Westward Large-Mesh Trawl Survey Abundance Estimation, Including Missing Data

Tyler Jackson
Alaska Department of Fish and Game, Commercial Fisheries Division, Kodiak, Alaska

1 Purpose

This document details a method for estimating Westward Large-Mesh Survey Tanner crab abundance that is robust to missing stations within a given year, without using the 'rolling block' ANOVA.

2 Estimation

The survey most closely represents a stratified design in which survey stations of varying area are individual strata, nested within subsections, sections, and districts - though subsections are not relevant for each district (more below). Tanner crab abundance within a section ($\hat{N}_\tau$) is estimated as

$$\hat{N}_\tau = \sum_{i=1}^{n} U_i A_i \tag{1}$$

where U is catch per unit effort (CPUE, or density) expressed as crab per square nautical mile, and A is area (nmi^2) for station i of n. Note, CPUE for this survey is often reported as crab per kilometer towed, but for the purpose of this document and estimation method it represents area-swept crab density. At face value, Equation 1 does not allow for missing stations, which confer and abundance of zero. However a simple algebraic rearrangement of Equation 1 would result in the same estimation mathematically, allow for missing stations, and provide an unbiased estimator of section (or subsection) CPUE:

$$\hat{N}_\tau = \bar{U} \sum_{i=1}^{n} A_i \tag{2}$$

where $\bar{U}$ is the unbiased estimator of section (or subsection) CPUE (i.e., the mean CPUE weighted by station area) estimated as

$$\bar{U} = \frac{\sum_{i=1}^{h} U_i A_i}{\sum_{i=1}^{h} A_i} \tag{3}$$

Here, $\bar{U}$ is estimated only using stations that were towed within the given survey year, so that $h \leq n$, but is then scaled to abundance using the summed area of **all** stations with the section (or subsection). This method achieves the same result mathematically since the rearrangement of Equation 1 comes down 1) bringing U outside of the summation, and 2) multiplying it by $\frac{\sum_{i=1}^{h} A_i}{\sum_{i=1}^{h} A_i} = 1$.

3 Comparison to Rolling Block ANOVA

Mature male and mature female abundance estimates where compared between this 'design based' method and those computed using the rolling block ANVOA.

Nearly all annual survey estimates among sections or subsections are very similar among methods - at least well within the level of precision possible with the survey design. This is expected since the rolling block anova employs a similar procedure, that the mean CPUE computed for a missing station is estimated from a subset of 10 stations, whereas the design based approach uses all stations within a district. Since the design based estimation method represents the unbiased estimator for a stratified design, I feel it is more defensible than selecting the subset of 10 stations based on previous years data. Obviously, neither method is appropriate when only an insufficient number of stations - or no stations - within a section are surveyed. To be clear, the design based method is appropriate for any size class of crab, or any other animal.

4 R workflow

The R script used in timeseries estimation is available on the ADF&G Commercial Fisheries GitHub repository,`westward_goa_crab`, https://github.com/commfish/westward_goa_crab. The two main survey datasets used are the haul summary dump from 1988 - 2021 and the all animal dump from 1988 - 2021 (the word 'catch' in the animal dump file name was changed to specimen to avoid masking the actual catch data dump which produces the same file name). I also included a lookup table csv dataset called 'tanner_standard_stations' that includes a list of all standard survey stations, the necessary predecessor stations, district, section, subsection, and area.

Survey estimates for stations not spanning the full timeseries were estimated by averaging the CPUE from available predecessor stations (called 'reference' stations in my R script because I have a hard time spelling predecessor). The function `f_timeseries_station_correct` carries out this procedure in 3 steps: 1) it creates a data frame of all combination of survey years and standard stations, 2) it identifies if a standard station was towed within a year, and then if not 3) averages among reference stations if applicable. Corrected CPUE estimates are then expanded as above. Note that in years without any missing data, estimates may be slightly off due to the number of significant digits used in estimation, since data were not pull directly from the database (e.g. sample fraction).

