

# GAMs - Acoustics Only - 25 km

Yvonne Barkley

9/26/2020

## Technical coding questions:

- what's the easiest way to make the partial residual plots using the raw data when I'm running the models with scaled data?
- I need help with the predict() function for predicting the test data. I thought it was straight-forward but I'm misunderstanding how it works.

Load libraries

```
library(tidyverse)
library(mgcv)
library(corrplot)
library(geoR)
library(tidymv)
library(here)
```

## Research question:

What environmental variables characterize sperm whale habitat?

Hypothesis: Sperm whales are found in deep, productive offshore waters.

Include more details about what to expect in this document

Load universal variables

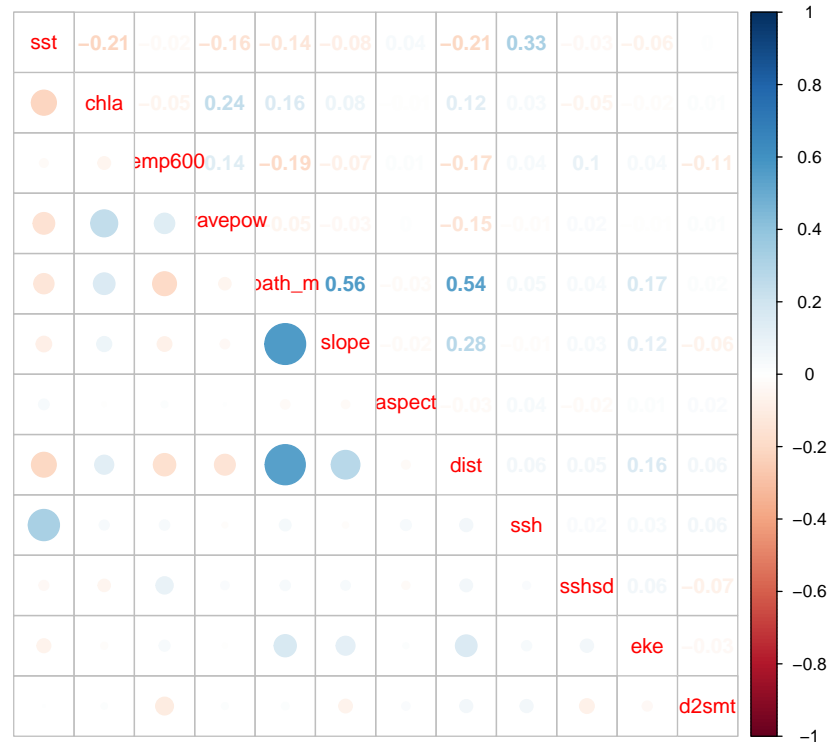
```
# Values used for file and directory names
survey = "AllSurveys"
gridsize = 25
loctype = "AcOnly"
loctype2 = "Ac"
```

Load data

```
PmScaled <- readRDS(here::here(paste0("output/models/", loctype,
  "/data/", "CompletePm_", gridsize, "km_", loctype2, "_scaled.rda")))
# add column for log effort as offset #
PmScaled$log.effort = log(PmScaled$EffArea)
```

Check correlation of covariates

```
require(corrplot)
corrplot.mixed(cor(PmScaled[, 18:29]), upper = "number", lower = "circle")
```



```
# Are all correlation coefficients < |0.6|?
abs(cor(PmScaled[, 18:29])) <= 0.6
```

	sst	chla	temp600	wavepow	bath_m	slope	aspect	dist	ssh	sshsd	eke
sst	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
chla	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
temp600	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
wavepow	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
bath_m	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
slope	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
aspect	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
dist	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
ssh	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
sshsd	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
eke	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE
d2smt	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

	d2smt
sst	TRUE
chla	TRUE
temp600	TRUE
wavepow	TRUE
bath_m	TRUE
slope	TRUE
aspect	TRUE

```

dist      TRUE
ssh       TRUE
sshsd     TRUE
eke       TRUE
d2smt    FALSE

```

## KS tests

I compared the distributions of environmental data between the whales and the absences. Plots are attached in separate powerpoint. In summary, temperature at 600 m, SSH, and chlorophyll were the only variables with significantly different distributions ( $p\text{-value} < 0.05$ ). However, the D statistics were close to zero ( $D \sim 0.1$ ) for each, indicating that although the distributions were different, they were not that far apart. The plots also show how similar the general shape of the distributions are between where the whales were observed and where they were not observed.

## Data Splitting

Split the data into train and test sets

```

require(dplyr)
splitdf <- function(dataframe, seed = NULL) {
  if (!is.null(seed))
    set.seed(seed)
  index <- 1:nrow(dataframe)
  trainindex <- sample(index, trunc(length(index) * 0.7))
  trainset <- dataframe[trainindex, ]
  testset <- dataframe[-trainindex, ]
  list(trainset = trainset, testset = testset)
}

trainAcOnly = NULL
testAcOnly = NULL
seed = 678
seed = 555
for (s in c(1641, 1303, 1604, 1705, 1706)) {

  trSub <- filter(PmScaled, survey == s)

  # subset for presences and split 70/30
  pres1 <- filter(trSub, pa > 0 & sid == 999) # & loc == 1) #for S999 versions
  listPres <- splitdf(pres1, seed) #output is list for train and test

  # subset for absences and split 70/30
  abs0 <- filter(trSub, pa == 0)
  listAbs <- splitdf(abs0, seed) #output is list for train and test

  # combine train data for presence and absence
  trainAll <- rbind(listPres$trainset, listAbs$trainset)

  # combine test data for presence and absence
  testAll <- rbind(listPres$testset, listAbs$testset)
}

```

```

trainAcOnly = rbind(trainAcOnly, trainAll)
testAcOnly = rbind(testAcOnly, testAll)

# trainAcOnly$log.effort <- log(trainAcOnly$EffArea)
# testAcOnly$log.effort <- log(testAcOnly$EffArea)
}
saveRDS(trainAcOnly, here::here(paste0("output/models/", loctype,
"/data/Train_", gridsize, "km_", loctype2, "_S999b.rda")))
saveRDS(testAcOnly, here::here(paste0("output/models/", loctype,
"/data/Test_", gridsize, "km_", loctype2, "_S999b.rda")))

# nrow(dplyr::filter(trainAcOnly, trainAcOnly$pa >0))
# nrow(dplyr::filter(testAcOnly, testAcOnly$pa >0))

```

## Generalized Additive Models

The data are treated as count data, number of sperm whale encounters per cell, and we used the Tweedie distribution since it has been shown to work well when fewer positive responses exist within the data. We used thin-plate regression splines (the default basis) for the smoothers of the environmental predictors. Each smoother was limited to 3 degrees of freedom ( $k=3$ ) to reduce overfitting parameters per recommendations from other studies building similar types of cetaceans distribution models. The log of the effort was included as an offset to account for the variation in effort per cell.

### 25 km spatial scale

- Knots constrained to  $k=3$  according to literature on cetacean distribution models.
- Automatic term selection uses an additional penalty term when determining the smoothness of the function ('select' argument = TRUE)..
- We excluded all non-significant variables ( $\alpha=0.05$ ) and refit the models until all variables were significant.
- REML is restricted maximum likelihood used to optimize the parameter estimates.

