Shifts in the size of fish from a culturally important recreational fishery

Mark Scheuerell

09 October 2020

Data

The original Tengu Derby data were provided to me by Tom Quinn on 16 June 2020 in the form of an MS Excel file titled `Tengu_derby_leaders through 2019 derby.xls`. I exported one worksheet of interest (`data in kg`) as `~/data/tengu_derby_data.csv`. The same Excel file also included a worksheet with information from WDFW on the number of natural- and hatchery-origin Chinook, and the mean mass of Chinook (`Losee Chiook data`), which I exported as `~/data/wdfw_data.csv`

```
## set data dir
datadir <- here::here("data")
## import raw Tengu data
tengu_data <- readr::read_csv(file.path(datadir, "tengu_derby_data.csv"))
```

```
## Parsed with column specification:
## cols(
##   derby = col_double(),
##   year = col_double(),
##   month = col_character(),
##   days = col_double(),
##   members = col_double(),
##   total_catch = col_double(),
##   n_over_10 = col_double(),
##   n_over_5 = col_double(),
##   size_1 = col_double(),
##   size_2 = col_double(),
##   size_3 = col_double(),
##   size_4 = col_double(),
##   size_5 = col_double()
## )
```

```
## import raw WDFW data
wdfw_data <- readr::read_csv(file.path(datadir, "wdfw_data.csv"))
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
```

```
##   NOR = col_double(),
##   HOR = col_double(),
##   total = col_double(),
##   size = col_double()
## )
```

Fishing effort

The Tengu Derby dataset lacks the necessary detail to calculate a proper index of the catch per unit effort (CPUE) because effort is ill-defined. That is, although we know the total number of days per year the derby was open and the total number of anglers that participated per year, we don't know how many days *each* angler fished. Therefore, we assumed that each angler fished every day the derby was open, which is almost certainly an overestimate of the true effort, but we have no reason to believe there would be any systematic change over time.

```
## CPUE
cpue <- tengu_data$total_catch / (tengu_data$days * tengu_data$members)
```

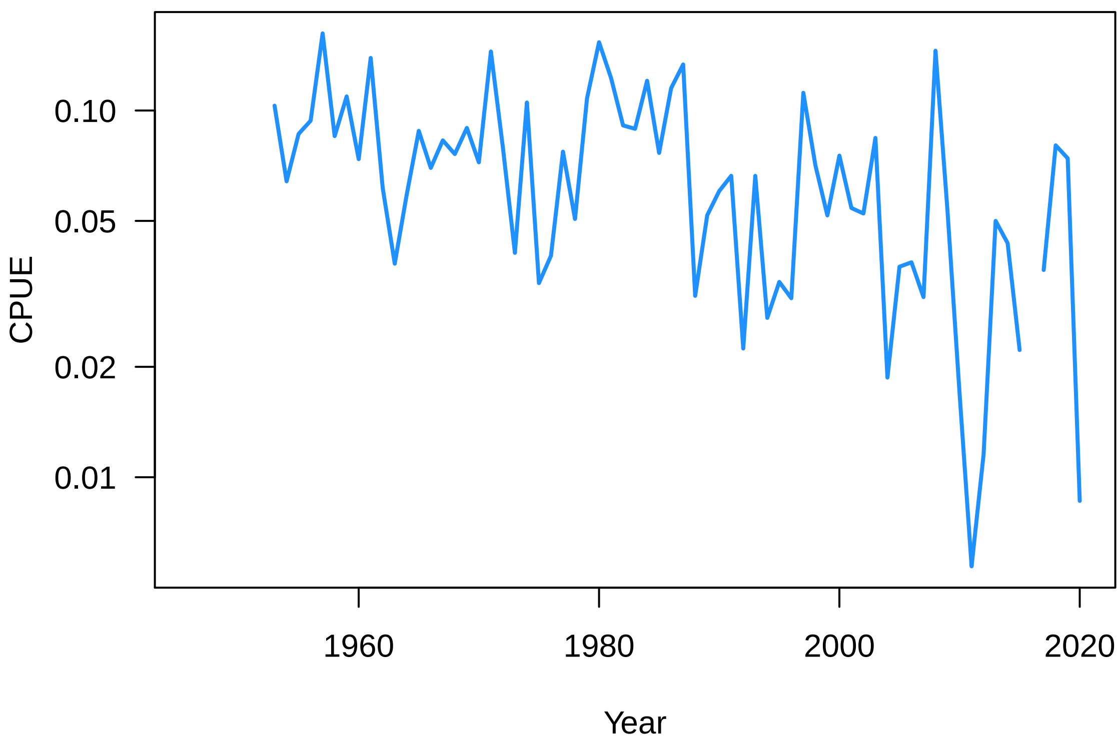

Figure 1. Time series of catch per unit effort of blackmouth from the Tengu Derby. Note that the y-axis is on a natural log scale.

Changes in catch over time

We can model CPUE data with a state-space model, such that the observed CPUE are treated as an index of the general abundance of blackmouth within this region of Puget Sound. Because the CPUE appears to generally decline over time, I fit two different state models:

1) a normal random walk; and
2) a biased random walk.