Survey estimates that used the rolling block ANOVA procedure to fill missing stations were pulled from the provided Excel spreadsheets, and combined into a single csv file.

5 Figures

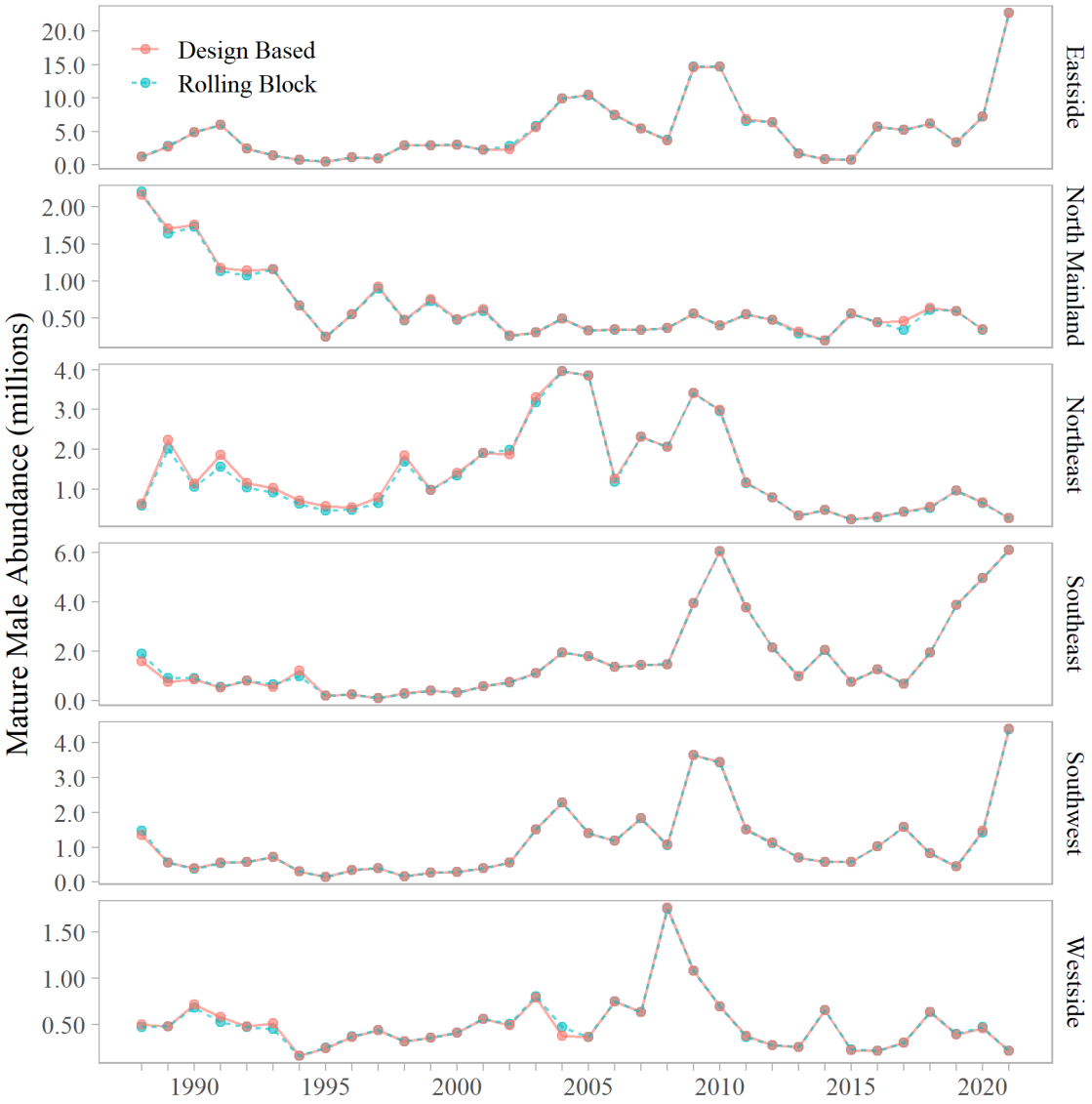

Figure 1: Timeseries of mature male abundance by section in the Kodiak district for the rolling block (blue) and design based (pink) method of estimation.