Load training and test data

```

# trainAcOnly <- readRDS(here::here(
# paste0('output/models/', loctype, '/data/Train_', gridsize,
# 'km_', loctype2, '.rda') )) testAcOnly <-
# readRDS(here::here( paste0('output/models/', loctype,
# '/data/Test_', gridsize, 'km_', loctype2, '.rda') ))
# trainLoc <- readRDS(here::here(
# paste0('output/models/', loctype, '/data/Train_', gridsize,
# 'km_', loctype2, '_Loc.rda') )) testLoc <-
# readRDS(here::here( paste0('output/models/', loctype,
# '/data/Test_', gridsize, 'km_', loctype2, '_Loc.rda') ))
# Encounters only detected by acoustics, sighting ID = S999
trainS999 <- readRDS(here::here(paste0("output/models/", loctype,
"/data/Train_", gridsize, "km_", loctype2, "_S999.rda")))
testS999 <- readRDS(here::here(paste0("output/models/", loctype,
"/data/Test_", gridsize, "km_", loctype2, "_S999.rda")))

# trainS999 <- readRDS(here::here(

```

```
# paste0('output/models/',loctype, '/data/Train_', gridsize,
# 'km_', loctype2, '_S999b.rda')) testS999 <-
# readRDS(here::here( paste0('output/models/',loctype,
# '/data/Test_', gridsize, 'km_', loctype2, '_S999b.rda')) )
```

## Model Selection

### FULL MODEL

- does not include spatial smoother

```
# * Does NOT include sighted acoustic encounters OR spatial
# smoother
require(mgcv)
twS999 <- gam(pa ~ s(bath_m, k = 3) + s(dist, k = 3) + s(d2smt,
  k = 3) + s(sst, k = 3) + s(chla, k = 3) + s(temp600, k = 3) +
  s(ssh, k = 3) + s(sshsd, k = 3) + s(eke, k = 3) + s(wavepow,
  k = 3) + offset(log.effort), data = trainS999, family = tw,
  link = "log", select = TRUE, method = "REML")
summary(twS999)
```

Family: Tweedie(p=1.01)

Link function: log

Formula:

```
pa ~ s(bath_m, k = 3) + s(dist, k = 3) + s(d2smt, k = 3) + s(sst,
  k = 3) + s(chla, k = 3) + s(temp600, k = 3) + s(ssh, k = 3) +
  s(sshsd, k = 3) + s(eke, k = 3) + s(wavepow, k = 3) + offset(log.effort)
```

Parametric coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) -22.9628    0.1142  -201.1  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(bath_m)	7.535e-01	2	1.534	0.04211 *
s(dist)	7.708e-05	2	0.000	1.00000
s(d2smt)	1.150e-04	2	0.000	0.66559
s(sst)	1.146e+00	2	1.392	0.08427 .
s(chla)	8.596e-01	2	2.919	0.00794 **
s(temp600)	1.749e+00	2	10.093	8.57e-06 ***
s(ssh)	8.916e-01	2	4.078	0.00225 **
s(sshsd)	8.440e-01	2	2.597	0.01280 *
s(eke)	7.982e-01	2	1.668	0.04011 *
s(wavepow)	7.147e-05	2	0.000	0.82655

