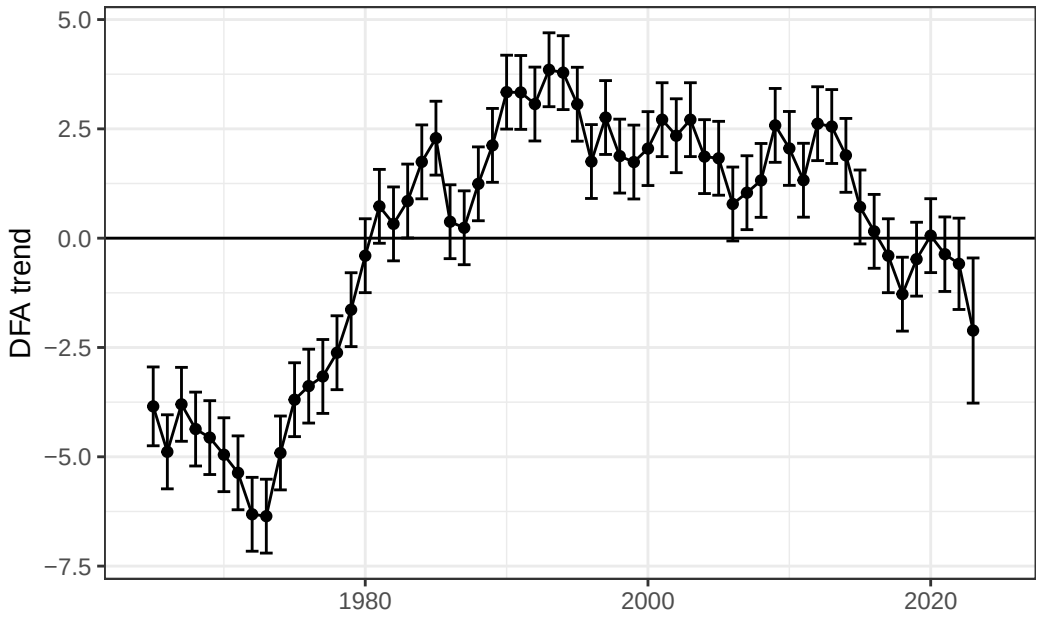
DLMs and GAMs for GOA salmon

```
library(MARSS)
library(mgcv)
library(dplyr)
library(forecast)
library(ggplot2)
library(lubridate)
library(marssTMB)
library(tidyverse)
```

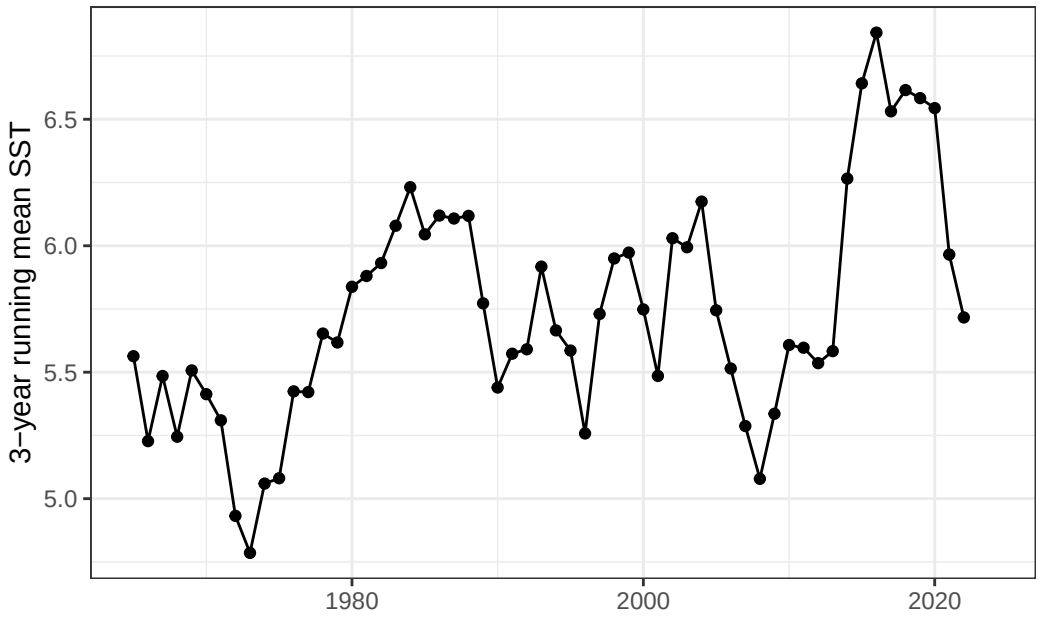
Data

```
# Load salmon timeseries collapsed into DFA trend
# salm.dat <- read.csv("GOA_salmon_case_study/Data/dfa_trend.csv") %>%
#  rename(year = t, salmon_DFA = estimate)

salm.dat <- readRDS(here::here("GOA_salmon_case_study/Data/dfa_trend.rds"))

# Load 3-year running mean winter SST timeseries
# sst.dat <- read.csv("GOA_salmon_case_study/Data/winterSST_3yr_running_mean.csv") %>%
#   dplyr::select(!X)

sst.dat <- readRDS(here::here("GOA_salmon_case_study/Data/winterSST_3yr_running_mean.rds"))

# Bind
right_join(salm.dat, sst.dat) -> dat

# Plot salmon DFA
ggplot(dat, aes(year, salmon_DFA)) +
  theme_bw() +
  geom_line() +
```

```
  geom_hline(yintercept = 0) +
  geom_point() +
  geom_errorbar(aes(x=year, ymin=conf.low, ymax=conf.high)) +
  xlab("") +
  ylab("DFA trend")
```