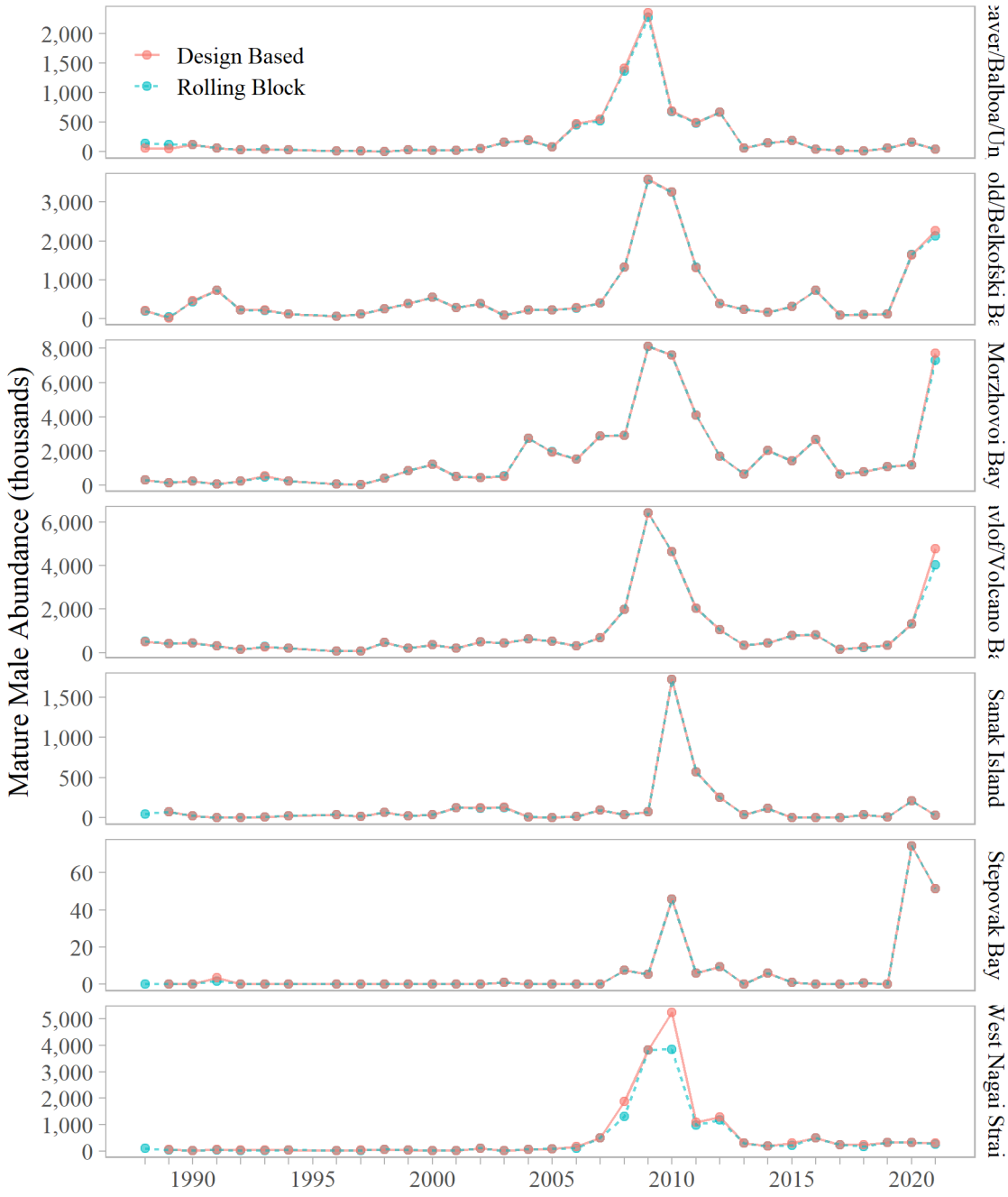

Figure 2: Timeseries of mature male abundance by subsection in the South Peninsula district for the rolling block (blue) and design based (pink) method of estimation.