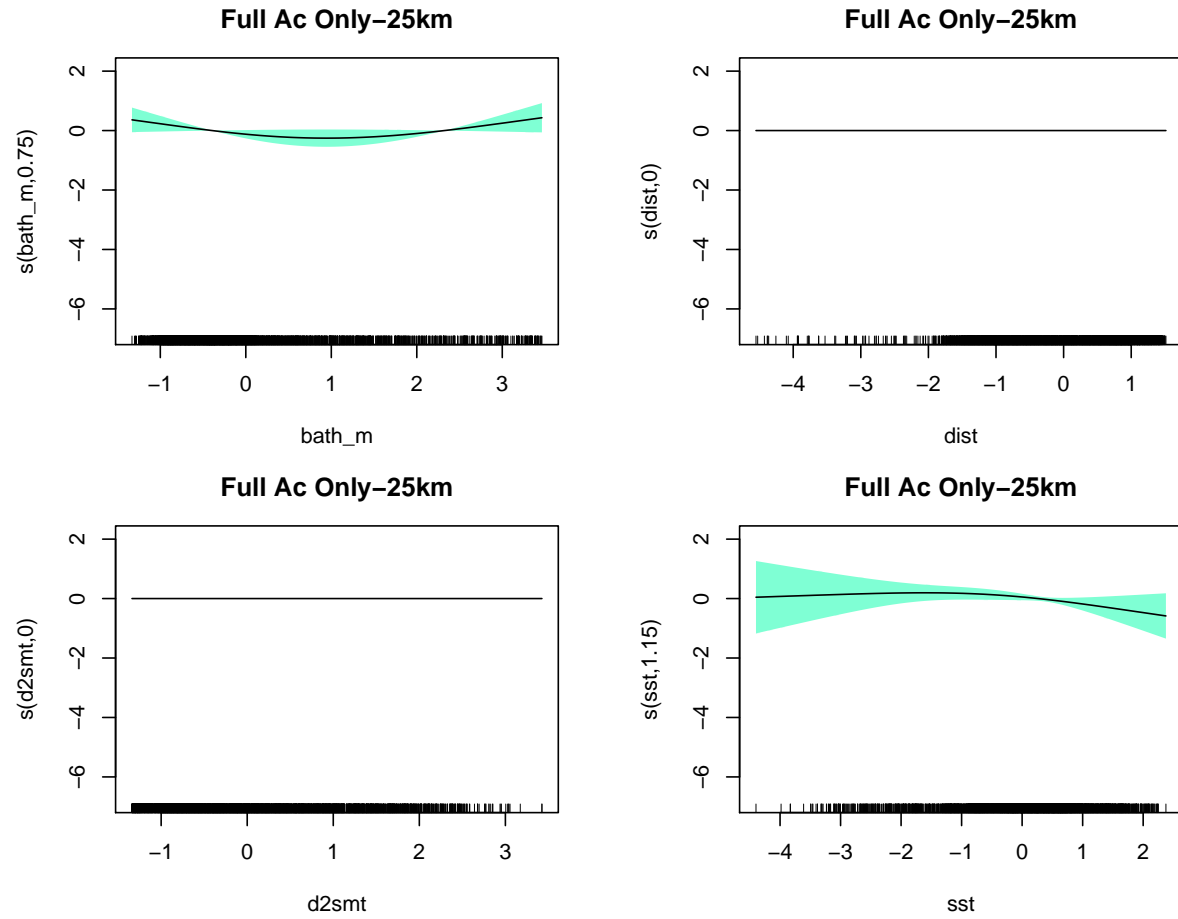
---

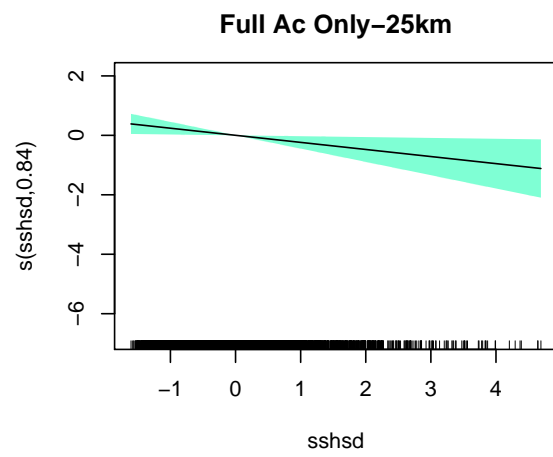
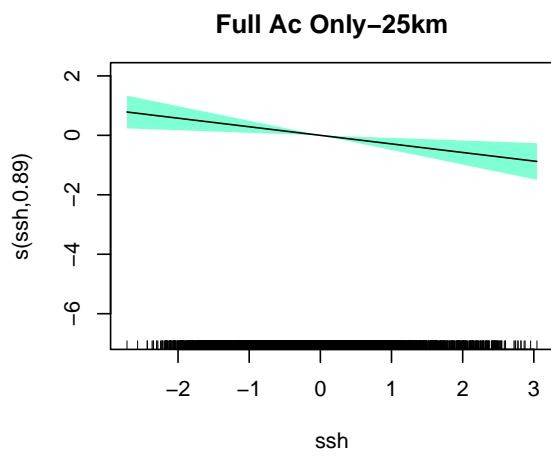
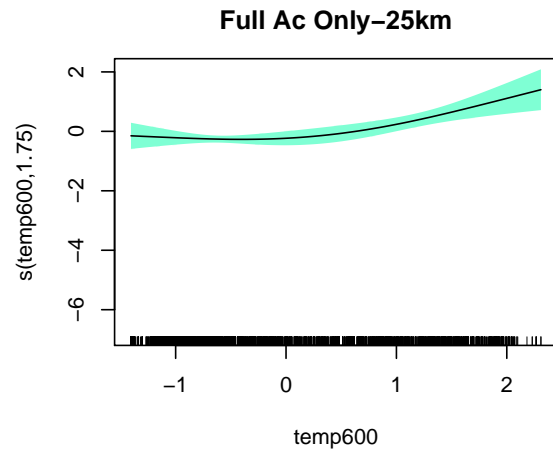
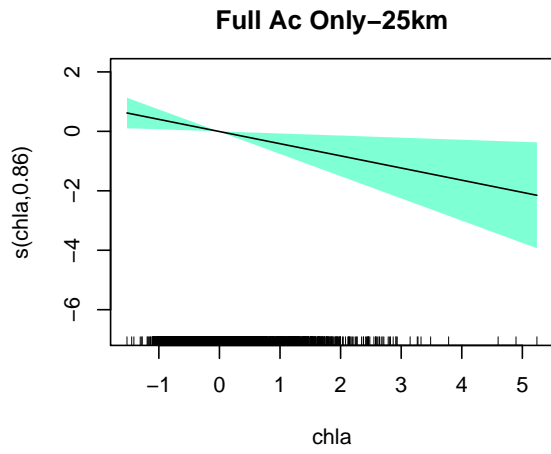
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

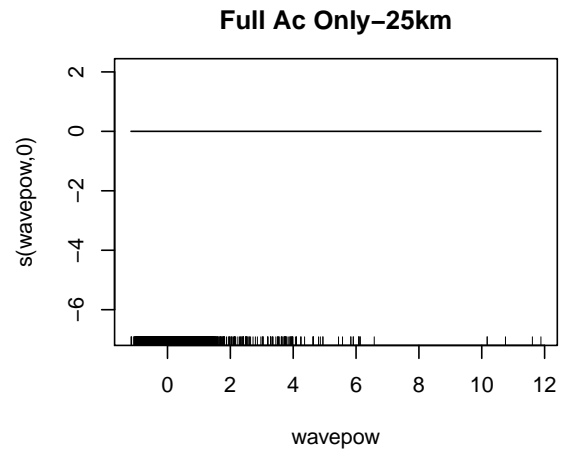
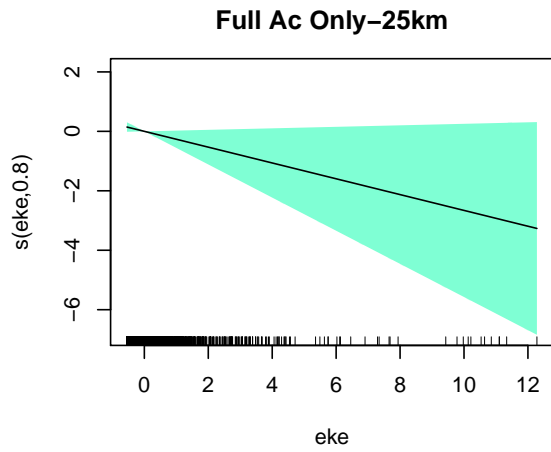
R-sq.(adj) = 0.0471 Deviance explained = 9.36%

-REML = 295.94 Scale est. = 1.0404 n = 3657

# + s(slope, k=3) + s(aspect, k=3) # removed 9/27

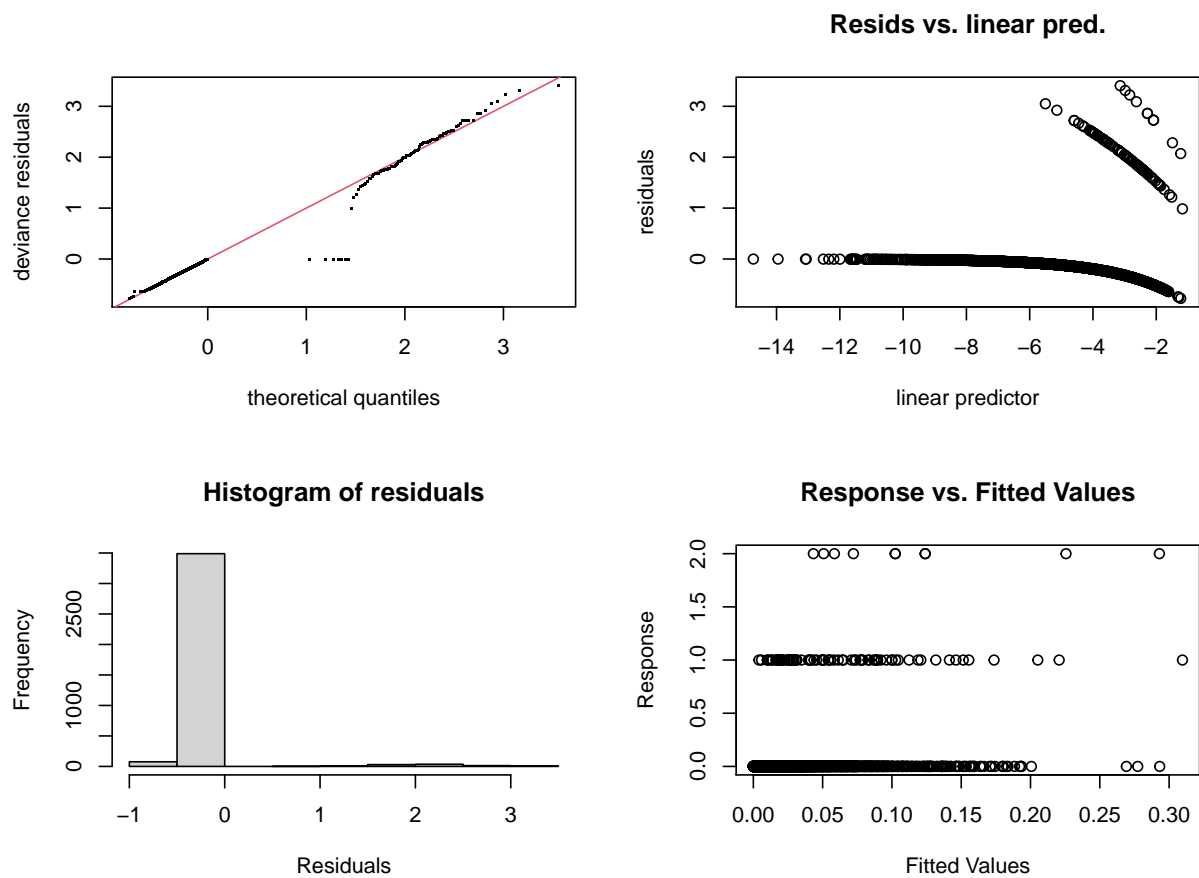






**MODEL DIAGNOSTICS** The stripe at the bottom left of the residuals vs. fitted values (linear predictor) corresponds to the zeros.