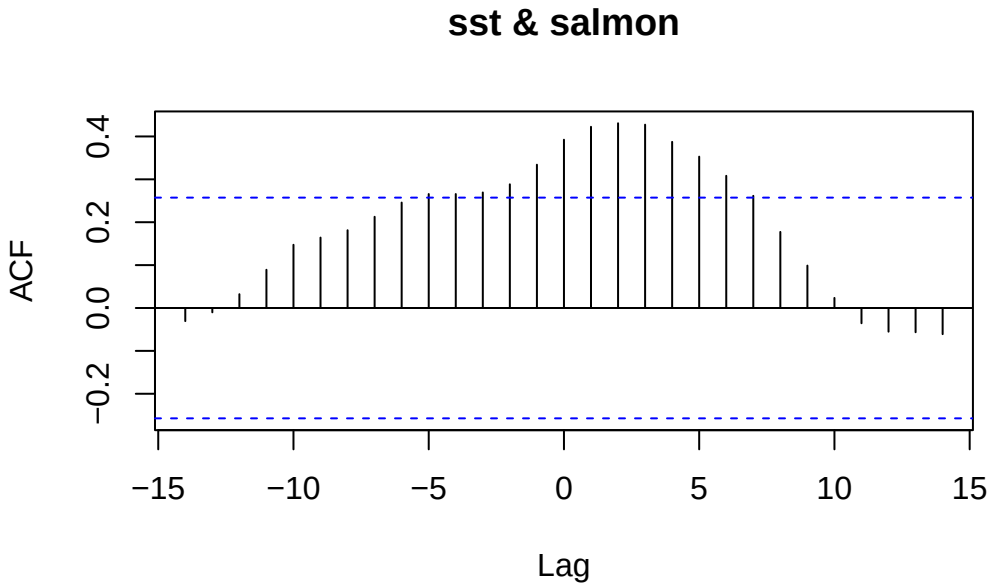

```
# Plot sst
ggplot(dat, aes(year, sst_3yr_running_mean)) +
  theme_bw() +
  geom_line() +
  geom_point() +
  xlab("") +
  ylab("3-year running mean SST")
```




```
# Scale predictor sst
dat %>%
  mutate(z_sst = scale(sst_3yr_running_mean)) %>%
  na.omit() -> dat2
```

Evaluate lags between trend estimate (response) and response relationships. 1-3 months strongly correlated

```
sst <- as.vector(dat2$z_sst)
salmon <- as.vector(dat2$salmon_DFA)

ccf(sst, salmon)
```



Use a lag of 3 months?

```
lag_n <- 0
dat2$lagged_sst <- c(rep(NA, lag_n), dat2$z_sst[1:(nrow(dat2)-lag_n)])
dat2 <- dat2[-c(1:lag_n),]
```

DLMs

Define the model.