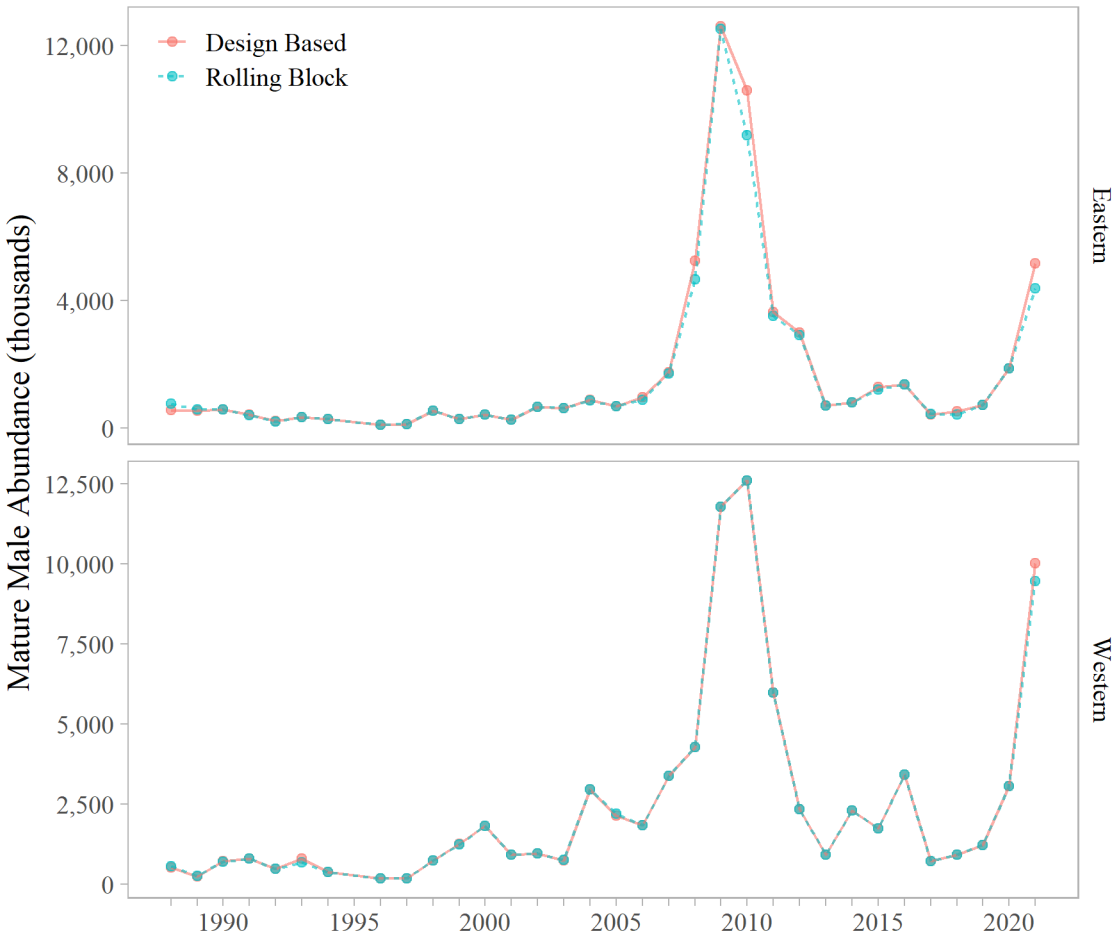

Figure 3: Timeseries of mature male abundance by section in the South Peninsula district for the rolling block (blue) and design based (pink) method of estimation.