Method: REML Optimizer: outer newton  
 full convergence after 23 iterations.  
 Gradient range [-0.0005605756,6.506142e-05]  
 (score 295.9409 & scale 1.040415).  
 eigenvalue range [-2.033627e-05,10199.49].  
 Model rank = 21 / 21

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(bath_m)	2.00e+00	7.54e-01	0.90	0.850
s(dist)	2.00e+00	7.71e-05	0.88	0.345
s(d2smt)	2.00e+00	1.15e-04	0.85	0.020 *
s(sst)	2.00e+00	1.15e+00	0.86	0.070 .
s(chla)	2.00e+00	8.60e-01	0.86	0.070 .
s(temp600)	2.00e+00	1.75e+00	0.75	<2e-16 ***
s(ssh)	2.00e+00	8.92e-01	0.89	0.525
s(sshsd)	2.00e+00	8.44e-01	0.84	0.005 **
s(eke)	2.00e+00	7.98e-01	0.88	0.320
s(wavepow)	2.00e+00	7.15e-05	0.81	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## REDUCE MODEL PARAMETERS

- Removed non-significant variables:
  - distance to land
  - distance to seamount
  - sst
  - wave power

```
# * Does NOT include sighted acoustic encounters
```

```
twS999b <- gam(pa ~ s(bath_m, k = 3) + s(chla, k = 3) + s(temp600,  
  k = 3) + s(ssh, k = 3) + s(sshsd, k = 3) + s(eke, k = 3) +  
  offset(log.effort), data = trainS999, family = tw, link = "log",  
  select = TRUE, method = "REML")  
summary(twS999b)
```

Family: Tweedie(p=1.01)

Link function: log

Formula:

```
pa ~ s(bath_m, k = 3) + s(chla, k = 3) + s(temp600, k = 3) +  
  s(ssh, k = 3) + s(sshsd, k = 3) + s(eke, k = 3) + offset(log.effort)
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-22.949	0.113	-203.2	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

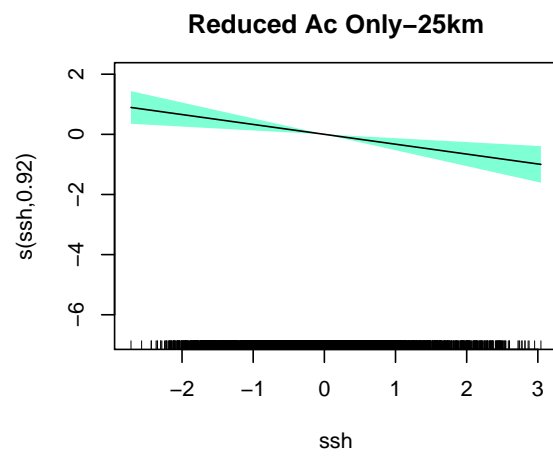
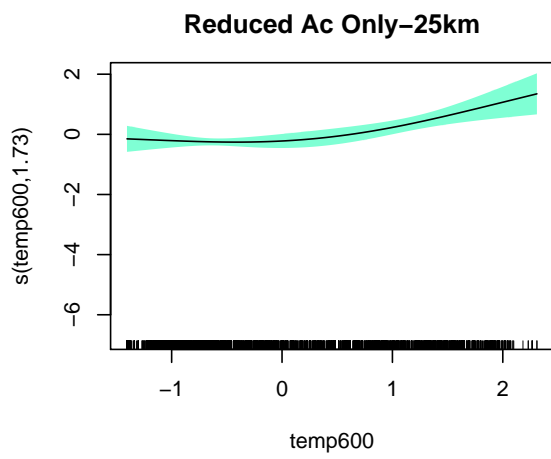
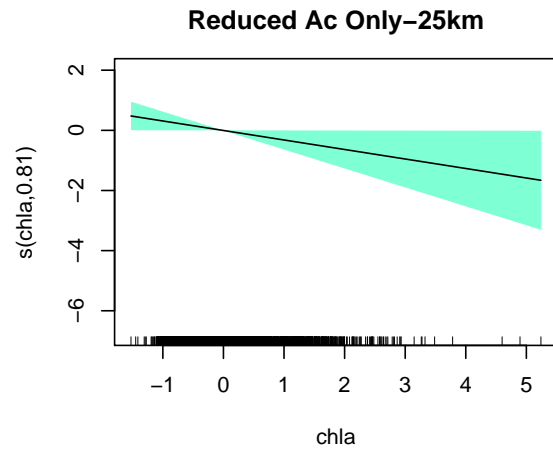
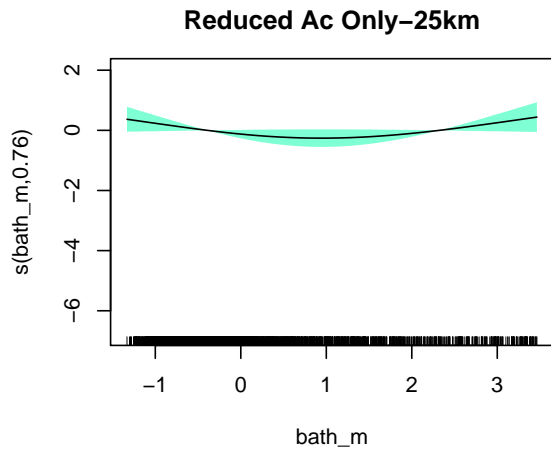
	edf	Ref.df	F	p-value
s(bath_m)	0.7599	2	1.588	0.04008 *
s(chla)	0.8115	2	2.034	0.02480 *
s(temp600)	1.7331	2	9.479	1.79e-05 ***
s(ssh)	0.9160	2	5.414	0.00058 ***
s(sshsd)	0.8394	2	2.507	0.01437 *
s(eke)	0.7968	2	1.655	0.04145 *

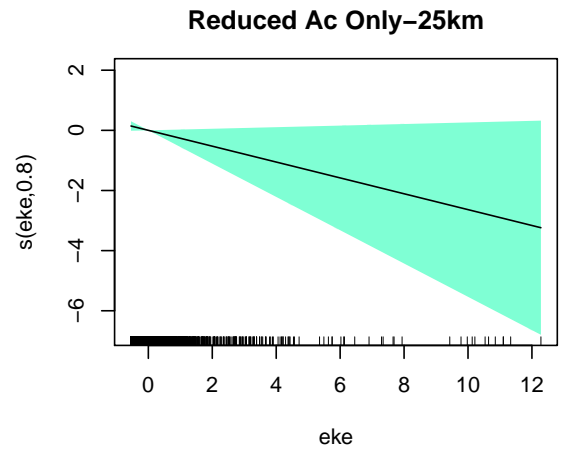
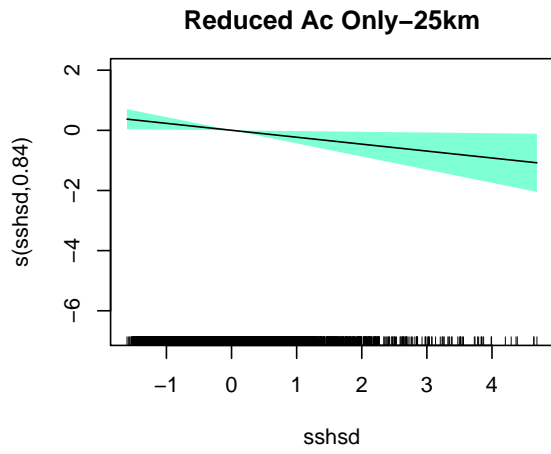
---

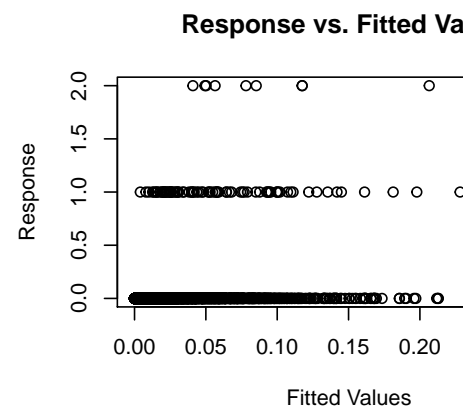
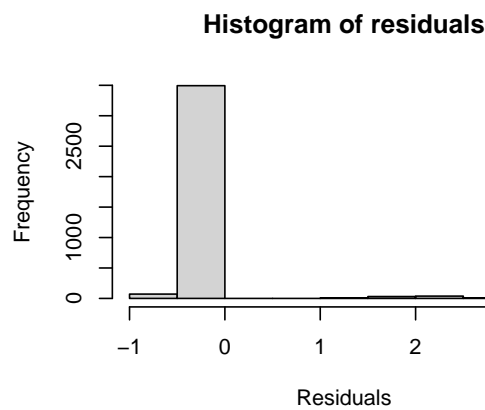
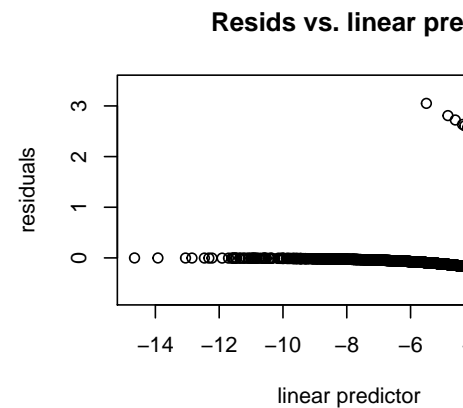
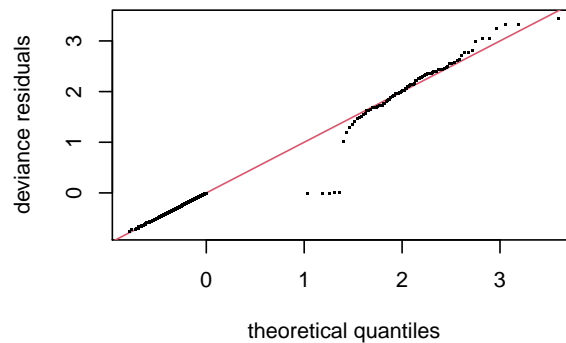
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0428 Deviance explained = 8.74%

-REML = 296.44 Scale est. = 1.0406 n = 3657







## MODEL DIAGNOSTICS

Method: REML Optimizer: outer newton  
 full convergence after 21 iterations.  
 Gradient range [-0.0006173768,7.132451e-05]  
 (score 296.4414 & scale 1.040571).  
 Hessian positive definite, eigenvalue range [1.653048e-05,10199.49].  
 Model rank = 13 / 13

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(bath_m)	2.000	0.760	0.90	0.80
s(chla)	2.000	0.812	0.86	0.05 *
s(temp600)	2.000	1.733	0.75	<2e-16 ***
s(ssh)	2.000	0.916	0.89	0.55
s(sshsd)	2.000	0.839	0.84	<2e-16 ***
s(eke)	2.000	0.797	0.88	0.26

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Include 2D Lat-Lon

Includes 2D Lat-Lon smoother to account for spatial structure in the data and fit the spatial variation not explained by the other predictors

\* Notice that the temperature at 600m is no longer significant compared to the previous models

\* Chlorophyll and SSHsd remain significant

+ Does this indicate that they aren't spatially structured and are independent of location?