Because CPUE is necessarily bounded by zero at a minimum, I fit models to the log-transformed CPUE. Specifically, the models are defined as

$$c_t = x_t + v_t \tag{1}$$

and

$$x_t = x_{t-1} + u + w_t \tag{2}$$
$$\text{or}$$
$$x_t = x_{t-1} + w_t \tag{3}$$

where c_t is the $\log_e$-transformed CPUE, x_t is the true, but unobserved abundance of blackmouth, u is a bias term, and $v_t \sim \text{N}(0, \sigma_y)$ and $w_t \sim \text{N}(0, \sigma_c)$.

```
## model setup
mod_list <- list(
  B = matrix(1),
  U = matrix("u"),
  Q = matrix("q"),
  Z = matrix(1),
  A = matrix(0),
  R = matrix("r")
)

## response
l_cpue <- matrix(log(cpue), nrow = 1)
```

Fit biased random walk

```
## fit model with bias (Eqn 2)
cpue_brw <- MARSS(l_cpue, model = mod_list)
```

```
## Success! abstol and log-log tests passed at 435 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
```

```
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 435 iterations.
## Log-likelihood: -58.58509
## AIC: 125.1702   AICc: 125.8051
##
##        Estimate
## R.r      0.3280
## U.u     -0.0184
## Q.q      0.0000
## x0.x0   -2.0836
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

```
## 95% CI for bias
cpue_brw <- MARSSparamCIs(cpue_brw)
```

Fit unbiased random walk

```
## fit model without bias (Eqn 3)
mod_list$U <- matrix(0)
cpue_rw <- MARSS(l_cpue, model = mod_list)
```

```
## Success! abstol and log-log tests passed at 106 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 106 iterations.
## Log-likelihood: -62.53197
## AIC: 131.0639   AICc: 131.4389
##
##        Estimate
## R.r     0.31263
## Q.q     0.00873
## x0.x0 -2.36803
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

The 95% confidence interval for the bias term (u) is (-0.03, -0.01), and the model with a bias term has an AICc value ~5.6 units lower than the model without a bias term, suggesting there is indeed some data support for a systematic downward trend in the log-transformed CPUE over time.

Model fit to CPUE

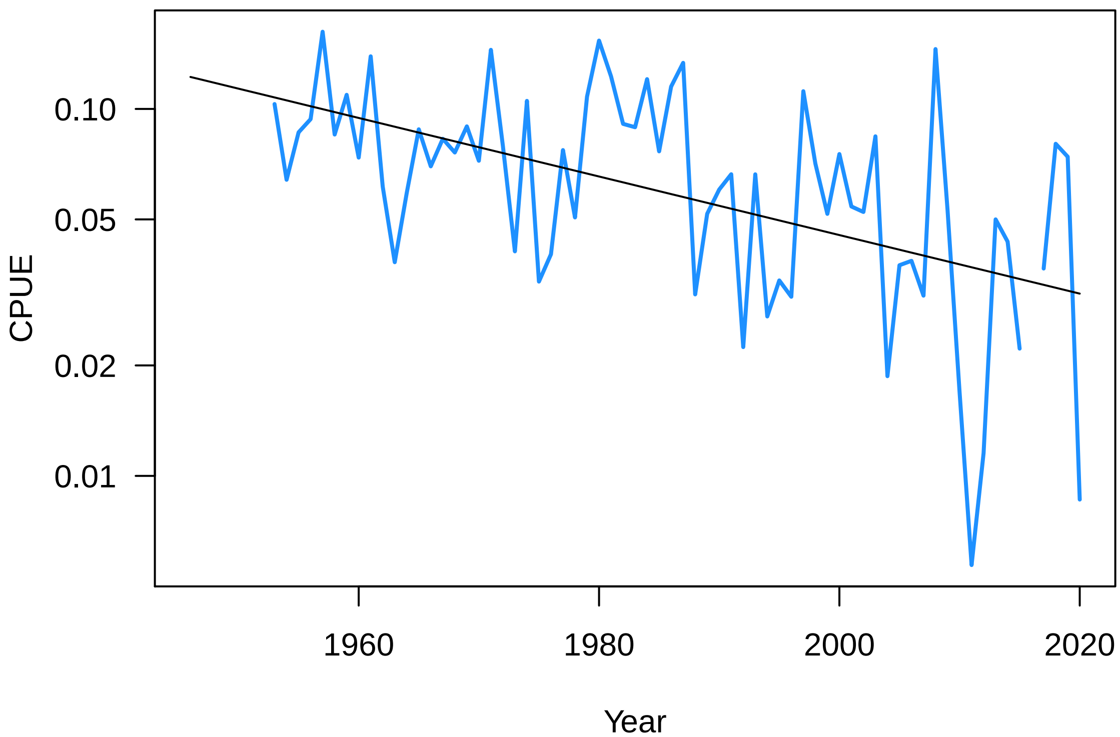

Figure 2. Time series of observed CPUE (blue) and the fit for the biased random walk model (Eqn 2; black).