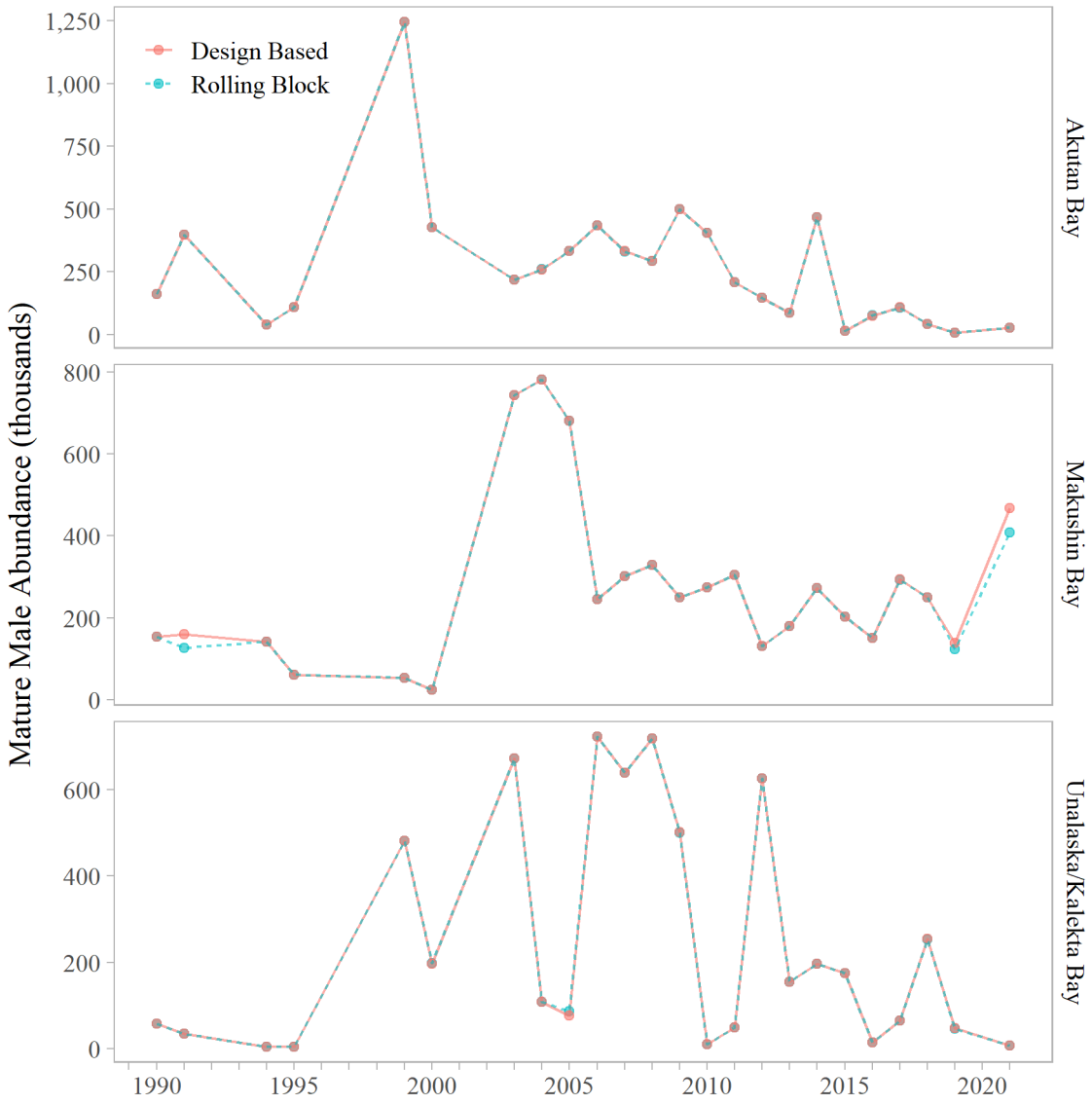

Figure 4: Timeseries of mature male abundance by section in the Eastern Aleutian district for the rolling block (blue) and design based (pink) method of estimation.