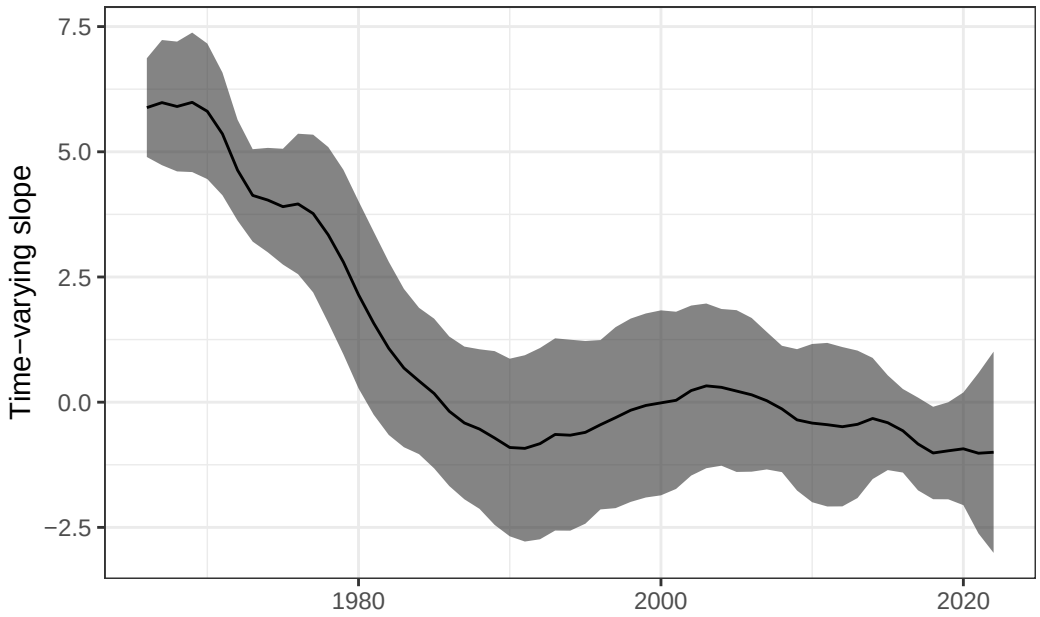
```
m <- 2
TT <- nrow(dat2)
B <- diag(m)  ## 2x2; Identity
U <- matrix(0, nrow = m, ncol = 1)  ## 2x1; both elements = 0
Q <- matrix(list(0), m, m)  ## 2x2; all 0 for now
diag(Q)[1] <- 0.0000001
diag(Q)[2] <- c("q.beta")
#diag(Q) <- c("q.alpha", "q.beta")  ## 2x2; diag = (q1,q2)
Z <- array(NA, c(1, m, TT))  ## NxMxT; empty for now
Z[1, 1, ] <- rep(1, TT)  ## Nx1; 1's for intercept
Z[1, 2, ] <- dat2$lagged_sst  ## Nx1; predictor variable
A <- matrix(0)  ## 1x1; scalar = 0
```

```
R <- matrix("r")  ## 1x1; scalar = r
## only need starting values for regr parameters
inits_list <- list(x0 = matrix(c(0, 0), nrow = m))

## list of model matrices & vectors
mod_list <- list(B = B, U = U, Q = Q, Z = Z, A = A, R = R)

# convert response to matrix
dat_mat <- matrix(dat2$salmon_DFA, nrow = 1)

# fit the model -- crank up the maxit to ensure convergence
dlm_1 <- MARSS(dat_mat, inits = inits_list, model = mod_list,
               control = list(maxit=4000), method="TMB")
```

```
MARSS fit is
Estimation method: TMB
Estimation converged in 14 iterations.
Log-likelihood: -111.193
AIC: 230.3861   AICc: 231.1553

         Estimate
R.r         2.027
Q.q.beta    0.368
x0.X1       1.324
x0.X2       5.881
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

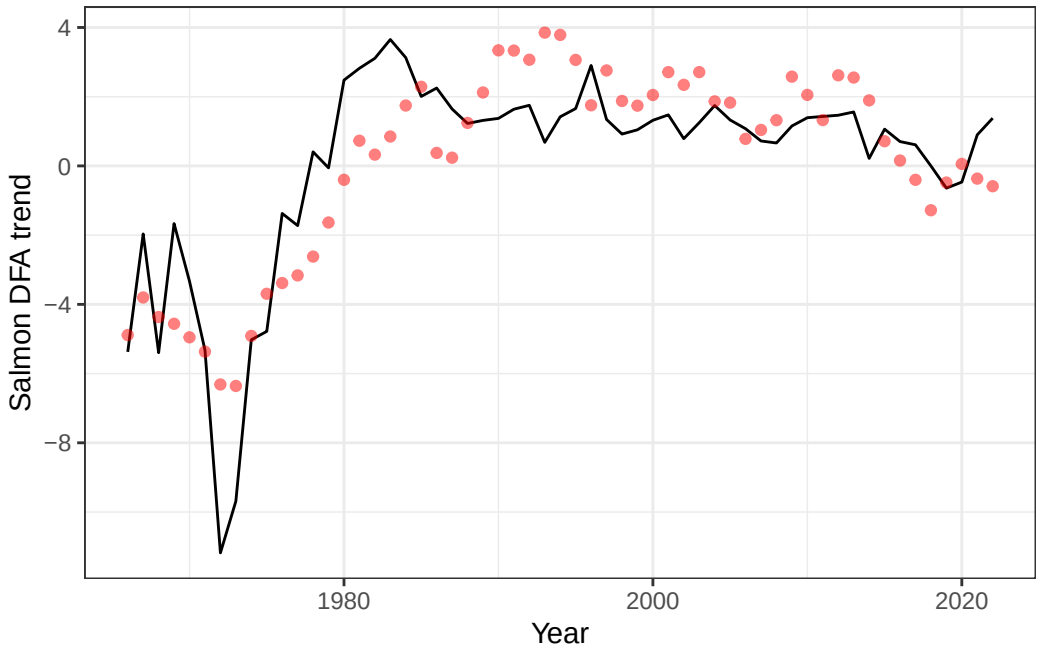
Put the state estimates back into the original dataframe and plot. The slope is generally positive, which is what we expect – and exhibits clear variation through time