Changes in fish size over time

The data set contains three different indicators of fish size over time:

1) the total number of fish over 10 pounds (~4.55 kg);
2) the total number of fish over 5 pounds (~2.27 kg); and
3) the masses (kg) of the 5 largest fish.

Clearly (1) and (2) will be correlated, as (1) is a subset of (2). To see how they relate to (3), I averaged the sizes of the 5 largest fish and compare that metric to (1) and (2). Here are the correlations among the three.

```
## total mass of 5 largest fish
mean_of_top_5 <- apply(tengu_data[,c(9:13)], 1, mean, na.rm = TRUE)
mean_of_top_5[is.nan(mean_of_top_5)] <- NA
```

```r
## 3 size metrics
fish_sizes <- cbind("N over 10 lbs" = tengu_data$n_over_10,
                    "N over 5 lbs" = tengu_data$n_over_5,
                    "Mean of top 5 (kg)" = mean_of_top_5) %>%
  ts(start = min(tengu_data$year))

## correlation among the 3 metrics
cor(fish_sizes, use = "pairwise.complete.obs") %>%
  round(2)
```

```
##                    N over 10 lbs N over 5 lbs Mean of top 5 (kg)
## N over 10 lbs               1.00         0.68               0.84
## N over 5 lbs                0.68         1.00               0.62
## Mean of top 5 (kg)          0.84         0.62               1.00
```

As suspected, all three metrics are correlated, with the total number of fish over 10 pounds and the total mass (kg) of the five largest fish being particularly so.

Here are plots of the three size metrics over time.

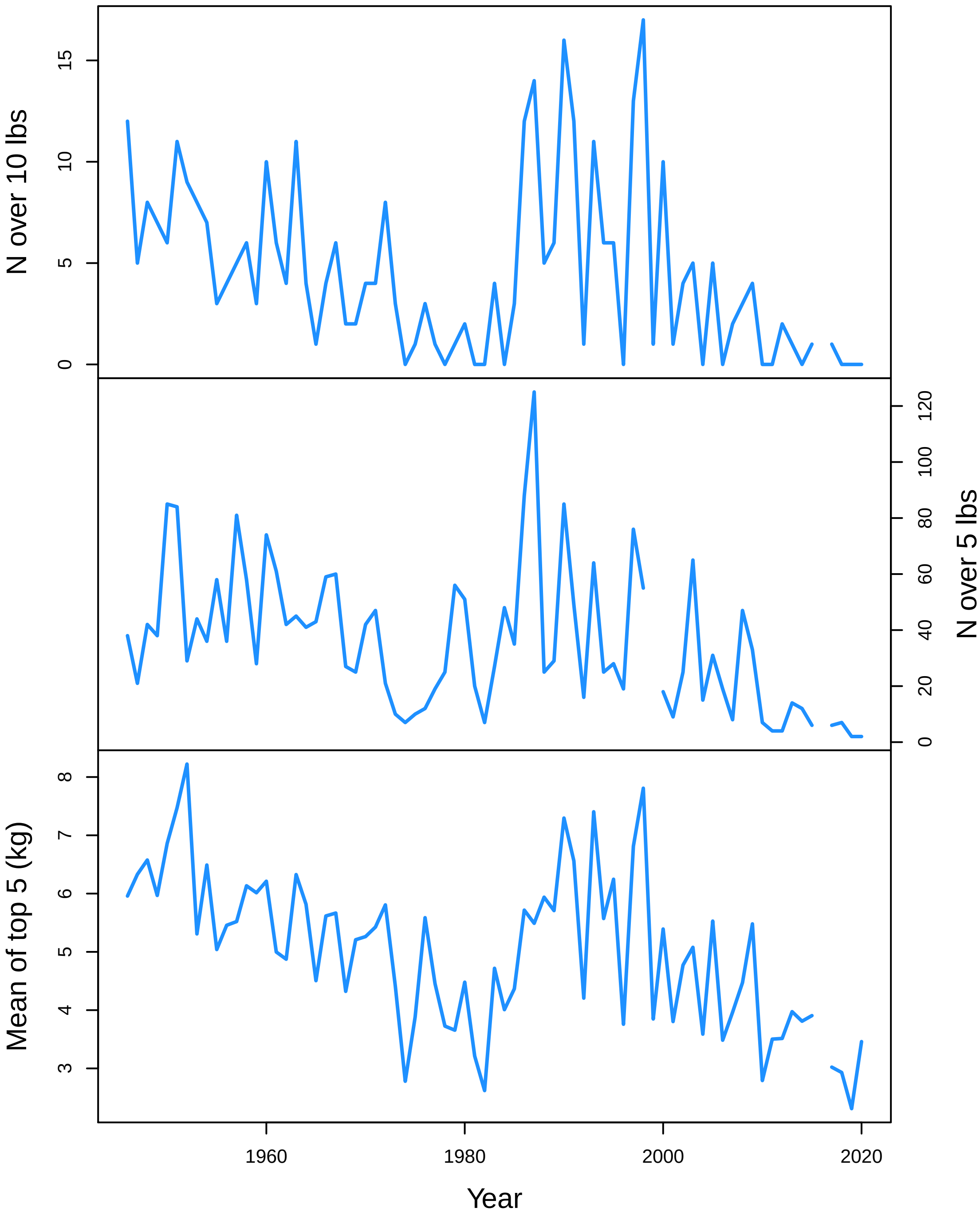

Figure 3. Time series of fish size over time, including the total number of fish over 10 pounds (top), the total number of fish over 5 pounds (middle), and the total mass (kg) of the five largest fish (bottom).