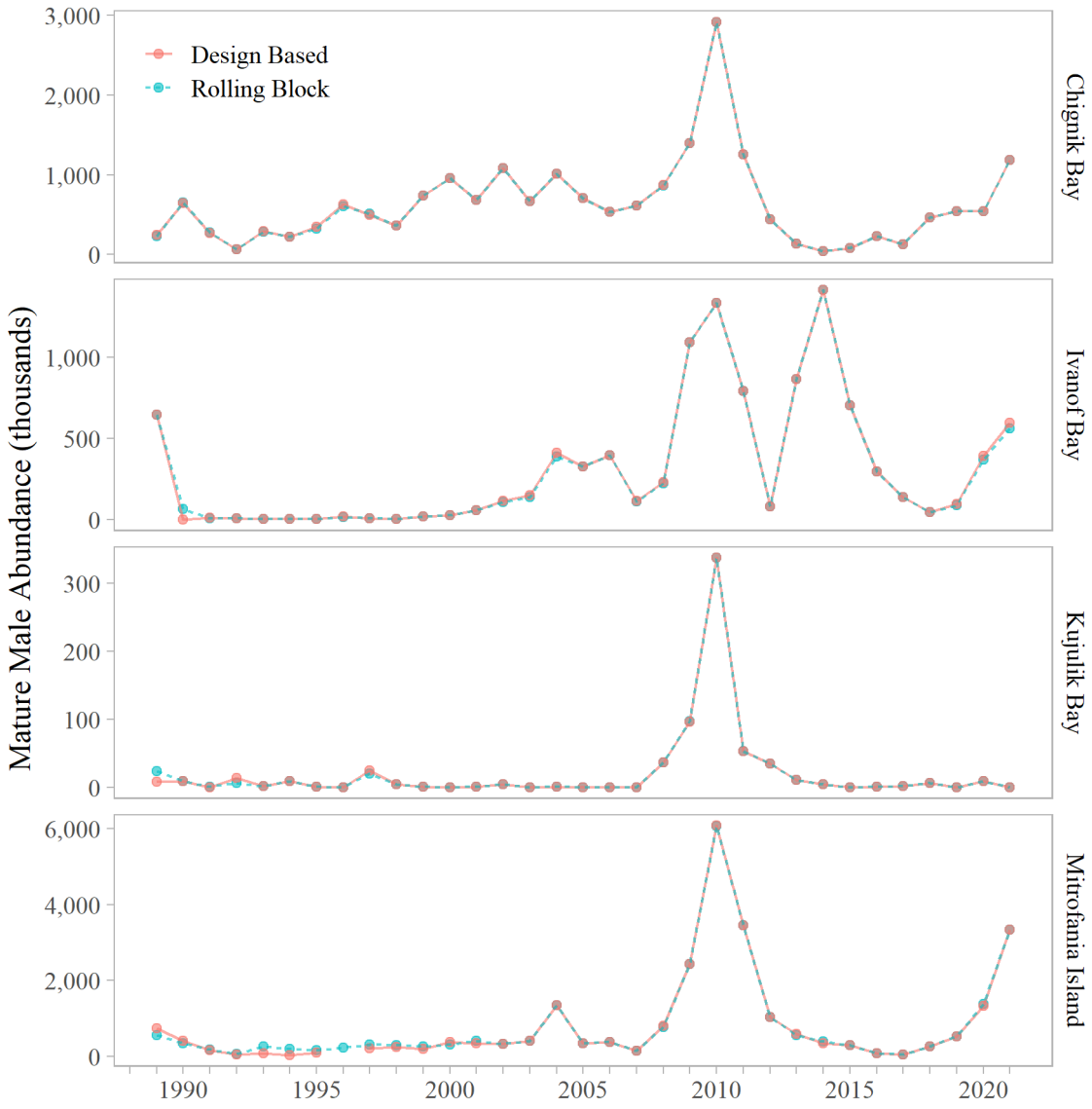

Figure 5: Timeseries of mature male abundance by section in the Chignik district for the rolling block (blue) and design based (pink) method of estimation.