```
dat2$int_est <- dlm_1$states[1,]
dat2$int_se <- dlm_1$states.se[1,]
dat2$slope_est <- dlm_1$states[2,]
dat2$slope_se <- dlm_1$states.se[2,]

ggplot(dat2, aes(year, slope_est)) +
  geom_ribbon(aes(ymin=slope_est-2*slope_se, ymax = slope_est+2*slope_se), alpha=0.6) +
  geom_line() + ylab("Time-varying slope") + xlab("") + theme_bw()
```



Plot the fitted values from the model, and some diagnostics

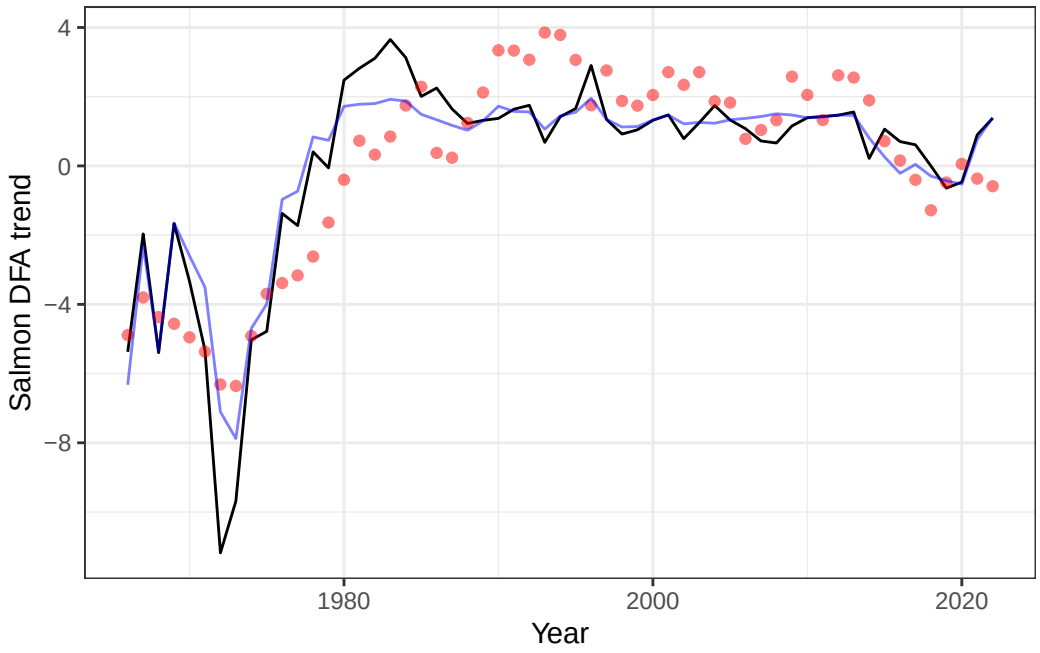
```
pred <- fitted(dlm_1)