There appears to be an overall decline in fish size from the mid 1940s until the early 1980s, when fish sizes increased rapidly before declining again until present.

Fit biased random walk

Just as I did with the CPUE data, I fit both biased and unbiased forms of random walk models to the fish size data. Here, however, I used the log-transformed sum of the two largest fish, instead of the top five, because fewer than five fish were recorded in some of the later years.

```
## total mass of 5 largest fish
sum_of_top_2 <- apply(tengu_data[,c(9:10)], 1, mean, na.rm = TRUE)
sum_of_top_2[is.nan(sum_of_top_2)] <- NA

## model setup
mod_list <- list(
  B = matrix(1),
  U = matrix("u"),
  Q = matrix("q"),
  Z = matrix(1),
  A = matrix(0),
  R = matrix("r")
)
## response
l_size <- matrix(log(sum_of_top_2), nrow = 1)
```

```
## fit model with bias (Eqn 1)
size_brw <- MARSS(l_size, model = mod_list)
```

```
## Success! abstol and log-log tests passed at 53 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 53 iterations.
## Log-likelihood: 2.901809
## AIC: 2.196381   AICc: 2.776091
##
##       Estimate
## R.r    0.04099
## U.u   -0.01011
## Q.q    0.00343
## x0.x0  2.01841
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

```
## 95% CI for bias
size_brw <- MARSSparamCIs(size_brw)
```

Fit unbiased random walk

```r
## fit model without bias (Eqn 2)
mod_list$U <- matrix(0)
size_rw <- MARSS(l_size, model = mod_list)
```

```
## Success! abstol and log-log tests passed at 43 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 43 iterations.
## Log-likelihood: 1.969284
## AIC: 2.061432   AICc: 2.40429
##
##        Estimate
## R.r     0.03994
## Q.q     0.00463
## x0.x0   1.97544
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

The 95% confidence interval for the bias term (u) is (-0.024, 0.004), and the model with a bias term has an AIC value that is only ~-0.4 units lower than the model without a bias term, suggesting there is essentially no data support for a systematic downward trend in the log-transformed size over time. This is not much of a surprise, however, given the apparent temporal patterns in the data.

Model fit to size data

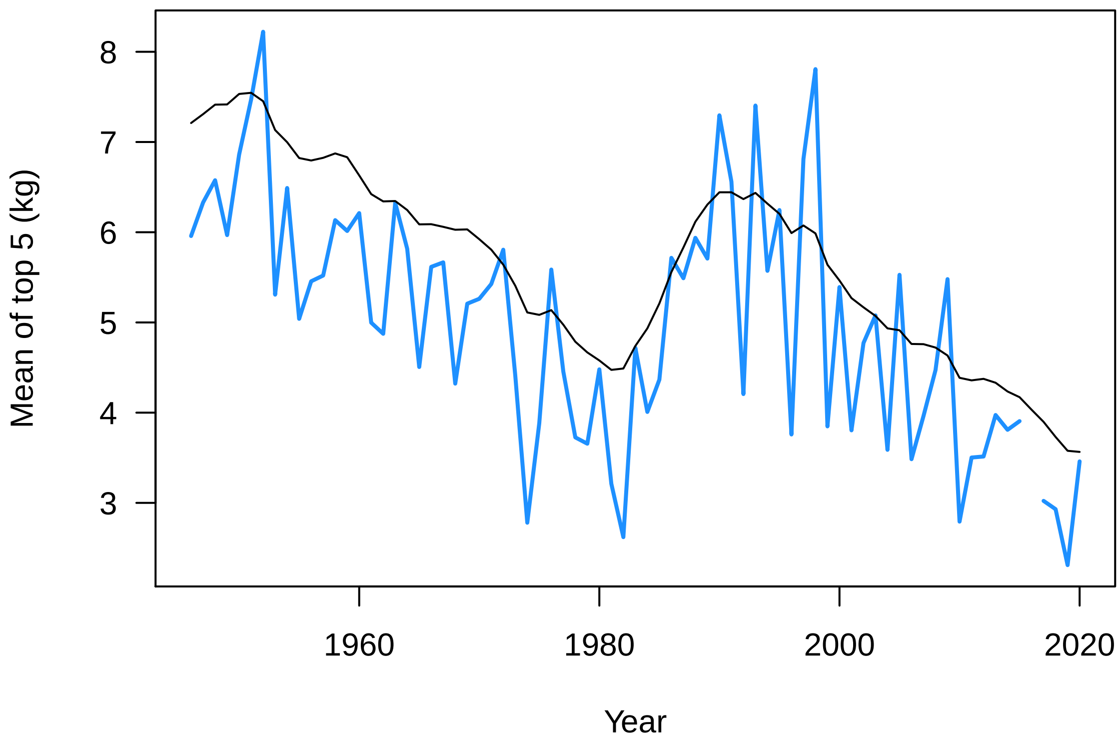

Figure 4. Time series of observed CPUE (blue) and the fit for the random walk model (Eqn 3; black).