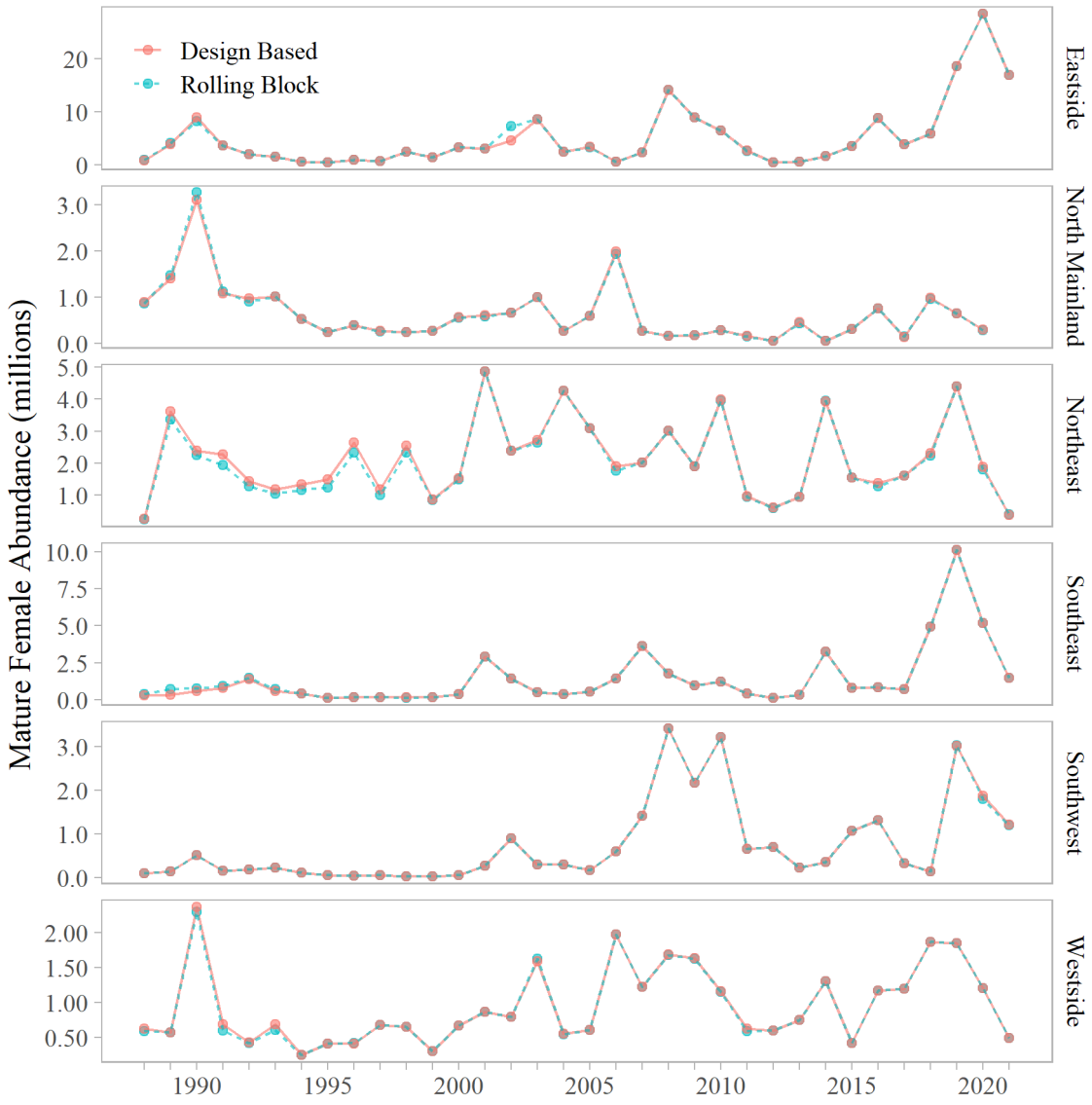

Figure 6: Timeseries of mature female abundance by section in the Kodiak district for the rolling block (blue) and design based (pink) method of estimation.

