"]],
                          upper = TmbList_Mod2[["Upper"]], getsd = TRUE,
                          savedir = DateFile_Mod2,
                          bias.correct = bias.correct, newtonsteps=2)
```

Save outputs from estimation

```
Report_Mod2 = Obj_Mod2$report()

Save_Mod2 = list("Opt"=Opt_Mod2, "Report"=Report_Mod2,
                 "ParHat"=Obj_Mod2$env$parList(Opt_Mod2$par),
                 "TmbData"=TmbData_Mod1)

save(Save_Mod2, file=paste0(DateFile_Mod2,"Save.RData"))
```

5 Compare Models

5.1 Check convergence

To evaluate convergence of our three candidate models we will look at convergence and the maximum gradient, which can both be accessed from the **Opt** object as **Opt**$convergence**and**Opt$**max_gradient**.

- Note: **Opt$convergence = 0** indicates relative convergence.

```
temp.table = NULL
temp.table$name = c('Null Model', 'Model 1', 'Model 2')
temp.table$effect = c('None','Tow Duration', 'Gear Temperature')
temp.table$convergence = c(Opt_null$convergence,
                           Opt_Mod1$convergence,
                           Opt_Mod2$convergence)
temp.table$max_gradient = c(Opt_null$max_gradient,
                            Opt_Mod1$max_gradient,
                            Opt_Mod2$max_gradient)
temp.table = data.frame(temp.table)
names(temp.table) = c('Model Name','Catchability Covariate','Convergence','Maximum Gradient')
#Print the table
# kable(temp.table)
pander::pandoc.table(temp.table)
```

```
##
## ----------------------------------------------------------------------
##  Model Name   Catchability Covariate   Convergence   Maximum Gradient
## ------------ ------------------------ ------------- ------------------
##  Null Model            None                 1           5.363e-08
##
##   Model 1          Tow Duration             1           4.406e-08
##
##   Model 2        Gear Temperature           1           2.296e-07
## ----------------------------------------------------------------------
```

5.2 Covariate effects

Now that we have fit these three alternative models and checked convergence, lets see what the estimates are for the effect of each covariate on catchability. We can access

5.2.1 Model 1 with tow duration

We can recall that the parameters describing the effect of **catchability** covariates on the encounter probability component of our delta-model is $\lambda_1(k)$, and the effect on the positive catch rate component is $\lambda_2(k)$.

So, in our mode output below we are intersted in:

- `lambda1_k` - Effect of tow duration on encounter probability.
- `lambda2_k` - Effect of tow duration on positive catch rate.

```
Opt_Mod1$SD
```

```
## sdreport(.) result
##                Estimate Std. Error
## ln_H_input    0.6484959 0.20259081
## ln_H_input    0.5634271 0.20730304
## beta1_ct      0.6140915 0.48733242
## beta1_ct      0.9318625 0.51231927
## beta1_ct      0.8132314 0.50142164
## beta1_ct      0.8128270 0.48973741
## beta1_ct      0.4739782 0.42532579
## beta1_ct     -0.0130992 0.42426490
## beta1_ct     -0.5695683 0.44196734
## beta1_ct     -0.0853260 0.42061392
## beta1_ct     -0.0439489 0.41884564
## beta1_ct     -0.0050696 0.42679197
## beta1_ct      0.3784831 0.42434129
## beta1_ct      0.3836690 0.42683946
## beta1_ct      0.3869232 0.42714117
## beta1_ct      0.3395083 0.42198412
## beta1_ct     -0.4581402 0.42663522
## lambda1_k     0.4918490 0.62780091
## L_omega1_z   -1.4661828 0.16984990
## L_epsilon1_z -0.5688669 0.06042288
## logkappa1    -4.4566656 0.12560931
## beta2_ct      7.4826175 0.34103275
## beta2_ct      7.4007719 0.35772046
## beta2_ct      7.4850346 0.34540474
## beta2_ct      7.6440895 0.34277878
## beta2_ct      7.5308517 0.25576914
## beta2_ct      7.3911293 0.25760257
## beta2_ct      7.0024922 0.27491334
## beta2_ct      7.1941383 0.25196179
## beta2_ct      7.1976865 0.24964853
## beta2_ct      7.1090461 0.25809140
## beta2_ct      7.5340791 0.25549983
## beta2_ct      7.5239343 0.25762247
## beta2_ct      7.6107924 0.25796148
## beta2_ct      7.2798694 0.25100725
## beta2_ct      6.9088770 0.26379554
## lambda2_k    -0.3634464 0.58631354
## L_omega2_z    0.5456585 0.07720787
## L_epsilon2_z  0.3362267 0.04953504
## logkappa2    -4.6092707 0.17616916
## logSigmaM     0.4572419 0.00950848
## Maximum gradient component: 4.40601e-08
```

5.2.2 Model 2 with gear temperature

For our second model we are interested in:

- `lambda1_k` - Effect of gear temperature on encounter probability.
- `lambda2_k` - Effect of gear temperature on positive catch rate.