Comparisons between the Tengu Derby & WDFW

Although the two data sources come from different times, places, and gear types, they both contain information on the temporal changes in size and CPUE over time. I investigated whether or not the temporal trends in the two data sources track one another (i.e., are representative of one "state of nature"). To do so, I used a multivariate state-space model of the general form

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t \tag{4}$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t \tag{5}$$

For both forms of the model, $\mathbf{y}_t$ is a $[2 \times 1]$ vector of the observed data from both sources, $\mathbf{a}$ is a $[2 \times 1]$ vector of offsets (intercepts), and $\mathbf{v}_t \sim \text{MVN}(\mathbf{0}, \mathbf{R})$. For both models, I assumed that the observation errors at time t ($\mathbf{v}_t$) are independent and differently distributed, such that

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$$

Fish size

The time series from WDFW begins in 1970 and runs through 2015, but the Tengy Derby data is missing size information for 2015, so I restricted my analysis to the 45 years from 1970-2014. Again I fit models to the log-transformed size data.

One pattern over time

For the model with only one state of nature, $\mathbf{Z}$ is a $[2 \times 1]$ vector of 1's, $\mathbf{x}_t$ is a $[1 \times 1]$ scalar of the true state, and $\mathbf{w}_t \sim \mathrm{N}(0, q)$, such that

$$\begin{bmatrix} y_{\text{Tengu}} \\ y_{\text{WDFW}} \end{bmatrix}_t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} a_{\text{Tengu}} \\ a_{\text{WDFW}} \end{bmatrix} + \begin{bmatrix} v_{\text{Tengu}} \\ v_{\text{WDFW}} \end{bmatrix}_t \tag{6}$$

$$x_t = x_{t-1} + w_t \tag{7}$$

```r
## select common data
yy <- cbind(tengu = sum_of_top_2[tengu_data$year >= 1970 & tengu_data$year <= 2014],
            wdfw = wdfw_data$size[wdfw_data$year >= 1970 & wdfw_data$year <= 2014])

## model defn for Eqns 6 & 7
mod_list <- list(
  B = matrix(1),
  U = matrix(0),
  Q = matrix("q"),
  Z = matrix(1, nrow = 2, ncol = 1),
  A = matrix(c("T", "W"), nrow = 2, ncol = 1),
  R = matrix(list("T", 0, 0, "W"), 2, 2)
)

## fit Eqns 6 & 7
size_both_1 <- MARSS(t(log(yy)), model = mod_list)
```

```
## Success! abstol and log-log tests passed at 74 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 74 iterations.
## Log-likelihood: 31.70933
## AIC: -51.41867   AICc: -50.40662
```

```
##
##         Estimate
## A.T   -0.174663
## A.W    0.165421
## R.T    0.071885
## R.W    0.009556
## Q.q    0.000398
## x0.x0  1.920929
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

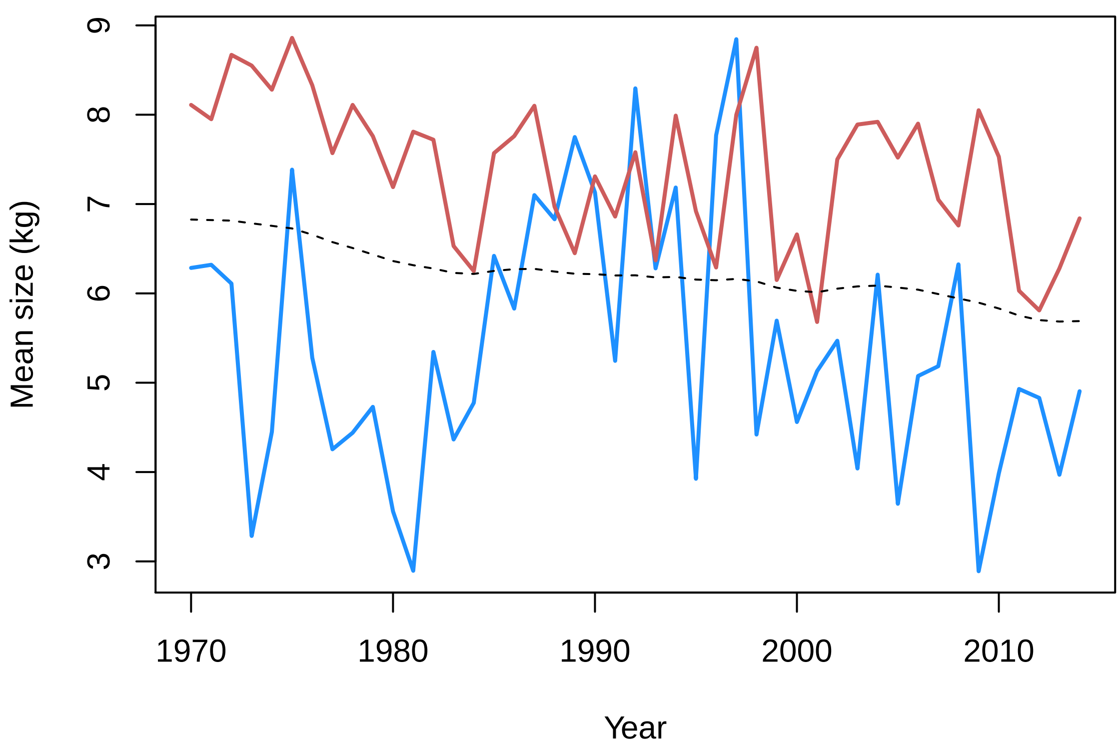

Figure 5. Time series of observed fish size from the Tengu derby (blue) and WDFW surveys (red), including fits from the multivariate random walk model (Eqns 6 & 7; dashed).

Two patterns over time

For the model with two different states of nature, $\mathbf{Z}$ is a $[2 \times 2]$ identity matrix, $\mathbf{x}_t$ is a $[2 \times 1]$ vector of the true states, and $\mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$, such that

$$
\begin{bmatrix} y_{\text{Tengu}} \\ y_{\text{WDFW}} \end{bmatrix}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{Tengu}} \\ x_{\text{WDFW}} \end{bmatrix}_t + \begin{bmatrix} a_{\text{Tengu}} \\ a_{\text{WDFW}} \end{bmatrix} + \begin{bmatrix} v_{\text{Tengu}} \\ v_{\text{WDFW}} \end{bmatrix}_t \tag{8}
$$

$$
\begin{bmatrix} x_{\text{Tengu}} \\ x_{\text{WDFW}} \end{bmatrix}_t = \begin{bmatrix} x_{\text{Tengu}} \\ x_{\text{WDFW}} \end{bmatrix}_{t-1} + \begin{bmatrix} w_{\text{Tengu}} \\ w_{\text{WDFW}} \end{bmatrix}_t \tag{9}
$$