```
twS999LL <- gam(pa ~ s(Longitude, Latitude) + s(bath_m, k = 3) +
  s(dist, k = 3) + s(d2smt, k = 3) + s(sst, k = 3) + s(chla,
  k = 3) + s(temp600, k = 3) + s(ssh, k = 3) + s(sshsd, k = 3) +
  s(eke, k = 3) + s(wavepow, k = 3) + offset(log.effort), data = PmScaled,
  family = tw, link = "log", select = TRUE, method = "REML")
summary(twS999LL)
```

Family: Tweedie(p=1.01)

Link function: log

Formula:

```
pa ~ s(Longitude, Latitude) + s(bath_m, k = 3) + s(dist, k = 3) +
  s(d2smt, k = 3) + s(sst, k = 3) + s(chla, k = 3) + s(temp600,
  k = 3) + s(ssh, k = 3) + s(sshsd, k = 3) + s(eke, k = 3) +
  s(wavepow, k = 3) + offset(log.effort)
```

Parametric coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -22.81564    0.09803  -232.7   <2e-16 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

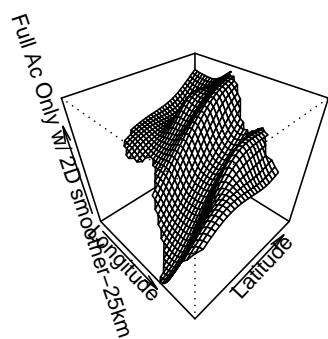
	edf	Ref.df	F	p-value
s(Longitude, Latitude)	1.537e+01	29	3.844	< 2e-16 ***
s(bath_m)	5.216e-01	2	0.548	0.131423
s(dist)	1.699e-04	2	0.000	1.000000
s(d2smt)	7.386e-01	2	1.409	0.040055 *
s(sst)	8.525e-01	2	2.872	0.006848 **
s(chla)	9.216e-01	2	5.690	0.000334 ***
s(temp600)	8.917e-01	2	4.088	0.001146 **
s(ssh)	1.663e-04	2	0.000	0.887862
s(sshsd)	1.757e+00	2	15.707	1.64e-08 ***
s(eke)	7.298e-01	2	1.191	0.067162 .
s(wavepow)	1.162e-04	2	0.000	0.733200

---

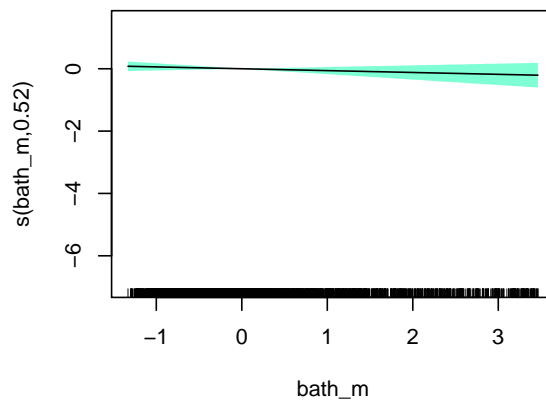
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.11 Deviance explained = 17.9%

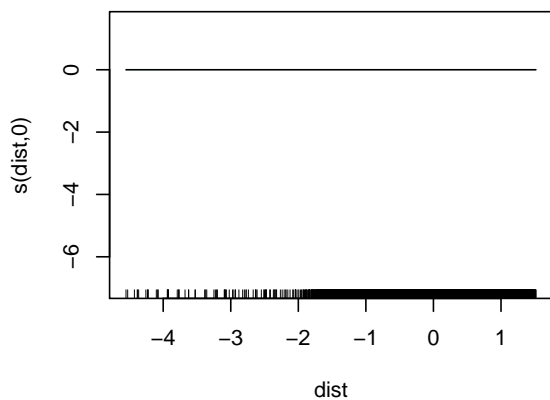
-REML = 517.38 Scale est. = 1.0345 n = 5287



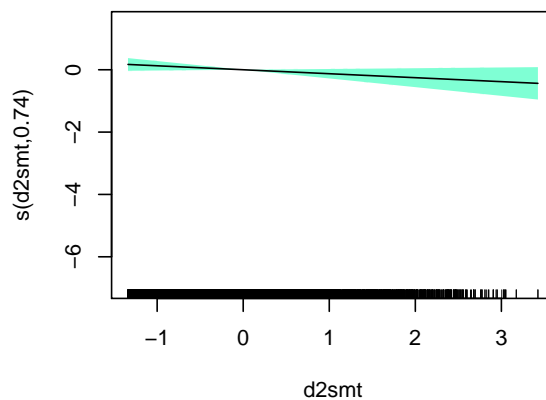
Full Ac Only w/ 2D smoother-25km

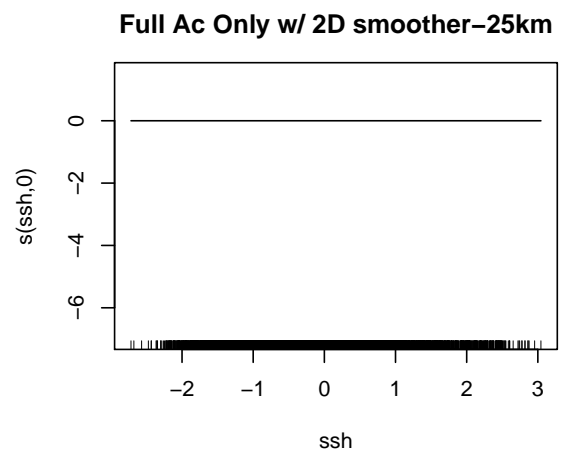