```
Opt_Mod2$SD
```

```
## sdreport(.) result
##               Estimate Std. Error
## ln_H_input    0.642764 0.20830246
## ln_H_input    0.579099 0.21117917
## beta1_ct     -1.101358 0.43652851
## beta1_ct     -0.686850 0.45572067
## beta1_ct     -0.705480 0.45271481
## beta1_ct     -0.707586 0.42913526
## beta1_ct     -1.127522 0.42728392
## beta1_ct     -1.491912 0.42102165
## beta1_ct     -2.335991 0.45332344
## beta1_ct     -1.844865 0.43199617
## beta1_ct     -1.736169 0.42854331
## beta1_ct     -1.466366 0.42162928
## beta1_ct     -1.087527 0.42175477
## beta1_ct     -1.215576 0.42925125
## beta1_ct     -1.170181 0.42929678
## beta1_ct     -1.457564 0.43390295
## beta1_ct     -2.171516 0.43506321
## lambda1_k     0.310637 0.02984861
## L_omega1_z   -1.410517 0.16305215
## L_epsilon1_z  0.604803 0.06086523
## logkappa1    -4.497075 0.12051365
## beta2_ct      8.552497 0.22234111
## beta2_ct      8.454392 0.23777976
## beta2_ct      8.424576 0.23635046
## beta2_ct      8.615643 0.21244914
## beta2_ct      8.536293 0.21117141
## beta2_ct      8.286690 0.20433573
## beta2_ct      8.089147 0.23694018
## beta2_ct      8.327692 0.22223155
## beta2_ct      8.274297 0.21700185
## beta2_ct      7.998512 0.20347849
## beta2_ct      8.431237 0.20271388
## beta2_ct      8.517860 0.21236049
## beta2_ct      8.593922 0.21434468
## beta2_ct      8.419032 0.22072325
## beta2_ct      7.953298 0.22572708
## lambda2_k    -0.193015 0.02346359
## L_omega2_z   -0.507525 0.07026177
## L_epsilon2_z -0.343588 0.05022916
## logkappa2    -4.353333 0.19186672
## logSigmaM     0.450651 0.00957048
## Maximum gradient component: 2.29581e-07
```

5.3 Compare AIC across models

One way to compare across our candidate models with and without catchability covariates is to use AIC. The AIC for each model can be accessed from **Opt$AIC**.

Lets compare AIC across models…