```
## model defn for Eqns 8 & 9
mod_list <- list(
  B = diag(2),
  U = matrix(0, nrow = 2, ncol = 1),
  Q = matrix(list("T", 0, 0, "W"), 2, 2),
  Z = diag(2),
  A = matrix(c("T", "W"), nrow = 2, ncol = 1),
  R = matrix(list("T", 0, 0, "W"), 2, 2)
)

## fit Eqns 8 & 9
size_both_2 <- MARSS(t(log(yy)), model = mod_list)
```

```
## Success! abstol and log-log tests passed at 118 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 118 iterations.
## Log-likelihood: 33.44122
## AIC: -50.88244   AICc: -49.10466
##
##              Estimate
## A.T         -0.012675
## A.W         -0.001470
## R.T          0.052382
## R.W          0.009137
## Q.T          0.004164
## Q.W          0.000477
## x0.X.tengu   1.704472
## x0.X.wdfw    2.098912
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

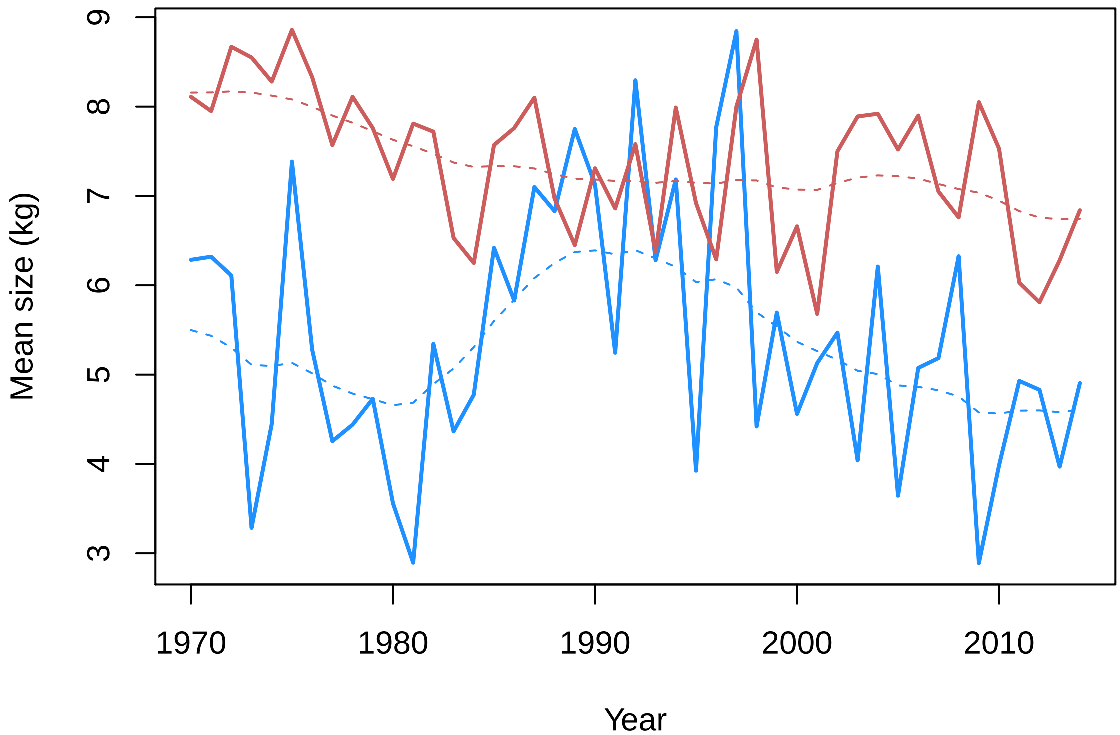

Figure 6. Time series of observed fish size from the Tengu derby (blue) and WDFW surveys (red), including fits from the multivariate random walk model (Eqns 8 & 9; dashed).

Summary of size comparison

The model with one common state has an AICc value of -50.4 and the model with two unique states has an AICc value of -49.1, which indicates rather modest support for two unique temporal patterns in the data.

CPUE

The time series of CPUE from WDFW begins in 1968 and runs through 2015, but the Tengy Derby data is missing catch information for 2015, so I restricted my analysis to the 47 years from 1968-2014. Because the units for the catch data differ for the two datasets and vary by several orders of magnitude, I log-transformed each of the time series and then normalized them to have mean zero and unit variance.

