VAST & RE model comparisons

Kelly Mistry

7/16/2021

Visual Comparison of VAST & random effects model predicted biomass results

```r
library(ggplot2)

# Plotting the VAST vs RE predicted biomass in WESTERN
for (sub in 1:length(subregion)) {
  for (species in 1:2) {
    RE_results <- get(paste0("RE_", subregion[sub], "_", stock_folder_names[species], "_results"))
    VAST_results <- get(paste0("VAST_", stock_folder_names[species], "_beta_4s_pred"))
    survey_data <- get(paste0(subregion[sub], "_", stock_folder_names[species], "_data"))

    plot <- ggplot(data = RE_results, aes(x = Years, y = Biomass)) +
      geom_line(aes(color = "red")) +
      geom_pointrange(aes(ymin = LCI, ymax = UCI, color = "red")) +
      geom_line(data = VAST_results, aes(x = Year, y = WESTERN, color = "black")) +
      geom_point(data = survey_data, aes(x = YEAR, y = AREA_BIOMASS, color = "green")) +
      scale_color_manual(values =c("black"="black","red"="red", "green"="green"),
                         labels = c("black" = "VAST", "red" = "RE", "green" = "Survey")) +
      labs(title = paste0(stock_names[species], " - ", subregion[sub], " subregion"))

    ggsave(plot = plot, filename = paste0(subregion[sub], "_", stock_folder_names[species], "_RE_v_VAST
  }
}

## The numbers don't match what I expected from the Index_Bioimass VAST plots, so trying importing the

# except that now I see that this data only has values for the survey years - and that is actually all

assign(paste0(subregion[1], "_", stock_folder_names[1], "_VAST_2_results"), read.csv(paste0("/Users/kel

x <- pivot_wider(data = get(paste0(subregion[1], "_", stock_folder_names[1], "_VAST_2_results")), names_

x <- na.omit(x)
colnames(x)[3:11] <- c(paste0(subregion[1], "_pred_biomass"), paste0(subregion[2], "_pred_biomass"), pas

# RE_WESTERN_Sebastes_alutus_results
#
plot <- ggplot(data = RE_WESTERN_Sebastes_alutus_results, aes(x = Years, y = Biomass)) +
  geom_line(aes(color = "red")) +
  geom_pointrange(aes(ymin = LCI, ymax = UCI, color = "red")) +
    geom_line(data = VAST_Sebastes_alutus_beta_4s_pred, aes(x = Year, y = WESTERN, color = "black")) +
```

```
    geom_point(data = WESTERN_Sebastes_alutus_data, aes(x = YEAR, y = AREA_BIOMASS, color = "green")) +
    geom_point(data = x, aes(x = Year, y = WESTERN_pred_biomass, color = "blue")) +
    scale_color_manual(values =c("black"="black","red"="red", "green"="green", "blue"="blue"),
                       labels = c("VAST.Index_ctl","RE", "Survey", "VAST.Table_SS3")) +
    labs(title = paste0(stock_names[1], " - ", subregion[1], " subregion")) +
    theme(legend.position="top")

ggsave(plot = plot, filename = "WESTERN_RE_v_VAST_results_2VASTs.png", path = here("Plots"))
```

First step: Model performance evaluation

Metric option 1: Mean absolute scaled error

$$\text{MASE} = \frac{\frac{1}{N}\sum_{n=1}^{N}|P_n - O_n|}{\frac{1}{T-1}\sum_{t=2}^{T}|O_t - O_{t-1}|}$$

```
### with RE predicted results first
# columns required: RE_results$Years, RE_results$Biomass and survey_data$YEAR, survey_data$AREA_BIOMASS
MASE_fun <- function(model_results, survey_data){
  survey_years <- sort(unique(survey_data$YEAR))
  N <- length(survey_years) # number of survey years
  T <- N-1 # number of time step

  inner_numerator_term <- abs(model_results$Biomass[model_results$Years %in% survey_years] - survey_data
  full_numerator_term <- (1/N)*sum(inner_numerator_term)

  inner_denominator_term <- vector()
  for(i in 1:T) {
    inner_denominator_term[i] <- abs(survey_data$AREA_BIOMASS[i+1] - survey_data$AREA_BIOMASS[i])
  }
  full_denominator_term <- 1/(T-1)*sum(inner_denominator_term)

  MASE <- full_numerator_term/full_denominator_term

  return(MASE)
}

# Using POP western region as initial example
MASE_fun(POP_re_results, WESTERN_Sebastes_alutus_data)
MASE_fun(POP_VAST_gamma_beta4_pred_biom, WESTERN_Sebastes_alutus_data)
```

Metric option 2: Mean relative error

This metric is the one used in the unpublished 2013 paper to, with this description: "Statistics selected to evaluate the performance of the various methods include the mean relative error (MRE) of biomass (relative error is defined here as estimate/true-1) and variability in these relative errors." Based on this description, here is version 1:

$$\text{MRE} = \frac{1}{N}\sum_{n=1}^{N}\frac{P_n}{O_n} - 1$$

```
MRE_v1_fun <- function(model_results, survey_data) {
  survey_years <- sort(unique(survey_data$YEAR))
  N <- length(survey_years) # number of survey years

  RE_vec <- model_results$Biomass[model_results$Years %in% survey_years]/(survey_data$AREA_BIOMASS - 1)

  MRE <- sum(RE_vec)*(1/N)

  return(list(MRE = MRE, RE_vec = RE_vec))
}

# Using POP western region as initial example
MRE_v1_fun(POP_re_results, WESTERN_Sebastes_alutus_data)
MRE_v1_fun(POP_VAST_gamma_beta4_pred_biom, WESTERN_Sebastes_alutus_data)
```

Version 2 is how CV is usually calculated:

$$\text{MRE} = \frac{1}{N} \sum_{n=1}^{N} \frac{|P_n - O_n|}{O_n}$$

```
MRE_v2_fun <- function(model_results, survey_data) {
  survey_years <- sort(unique(survey_data$YEAR))
  N <- length(survey_years) # number of survey years

  RE_vec <- abs(model_results$Biomass[model_results$Years %in% survey_years] - survey_data$AREA_BIOMASS

  MRE <- sum(RE_vec)*(1/N)

  return(list(MRE = MRE, RE_vec = RE_vec))
}

MRE_v2_fun(POP_re_results, WESTERN_Sebastes_alutus_data)
MRE_v2_fun(POP_VAST_gamma_beta4_pred_biom, WESTERN_Sebastes_alutus_data)
```