```r
aic.table <- NULL
aic.table$name = c('Null Model', 'Model 1', 'Model 2')
aic.table$effect = c('None','Tow Duration', 'Gear Temperature')
aic.table$AIC = c(Opt_null$AIC, Opt_Mod1$AIC, Opt_Mod2$AIC)
#Calculate dAIC
aic.table$dAIC <- aic.table$AIC - min(aic.table$AIC)
#Data frame
aic.table <- data.frame(aic.table)
names(aic.table) <- c('Model Name', 'Catchability Covariate', 'AIC', 'dAIC')
#Print the table
kable(aic.table)
```

Model Name	Catchability Covariate	AIC	dAIC
Null Model	None	66056.0	178.225
Model 1	Tow Duration	66059.0	181.229
Model 2	Gear Temperature	65877.7	0.000

```r
# pander::pandoc.table(aic.table)
```

6 General conclusions

Here are some general conclusions regarding GOA Pacific Cod:

- For **Model 1** the effect of tow duration is highly uncertain with CV>1 estimated for effects of this covariate on both encounter probability `lambda1_k` and positive catch rate `lambda2_k`.
- For **Model 2** estimated catchability covariate effects have lower uncertainty
- Encounter probability is estimated to **increase** with gear temperature.
- Positive catch rate is estimated to **decrease** with gear temperature.
- It appears that **Model 2** which incorporates **gear temperature** as a **catchability** covariate provides a more parsimonious fit to the survey data.
- It should be noted that a Poisson-link delta-model may be a better way to correct for differences in tow duration.
- This may be specified with `ObsModel = c(1,1)`.

---

7 Specifying catchability covariates for positive catch rate only

Given you assumptions about the sampling process, it may make more sense to estimate the effect of the **catchability** covariates on the **positive catch rate** component of the delta-model only. To do so, we must:

- Modify the `Tmb_List_...` object by…
- Extracting the `$Map` object, as: `Map = TmbList$Map`
- Modifying `Map` to turn off estimation of `lambda1_k`, as: `Map[["lambda1_k"]] = rep(NA, length(TmbList$Parameters$lambda1_k))`
- Recompile the `TmbList_...` with `VAST::BuildTMB_Fn()`

As an example we will update, recompile, and re-fit **Model 1** and **Model 2**, so that the effect of **tow duration** and **gear temperature** is only linked to encounter probability. We will refer to these as **Model 1b** and **Model 2b**.

7.1 Model 1b: tow duration effect on encounter probability only

First, lets take the existing `TmbList_Mod1` from **Model 1** and extract `$Map`

```
Map_1b = TmbList_Mod1$Map
```

Next, lets turn off estimation of `lambda2_k`, lambda__k being the effect of **catchability** covariates, and **2** indicating this is for the *positive catch rate* component of the delta-model.

```
Map_1b[["lambda1_k"]] = factor(rep(NA, length(TmbList_Mod1$Parameters$lambda1_k)))
```

Finally, lets recompile the model.

- Note: We

```
#New file
DateFile_Mod1b = paste0(DateFile,"/Model 1b/")
dir.create(DateFile_Mod1b)
```

```
## Warning in dir.create(DateFile_Mod1b): '/Users/
## curryc2/Documents/NOAA/2017/VAST Evaluation/
## AFSC_Spatio-temporal_Workshop/testing/GOA_Bottom-
## Trawl_Single-Species_Catchability-Covar/
## VAST_output//Model 1b' already exists
```

```
#Recompile
TmbList_Mod1b = VAST::Build_TMB_Fn(TmbData = TmbData_Mod1, RunDir = DateFile_Mod1b,
                                Version = Version, RhoConfig = RhoConfig,
                                loc_x = Spatial_List$loc_x,
                                Method = Method, Q_Config=TRUE, CovConfig=FALSE, Map=Map_1b)
```

```
## Note: Using Makevars in /Users/curryc2/.R/Makevars

## List of estimated fixed and random effects:

##     Coefficient_name Number_of_coefficients
## 1           beta1_ct                     15
## 2           beta2_ct                     15
## 3        L_epsilon1_z                     1
## 4        L_epsilon2_z                     1
## 5          L_omega1_z                     1
## 6          L_omega2_z                     1
## 7           lambda2_k                     1
```