One pattern over time

Here is the model for one common temporal trend, analogous to Equations 6 & 7 above.

```
## select common data
yy <- cbind(tengu = cpue[tengu_data$year >= 1968 & tengu_data$year <= 2015],
            wdfw = wdfw_data$total[wdfw_data$year >= 1968 & wdfw_data$year <= 2015])
```

```r
## standardize the mean and variance
yy <- scale(log(yy))

## model defn for Eqns 6 & 7
mod_list <- list(
  B = matrix(1),
  U = matrix(0),
  Q = matrix("q"),
  Z = matrix(1, nrow = 2, ncol = 1),
  A = matrix(c("T", "W"), nrow = 2, ncol = 1),
  R = matrix(list("T", 0, 0, "W"), 2, 2)
)

## fit Eqns 6 & 7
cpue_both_1 <- MARSS(t(yy), model = mod_list)
```

```
## Success! abstol and log-log tests passed at 33 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 33 iterations.
## Log-likelihood: -130.8703
## AIC: 273.7407   AICc: 274.6952
##
##       Estimate
## A.T   -0.01198
## A.W    0.00618
## R.T    0.69813
## R.W    0.72203
## Q.q    0.10310
## x0.x0  0.28467
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

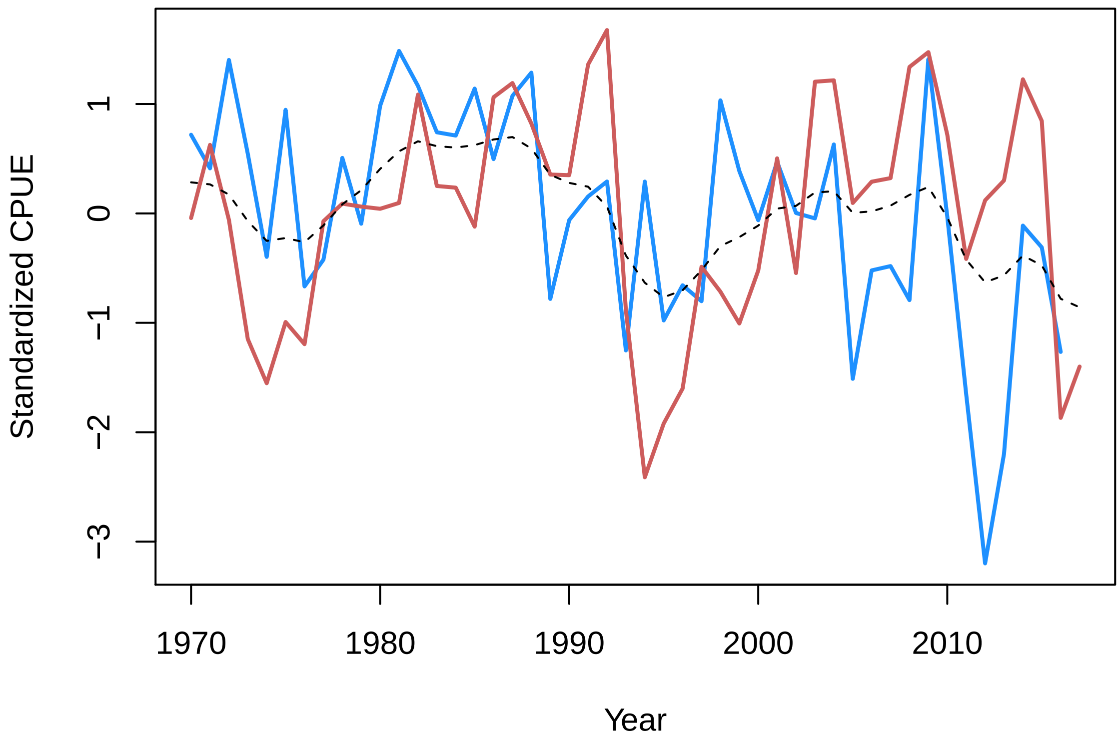

Figure 7. Time series of observed CPUE from the Tengu derby (blue) and WDFW surveys (red), including fits from the multivariate random walk model (Eqns 6 & 7; dashed).

Two patterns over time