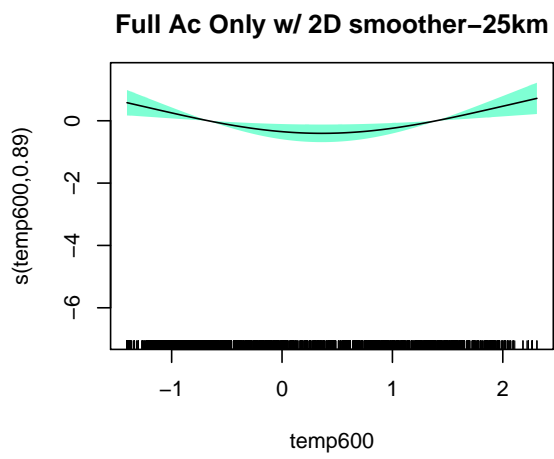
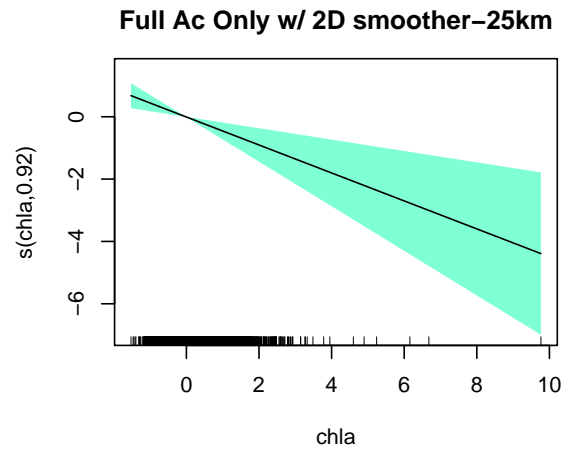
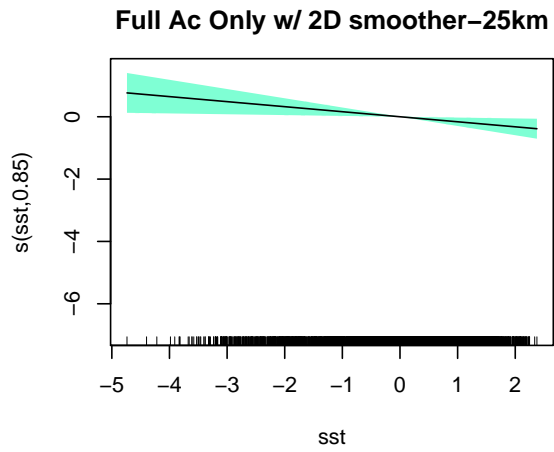


Full Ac Only w/ 2D smoother-25km

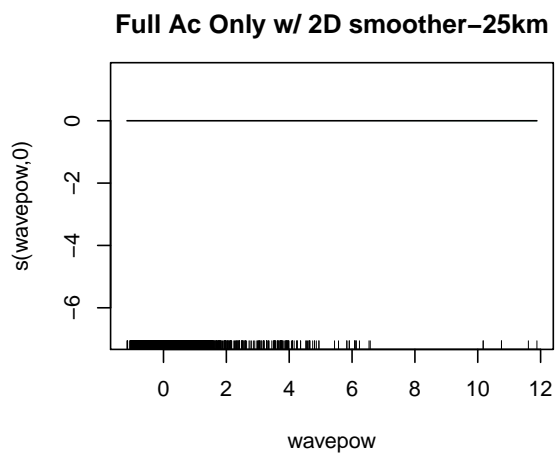
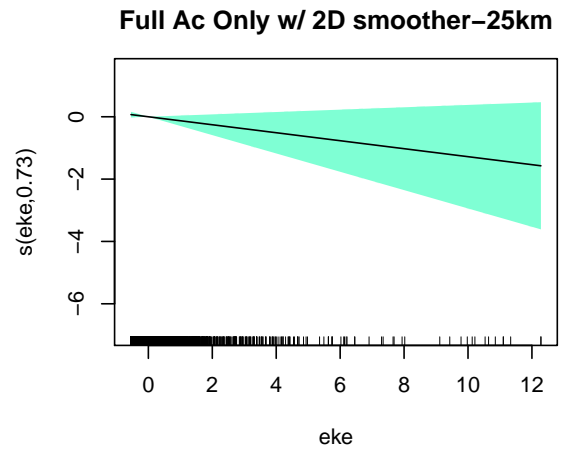
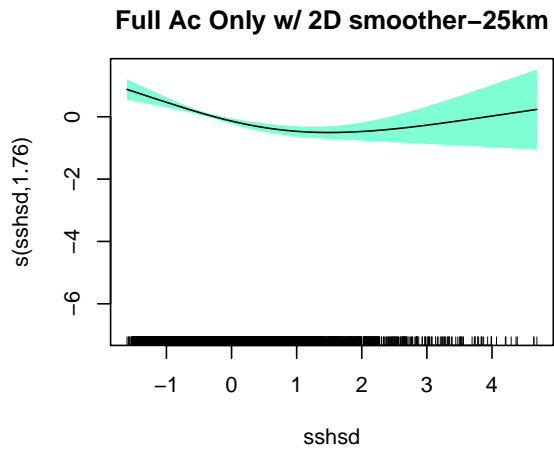


Full Ac Only w/ 2D smoother-25km









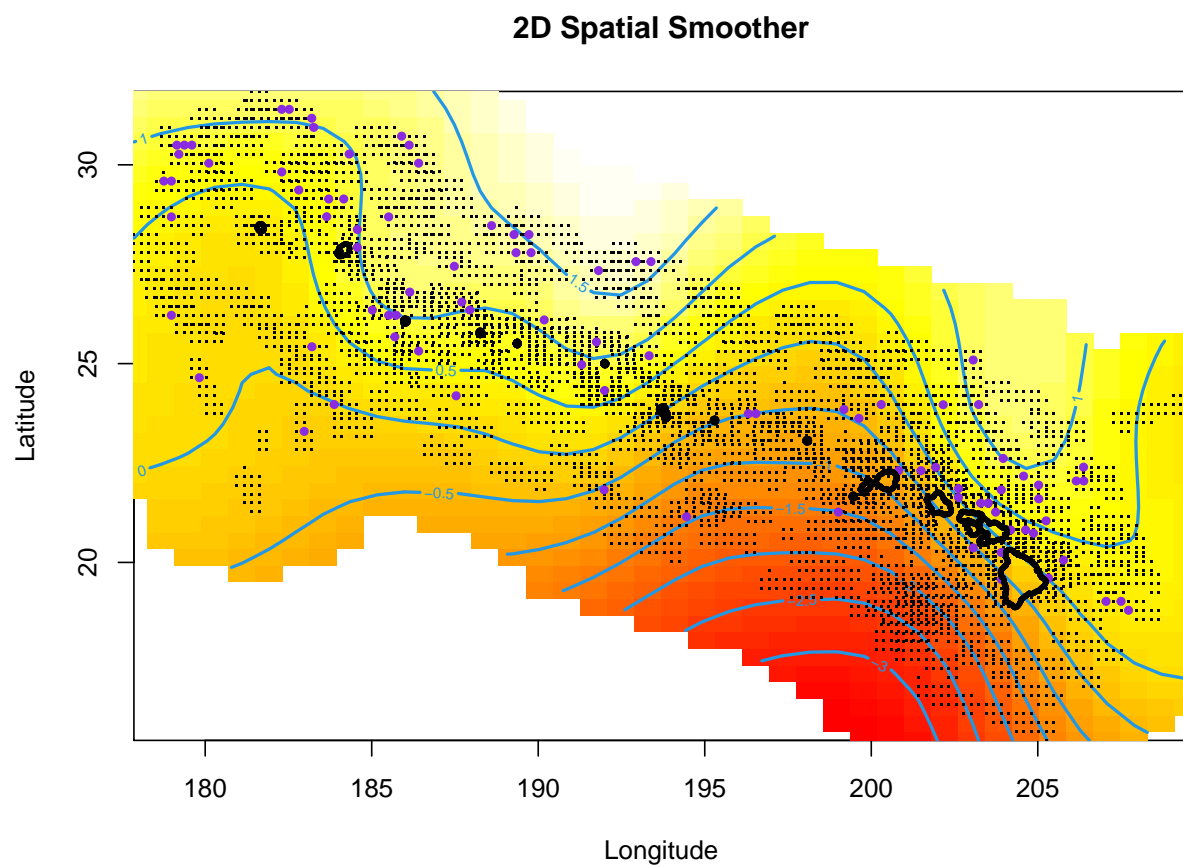
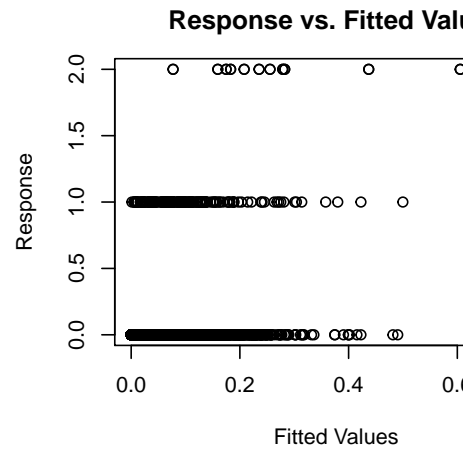
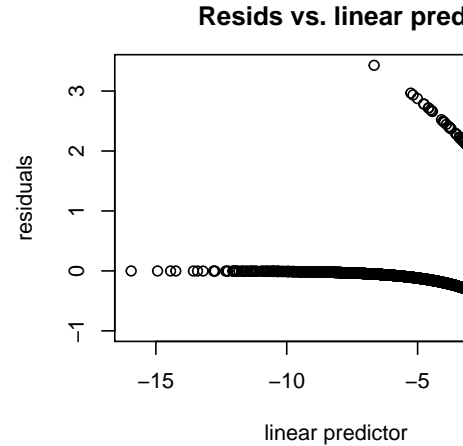
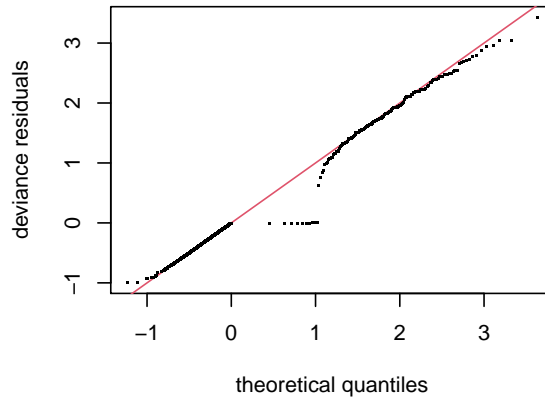


Figure 1: Purple dots represent acoustically detected encounters. Black dots are all data points(grid centroids)



## MODEL DIAGNOSTICS

Method: REML Optimizer: outer newton  
 full convergence after 30 iterations.  
 Gradient range [-0.001068839,0.000143211]  
 (score 517.3793 & scale 1.034495).  
 Hessian positive definite, eigenvalue range [2.214768e-05,21499.47].  
 Model rank = 50 / 50

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(Longitude,Latitude)	2.90e+01	1.54e+01	0.84	<2e-16 ***
s(bath_m)	2.00e+00	5.22e-01	0.86	0.010 **
s(dist)	2.00e+00	1.70e-04	0.83	<2e-16 ***
s(d2smt)	2.00e+00	7.39e-01	0.82	<2e-16 ***
s(sst)	2.00e+00	8.53e-01	0.82	<2e-16 ***
s(chla)	2.00e+00	9.22e-01	0.83	<2e-16 ***
s(temp600)	2.00e+00	8.92e-01	0.83	<2e-16 ***
s(ssh)	2.00e+00	1.66e-04	0.86	0.025 *
s(sshsd)	2.00e+00	1.76e+00	0.81	<2e-16 ***
s(eke)	2.00e+00	7.30e-01	0.83	<2e-16 ***
s(wavepow)	2.00e+00	1.16e-04	0.83	<2e-16 ***
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## REDUCE MODEL PARAMETERS

- Removed non-significant variables:

- depth
- distance to land
- distance to seamount
- sst
- temp at 600 m
- SSH
- EKE
- wave power