dat2$pred <- pred$.fitted
ggplot(dat2, aes(year, pred)) +
  geom_line() +
  ylab("Salmon DFA trend") +
  xlab("Year") +
  theme_bw() +
  geom_point(aes(year,salmon_DFA), col="red",alpha=0.5)
```



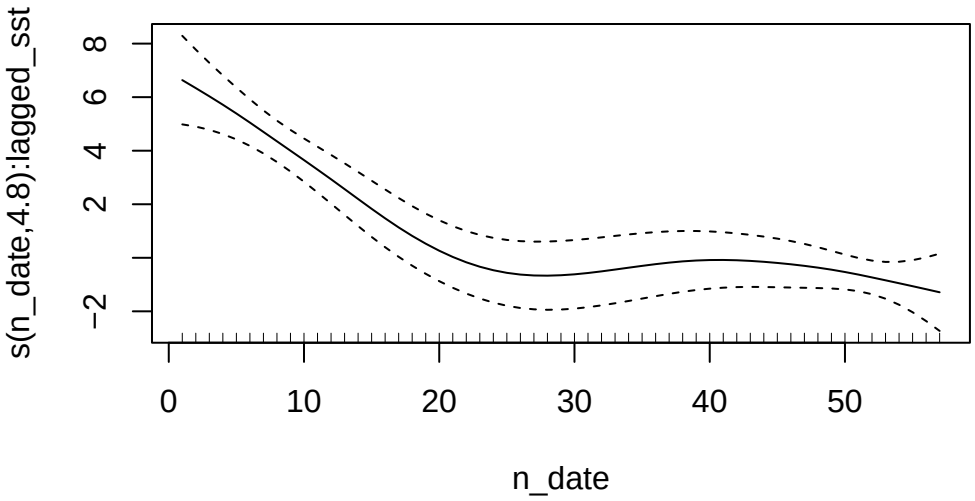
GAMs

Let's fit the same model with a GAM

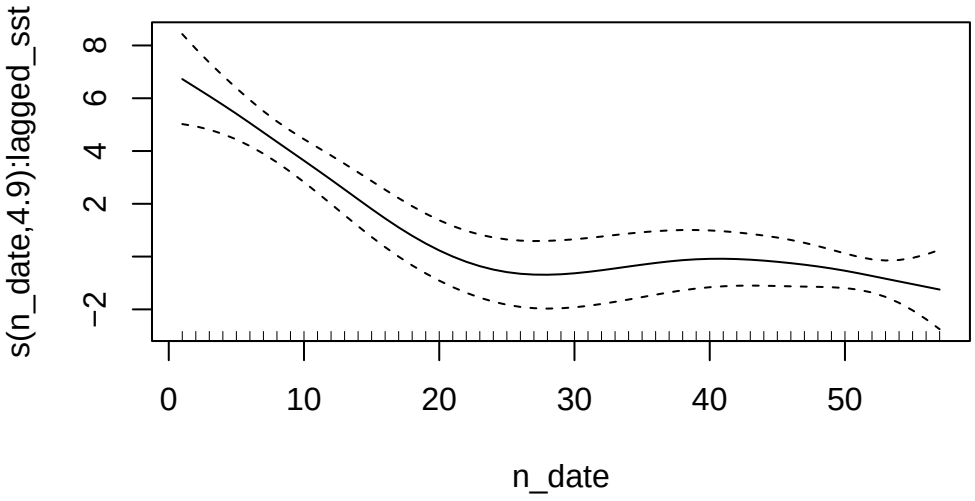
```
# The 'sp' parameter also controls the penalty on smoothing (below)
dat2$n_date <- as.numeric(as.factor(dat2$year))
gam_cr <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "cr",k=nrow(dat2)),
              data = dat2)
gam_tp <- gam(salmon_DFA ~ s(n_date, by = lagged_sst,k=nrow(dat2)),
              data = dat2) # k is ignored
gam_gp <- gam_cr <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "gp",k=nrow(dat2)),
                        data = dat2)

dat2$pred_gam <- predict(gam_tp)

