Probability of occurrence

Charles Perretti

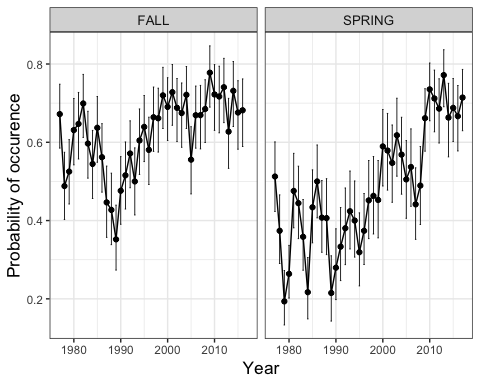
8/14/2018

## Probability of occurence of summer flounder

Probability of occurence of summer flounder is estimated in each season and year using a logistic regression, i.e., a GLM with a bionomial response and log-odds link. Error bars represent the 95% confidence interval.

This can be used as a point of comparison to the habitat-based probability of occurence plot.

# Load packages  
library(dplyr)  
library(ggplot2)  
  
# Function to calculate probability of occurence  
calc\_p <- function(df) {  
 m <-   
 glm(formula = pres ~ 1,  
 data = df,   
 family = binomial)  
   
 preds <-  
 predict(object = m,  
 type = "link",   
 se.fit = TRUE,  
 newdata = df[1,])  
   
 upr <- preds$fit + 1.96\*preds$se.fit  
 lwr <- preds$fit - 1.96\*preds$se.fit  
 fit <- preds$fit  
   
 p <- m$family$linkinv(fit)  
 p\_high <- m$family$linkinv(upr)  
 p\_low <- m$family$linkinv(lwr)  
   
 data.frame(p, p\_high, p\_low)  
}  
  
# Load summer flounder data  
load("df\_survdat.rda")  
  
# Choose strata to include (from ADIOS)  
strata\_fall <- c(1010, 1050, 1090, 1610, 1650, 1690,  
 1730, 3010, 3020, 3030, 3040, 3050,  
 3060, 3070, 3080, 3090, 3100, 3110,  
 3120, 3130, 3140, 3150, 3160, 3170,  
 3180, 3190, 3200, 3210, 3220, 3230,  
 3240, 3250, 3260, 3270, 3280, 3290,  
 3300, 3310, 3320, 3330, 3340, 3350,  
 3360, 3370, 3380, 3390, 3400, 3410,  
 3420, 3430, 3440, 3450, 3460, 3470,  
 3480, 3490, 3500, 3510, 3520, 3530,  
 3540, 3550, 3560, 3570, 3580, 3590,  
 3600, 3610)  
  
strata\_spring <- c(1010, 1020, 1030, 1040, 1050, 1060,  
 1070, 1080, 1090, 1100, 1110, 1120,  
 1610, 1620, 1630, 1640, 1650, 1660,  
 1670, 1680, 1690, 1700, 1710, 1720,  
 1730, 1740, 1750, 1760)  
  
# Filter for each season  
df\_survdat\_spring <-  
 df\_survdat %>%  
 dplyr::filter(YEAR > 1976,  
 SEASON == "SPRING",  
 STRATUM %in% strata\_spring)  
  
df\_survdat\_fall <-  
 df\_survdat %>%  
 dplyr::filter(YEAR > 1976,  
 SEASON == "FALL",  
 STRATUM %in% strata\_fall)  
  
# Combine seasons  
df\_survdat2use <-  
 rbind(df\_survdat\_spring,  
 df\_survdat\_fall) %>%  
 dplyr::mutate(pres = BIOMASS > 0)  
  
  
# Calculate probability of occurence in each season x year  
df\_prob\_occ <-  
 df\_survdat2use %>%  
 dplyr::group\_by(SEASON, YEAR) %>%  
 dplyr::do(calc\_p(df = .))  
  
# Make plot  
ggplot(df\_prob\_occ,   
 aes(x = YEAR, y = p)) +  
 geom\_point() +  
 geom\_errorbar(aes(ymin = p\_low, ymax = p\_high),   
 size = 0.3, width = 0.3) +  
 geom\_line() +  
 xlab("Year") +  
 ylab("Probability of occurence") +  
 theme\_bw() +  
 facet\_wrap(~SEASON) +  
 theme(axis.title = element\_text(size = 13),  
 strip.text = element\_text(size = 10))



ggsave("./prob\_occ.jpg", width = 5, height = 5)