```
twS999LLb <- gam(pa ~ s(Longitude, Latitude) + s(chla, k = 3) +  
  s(sshsd, k = 3) + offset(log.effort), data = trainS999, family = tw,  
  link = "log", select = TRUE, method = "REML")  
summary(twS999LLb)
```

Family: Tweedie(p=1.01)

Link function: log

Formula:

```
pa ~ s(Longitude, Latitude) + s(chla, k = 3) + s(sshsd, k = 3) +  
  offset(log.effort)
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-23.1397	0.1322	-175	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Longitude, Latitude)	11.4515	29	2.298	1.46e-12 ***
s(chla)	0.7884	2	1.735	0.031557 *
s(sshsd)	1.3692	2	5.023	0.000972 ***

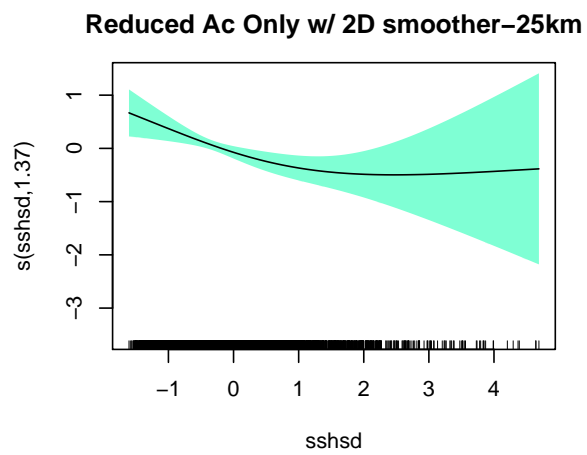
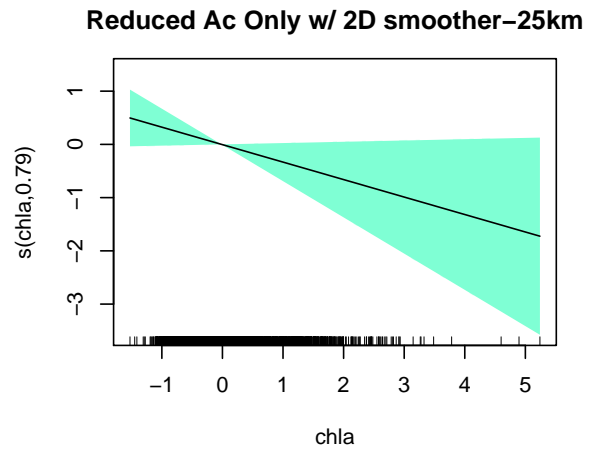
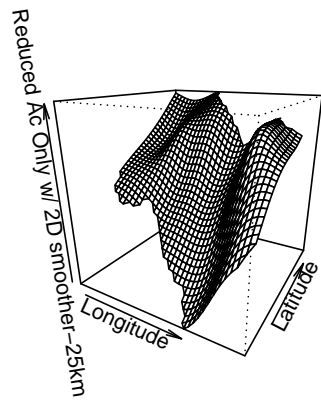
---