ggplot(dat2, aes(year, pred)) +
  geom_line()+
  ylab("Salmon DFA trend") +
  xlab("Year") +
  theme_bw() +
  geom_point(aes(year,salmon_DFA), col="red",alpha=0.5) +
  geom_line(aes(year,pred_gam), col="blue",alpha=0.5)
```



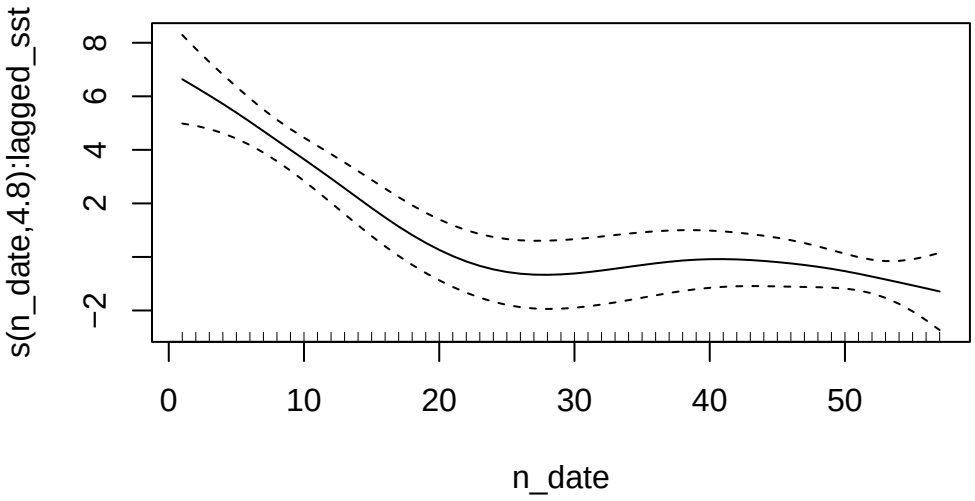

```
# extract the smooth estimates with SEs to compare to DLM
plot_cr <- plot(gam_cr, seWithMean = TRUE, n = nrow(dat2))
```



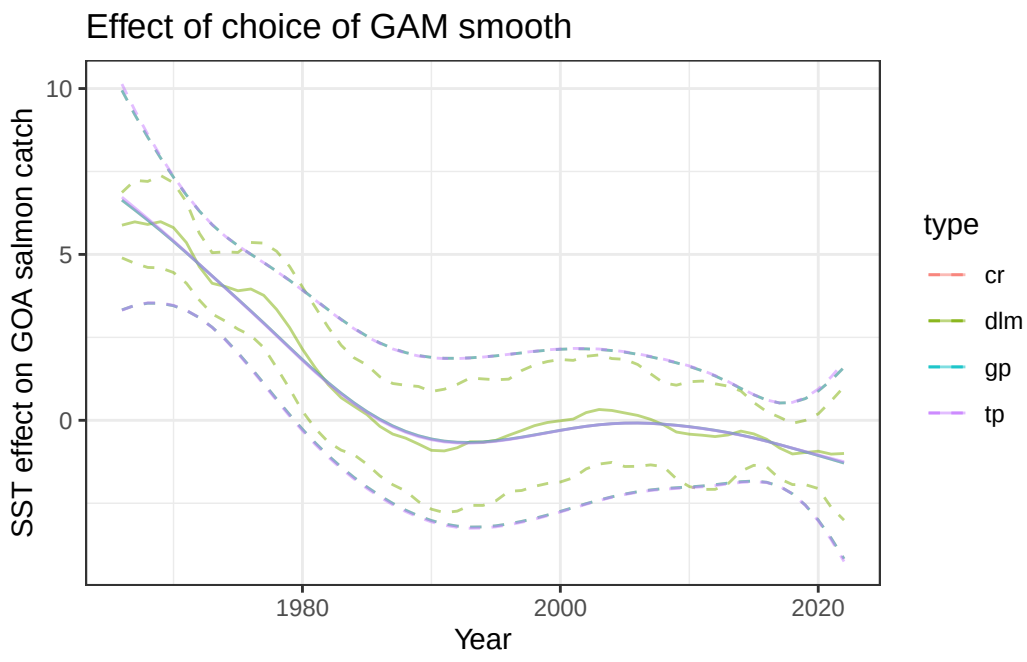
```r
plot_tp <- plot(gam_tp, seWithMean = TRUE, n = nrow(dat2))
```



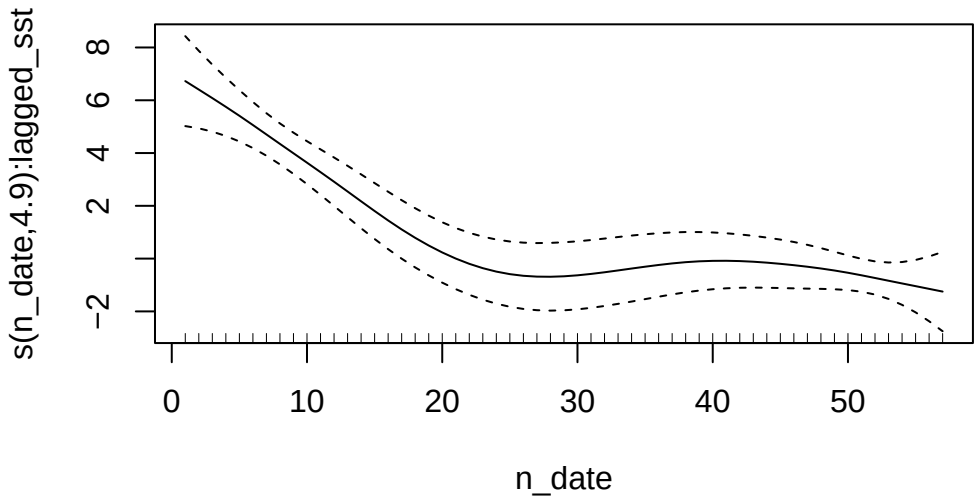