```
## 8         ln_H_input                     2
## 9          logkappa1                     1
## 10         logkappa2                     1
## 11         logSigmaM                     1
## 12 Epsiloninput1_sft                  3944
## 13 Epsiloninput2_sft                  3944
## 14    Omegainput1_sf                   116
## 15    Omegainput2_sf                   116
##      Type
## 1   Fixed
## 2   Fixed
## 3   Fixed
## 4   Fixed
## 5   Fixed
## 6   Fixed
## 7   Fixed
## 8   Fixed
## 9   Fixed
## 10  Fixed
## 11  Fixed
## 12 Random
## 13 Random
## 14 Random
## 15 Random
```

```
Obj_Mod1b = TmbList_Mod1b$Obj
```

Fit VAST model to the data by optimizing the TMB function.

```
Opt_Mod1b = TMBhelper::Optimize(obj = Obj_Mod1b, lower = TmbList_Mod1b[["Lower"]],
                          upper = TmbList_Mod1b[["Upper"]], getsd = TRUE,
                          savedir = DateFile_Mod1b,
                          bias.correct = bias.correct, newtonsteps=2)
```

Save outputs from estimation

```
Report_Mod1b = Obj_Mod1b$report()

Save_Mod1b = list("Opt"=Opt_Mod1b, "Report"=Report_Mod1b,
                 "ParHat"=Obj_Mod1b$env$parList(Opt_Mod1b$par),
                 "TmbData"=TmbData_Mod1)

save(Save_Mod1b, file=paste0(DateFile_Mod1b,"Save.RData"))
```

7.2 Model 2b: tow duration effect on encounter probability only

First, lets take the existing `TmbList_Mod2` from **Model 2** and extract `$Map`

```
Map_2b = TmbList_Mod2$Map
```

Next, lets turn off estimation of `lambda2_k`, lambda_k being the effect of **catchability** covariates, and **2** indicating this is for the *positive catch rate* component of the delta-model.

```
Map_2b[["lambda1_k"]] = factor(rep(NA, length(TmbList_Mod2$Parameters$lambda1_k)))
```

Finally, lets recompile the model.

```r
#New file
DateFile_Mod2b = paste0(DateFile,"/Model 2b/")
dir.create(DateFile_Mod2b)
```

```
## Warning in dir.create(DateFile_Mod2b): '/Users/
## curryc2/Documents/NOAA/2017/VAST Evaluation/
## AFSC_Spatio-temporal_Workshop/testing/GOA_Bottom-
## Trawl_Single-Species_Catchability-Covar/
## VAST_output//Model 2b' already exists
```

```r
#Recompile
TmbList_Mod2b = VAST::Build_TMB_Fn(TmbData = TmbData_Mod2, RunDir = DateFile_Mod2b,
                                Version = Version, RhoConfig = RhoConfig,
                                loc_x = Spatial_List$loc_x,
                                Method = Method, Q_Config=TRUE, CovConfig=FALSE, Map=Map_2b)
```

```
## Note: Using Makevars in /Users/curryc2/.R/Makevars
```

```
## List of estimated fixed and random effects:
```

```
##     Coefficient_name Number_of_coefficients
## 1           beta1_ct                     15
## 2           beta2_ct                     15
## 3       L_epsilon1_z                      1
## 4       L_epsilon2_z                      1
## 5         L_omega1_z                      1
## 6         L_omega2_z                      1
## 7          lambda2_k                      1
## 8         ln_H_input                      2
## 9          logkappa1                      1
## 10         logkappa2                      1
## 11         logSigmaM                      1
## 12 Epsiloninput1_sft                   3944
## 13 Epsiloninput2_sft                   3944
## 14    Omegainput1_sf                    116
## 15    Omegainput2_sf                    116
##      Type
## 1   Fixed
## 2   Fixed
## 3   Fixed
## 4   Fixed
## 5   Fixed
## 6   Fixed
## 7   Fixed
## 8   Fixed
## 9   Fixed
## 10  Fixed
## 11  Fixed
## 12 Random
## 13 Random
## 14 Random
## 15 Random
```

```r
Obj_Mod2b = TmbList_Mod2b$Obj
```

Fit VAST model to the data by optimizing the TMB function.