```
## model defn for Eqns 8 & 9
mod_list <- list(
  B = diag(2),
  U = matrix(0, nrow = 2, ncol = 1),
  Q = matrix(list("T", 0, 0, "W"), 2, 2),
  Z = diag(2),
  A = matrix(c("T", "W"), nrow = 2, ncol = 1),
  R = matrix(list("T", 0, 0, "W"), 2, 2)
)

## fit Eqns 8 & 9
cpue_both_2 <- MARSS(t(yy), model = mod_list)
```

```
## Success! abstol and log-log tests passed at 292 iterations.
## Alert: conv.test.slope.tol is 0.5.
## Test with smaller values (<0.1) to ensure convergence.
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Estimation converged in 292 iterations.
```

```
## Log-likelihood: -124.0529
## AIC: 264.1058   AICc: 265.7802
##
##              Estimate
## A.T         -0.020799
## A.W          0.001867
## R.T          0.653150
## R.W          0.056181
## Q.T          0.033660
## Q.W          0.680011
## x0.X.tengu   0.551907
## x0.X.wdfw   -0.000225
## Initial states (x0) defined at t=0
##
## Standard errors have not been calculated.
## Use MARSSparamCIs to compute CIs and bias estimates.
```

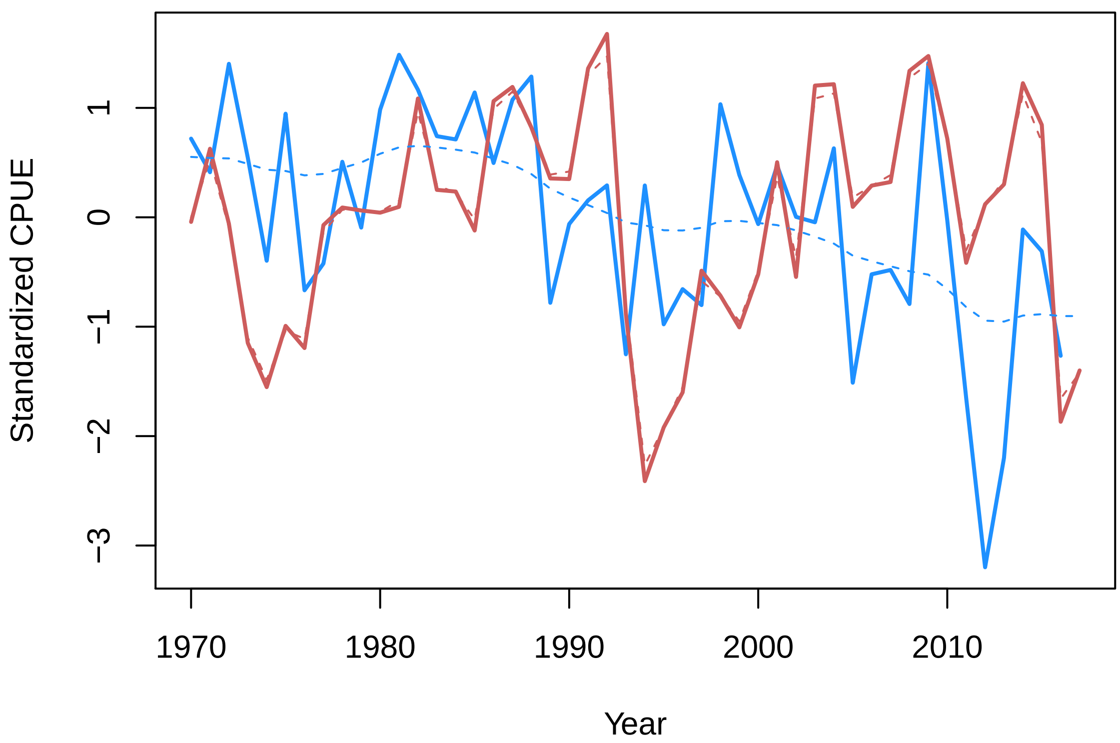

Figure 8. Time series of observed fish CPUE from the Tengu derby (blue) and WDFW surveys (red), including fits from the multivariate random walk model (Eqns 8 & 9; dashed).

Summary of CPUE comparison

The model with one common state has an AICc value of 274.7 and the model with two unique states has an AICc value of 265.8, which provides reasonable support for two unique temporal patterns in the data.