```r
plot_gp <- plot(gam_gp, seWithMean = TRUE, n = nrow(dat2))
```



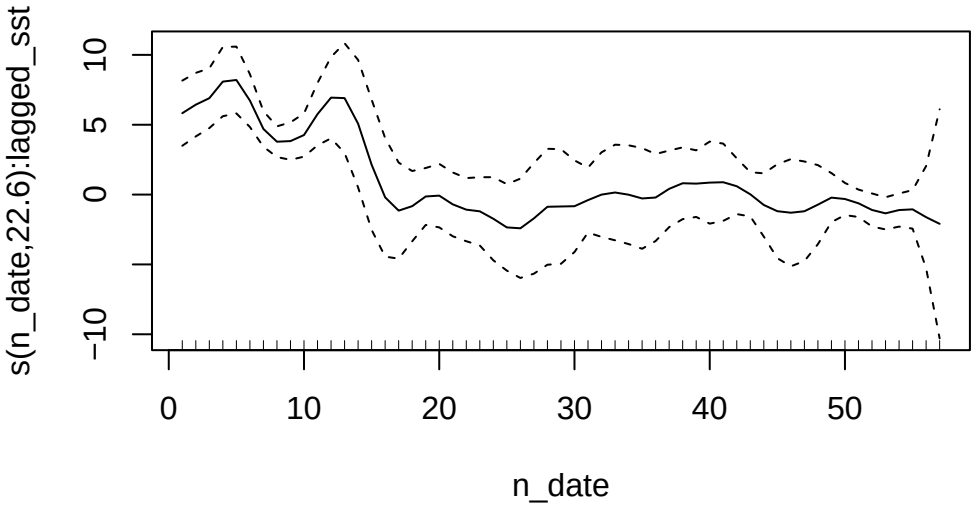
```r
df_cr <- data.frame(year = dat2$year,
                    est = plot_cr[[1]]$fit,
                    se = plot_cr[[1]]$se, type="cr")
df_tp <- data.frame(year = dat2$year,
                    est = plot_tp[[1]]$fit,
                    se = plot_tp[[1]]$se, type="tp")
df_gp <- data.frame(year = dat2$year,
                    est = plot_gp[[1]]$fit,
                    se = plot_gp[[1]]$se, type="gp")
df_dlm <- data.frame(year = dat2$year,
                     est = dat2$slope_est,
                     se = dat2$slope_se, type="dlm")
df_smooth <- rbind(df_cr, df_tp, df_gp, df_dlm)

ggplot(df_smooth, aes(year, est, col=type, fill=type)) +
  geom_line(aes(year,est-2*se), alpha=0.5, linetype=2) +
  geom_line(aes(year,est+2*se), alpha=0.5, linetype=2) +
  geom_line(aes(year,est), alpha=0.5) +
  theme_bw() +
  xlab("Year") +
  ylab("SST effect on GOA salmon catch") +
  ggtitle("Effect of choice of GAM smooth")
```



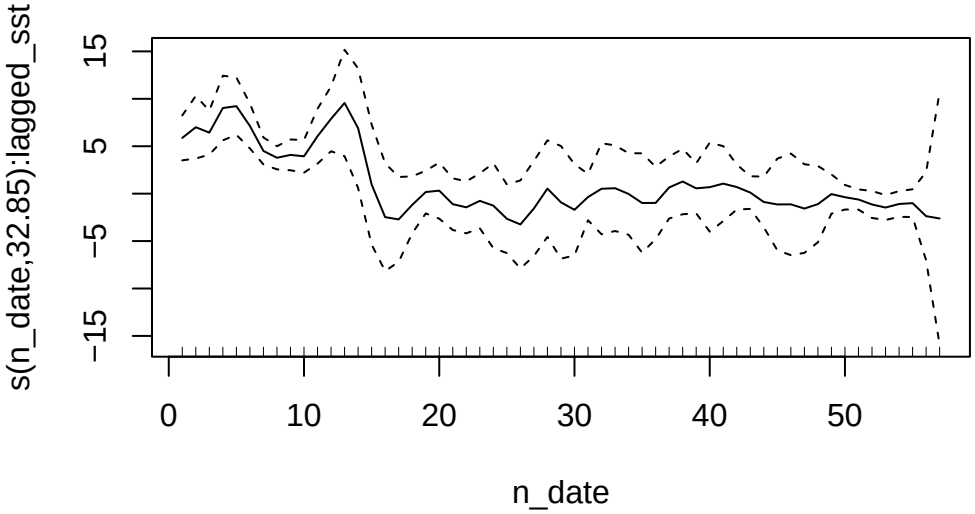
```
gam_cr30 <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "cr",
                           k=nrow(dat2)), data = dat2)
gam_cr60 <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "cr",
                           k=nrow(dat2),sp=16), data = dat2)
gam_cr90 <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "cr",
                           k=nrow(dat2),sp=2.5), data = dat2)
gam_cr120 <- gam(salmon_DFA ~ s(n_date, by = lagged_sst, bs = "cr",
                            k=nrow(dat2),sp=0.3), data = dat2)