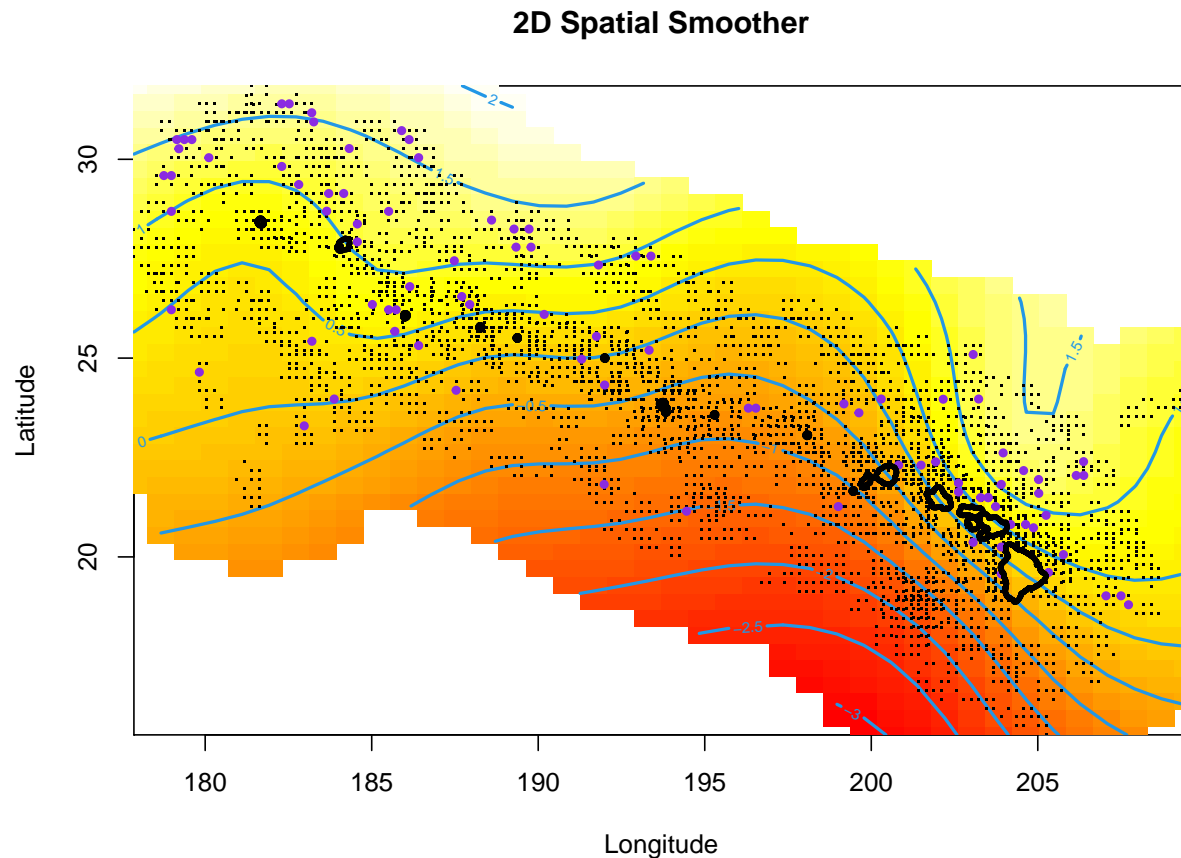
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0642 Deviance explained = 15.6%

-REML = 285.19 Scale est. = 1.0386 n = 3657

```
# twS999LLfull <- gam(pa ~ s(Longitude, Latitude) + s(chla,  
# k=3) + s(sshsd, k=3) + offset(log.effort), data = PmScaled,  
# family = tw, link = 'log', select = TRUE, method = 'REML')  
# summary(twS999LLfull)
```





## Conclusions

The first reduced Acoustics Only model included more dynamic variables for the significant smooth terms compared to both models that included a 2D spatial smoother. The former addressed my hypothesis more clearly given that depth, chl a, and SSH/SSHsd were included as important variables. The introduction of the 2D smoother reduces all dynamic variables except for chl a and SSHsd. Acoustic encounters of sperm whales indicate sperm whale occurrence is related to productivity in some way whether or not a 2D smoother is included. If the 2D smoother is left out of the model, I should acknowledge how certain important variables may also be spatially autocorrelated. Some papers state that spatial autocorrelation exists, calculate Moran's I to determine the magnitude, and don't do anything more about it, such as Forney et al. 2015.

**NEED HELP HERE:** Trying to sort out how to use the test data set for prediction purposes and how to use those results to evaluate model performance.

## Predict Test Data

```

# https://gist.github.com/aperium/9fc737ea311a758328eadf27c2426e47

require(magrittr)
require(dplyr)

# pulling the prediction and residual data from the model
trainS999 %<>% mutate(resid = resid(twS999LLb), predict = predict(twS999LLb))

predTrain <- predict.gam(twS999LLb, type = "response") #calculate MSE for these to compare with test s
trainS999$fit <- predTrain

# using scale of 0,1,2 makes this hard to interpret
mean((trainS999$pa - trainS999$fit)^2) #MSE
mean(abs((trainS999$pa - trainS999$fit))) #Mean absolute error
## Calculate MSE AFTER transforming the predictions back to
## the same scale as the observed data

# plot %<>% + geom_smooth(aes(y = predict), method = gam,
# formula = y ~ s(x, bs = 'ps'), data = filter(sa_factor,
# climate_scenario == unique(sa_factor$climate_scenario)[1]),
# size = 1, se = FALSE)

predSW <- predict.gam(twS999LLb, newdata = testS999, type = "response",
  se.fit = TRUE)
testFinal <- data.frame(testS999, fit = predSW$fit, se.fit = predSW$se.fit)
mean((testFinal$pa - testFinal$fit)^2) #MSE

mean(abs((testFinal$pa - testFinal$fit))) #Mean absolute error

plot(testS999$chla, testFinal$predSW)

predSWtidy <- predict_gam(twS999LLb) #, newdata = testS999)#, type = 'response', se.fit = TRUE)

ggplot(testFinal, aes(chla, fit)) + geom_smooth_ci()

lons <- seq(180, 210, 1/4)
lats <- seq(15, 33, 1/4)
zz <- matrix(NA, nrow = 98, ncol = 98) #nrow and ncol are sqrt(max(effCell))
zz[trainS999$effCells] <- trainS999$predicted

image(lons, lats, zz)

```

**Models including only static variables** What are the effects of the static/geographic variables on sperm whale occurrence? \* Resulted in only ~6% deviance explained for any iteration \* Depth, slope, distance to seamount were significant, no aspect included

**Models only including dynamic variables to evaluate their explanatory power alone.**

- Resulted in 8.6% explained deviance, all variables significant except wave power