```r
Opt_Mod2b = TMBhelper::Optimize(obj = Obj_Mod2b, lower = TmbList_Mod2b[["Lower"]],
                           upper = TmbList_Mod2b[["Upper"]], getsd = TRUE,
                           savedir = DateFile_Mod2b,
                           bias.correct = bias.correct, newtonsteps=2)
```

Save outputs from estimation

```r
Report_Mod2b = Obj_Mod2b$report()

Save_Mod2b = list("Opt"=Opt_Mod2b, "Report"=Report_Mod2b,
                  "ParHat"=Obj_Mod2b$env$parList(Opt_Mod2b$par),
                  "TmbData"=TmbData_Mod2)

save(Save_Mod2b, file=paste0(DateFile_Mod2b,"Save.RData"))
```

7.3 Convergence of five models

```r
temp.table = NULL
temp.table$name = c('Null Model', 'Model 1', 'Model 2', 'Model 1b', 'Model 2b')
temp.table$effect = c('None','Tow Duration', 'Gear Temperature', 'Tow Duration (PCR-only)', 'Gear Tempe
temp.table$convergence = c(Opt_null$convergence,
                           Opt_Mod1$convergence,
                           Opt_Mod2$convergence,
                           Opt_Mod1b$convergence,
                           Opt_Mod2b$convergence)
temp.table$max_gradient = c(Opt_null$max_gradient,
                            Opt_Mod1$max_gradient,
                            Opt_Mod2$max_gradient,
                            Opt_Mod1b$max_gradient,
                            Opt_Mod2b$max_gradient)
temp.table = data.frame(temp.table)
names(temp.table) = c('Model Name','Catchability Covariate','Convergence','Maximum Gradient')
#Print the table
# kable(temp.table)
pander::pandoc.table(temp.table)
```

```
##
## ---------------------------------------------------------------------------
##  Model Name     Catchability Covariate      Convergence   Maximum Gradient
## ------------ ----------------------------- ------------- ------------------
##  Null Model               None                   1           5.363e-08
##
##   Model 1            Tow Duration                1           4.406e-08
##
##   Model 2          Gear Temperature              1           2.296e-07
##
##   Model 1b      Tow Duration (PCR-only)          1           4.453e-08
##
##   Model 2b    Gear Temperature (PCR-only)        0           3.413e-07
## ---------------------------------------------------------------------------
```

7.4 Compare AIC for five models

Lets compare AIC across models…

```
aic.table <- NULL
aic.table$name = c('Null Model', 'Model 1', 'Model 2', 'Model 1b', 'Model 2b')
aic.table$effect = c('None','Tow Duration', 'Gear Temperature', 'Tow Duration (PCR-only)', 'Gear Tempera
aic.table$AIC = c(Opt_null$AIC, Opt_Mod1$AIC, Opt_Mod2$AIC, Opt_Mod1b$AIC, Opt_Mod2b$AIC)
#Calculate dAIC
aic.table$dAIC <- aic.table$AIC - min(aic.table$AIC)
#Data frame
aic.table <- data.frame(aic.table)
names(aic.table) <- c('Model Name', 'Catchability Covariate', 'AIC', 'dAIC')
#Print the table
kable(aic.table)
```

Model Name	Catchability Covariate	AIC	dAIC
Null Model	None	66056.0	178.225
Model 1	Tow Duration	66059.0	181.229
Model 2	Gear Temperature	65877.7	0.000
Model 1b	Tow Duration (PCR-only)	66057.6	179.841
Model 2b	Gear Temperature (PCR-only)	65990.9	113.125

```
# pander::pandoc.table(aic.table)
```