plot_cr30 <- plot(gam_cr30, seWithMean = TRUE, n = nrow(dat2))
```



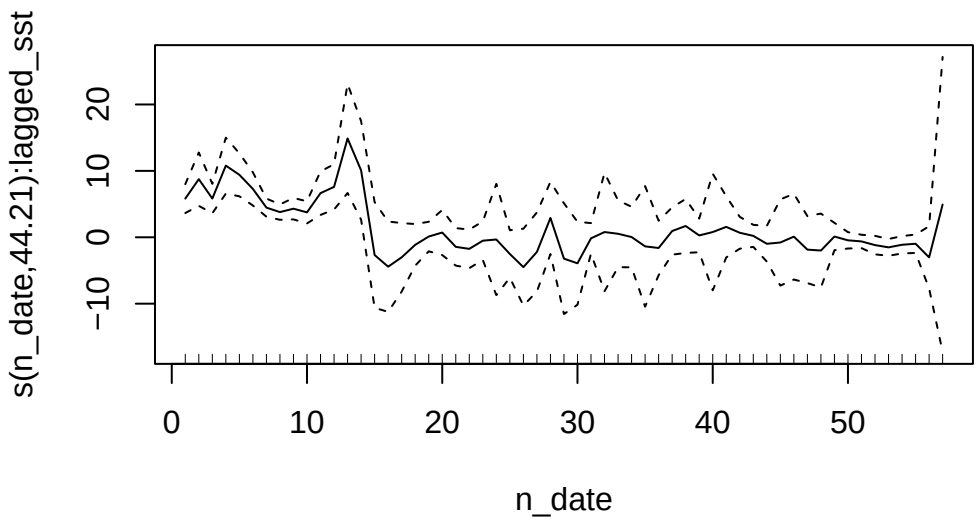

```
plot_cr60 <- plot(gam_cr60, seWithMean = TRUE, n = nrow(dat2))
```




```
plot_cr90 <- plot(gam_cr90, seWithMean = TRUE, n = nrow(dat2))
```



```
plot_cr120 <- plot(gam_cr120, seWithMean = TRUE, n = nrow(dat2))
```




```
df_30 <- data.frame(year = dat2$year,
                    est = plot_cr30[[1]]$fit,
                    se = plot_cr30[[1]]$se, knots="30")
df_60 <- data.frame(year = dat2$year,
                    est = plot_cr60[[1]]$fit,
                    se = plot_cr60[[1]]$se, knots="60")
df_90 <- data.frame(year = dat2$year,
                    est = plot_cr90[[1]]$fit,
                    se = plot_cr90[[1]]$se, knots="90")
df_120 <- data.frame(year = dat2$year,
                     est = plot_cr120[[1]]$fit,
                     se = plot_cr120[[1]]$se, knots="120")
df_smooth <- rbind(df_30, df_60, df_90, df_120)

ggplot(df_smooth, aes(year, est, col=knots, fill=knots)) +
  geom_line(aes(year,est-2*se), alpha=0.5, linetype=2) +
  geom_line(aes(year,est+2*se), alpha=0.5, linetype=2) +
  geom_line(aes(year,est), alpha=0.5) +
  theme_bw() +
  xlab("Year") +
```

```
  ylab("SST effect on GOA salmon catch") +
  ggtitle("Effect of smoothing DF on SEs")
```