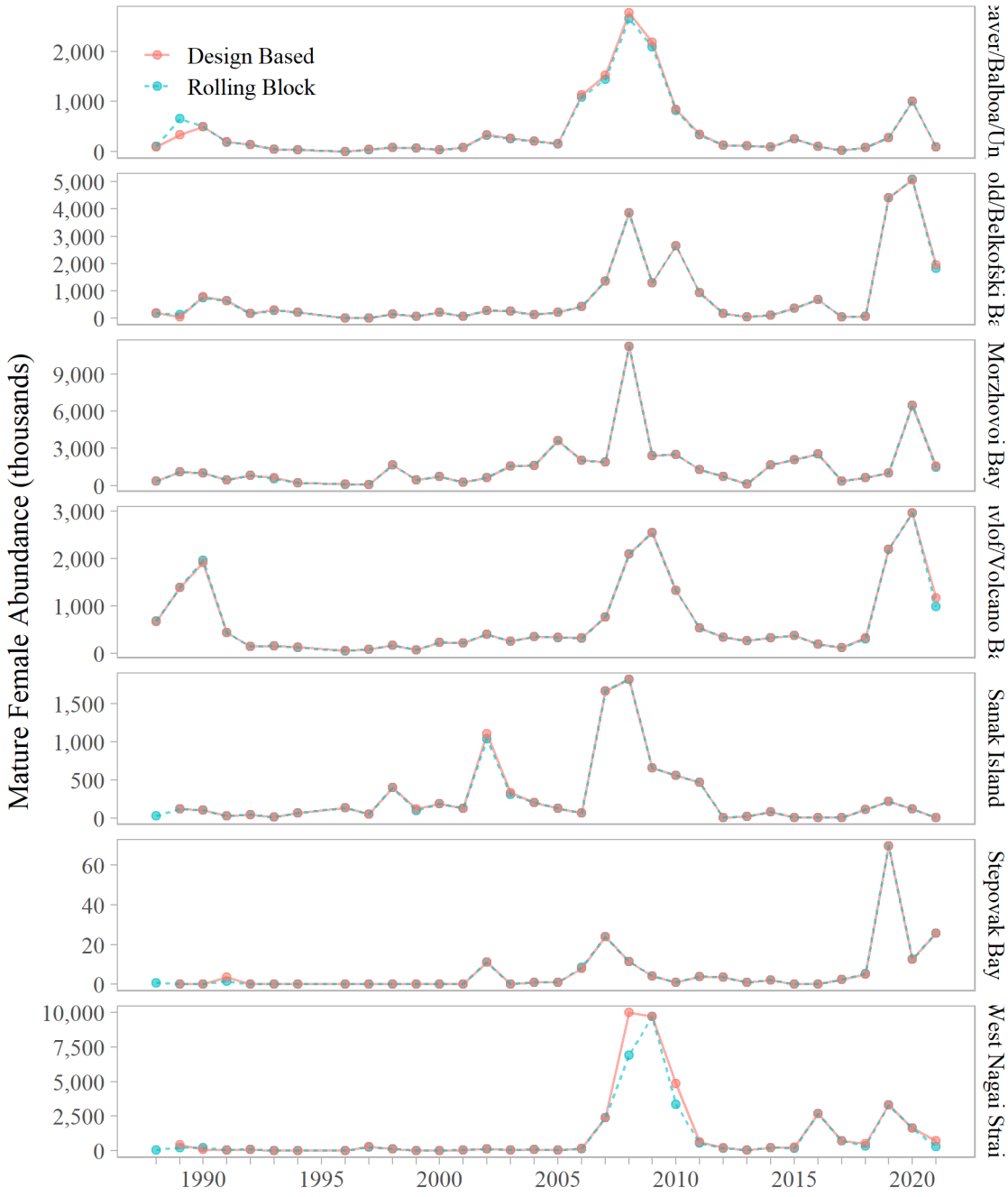

Figure 7: Timeseries of mature female abundance by subsection in the South Peninsula district for the rolling block (blue) and design based (pink) method of estimation.

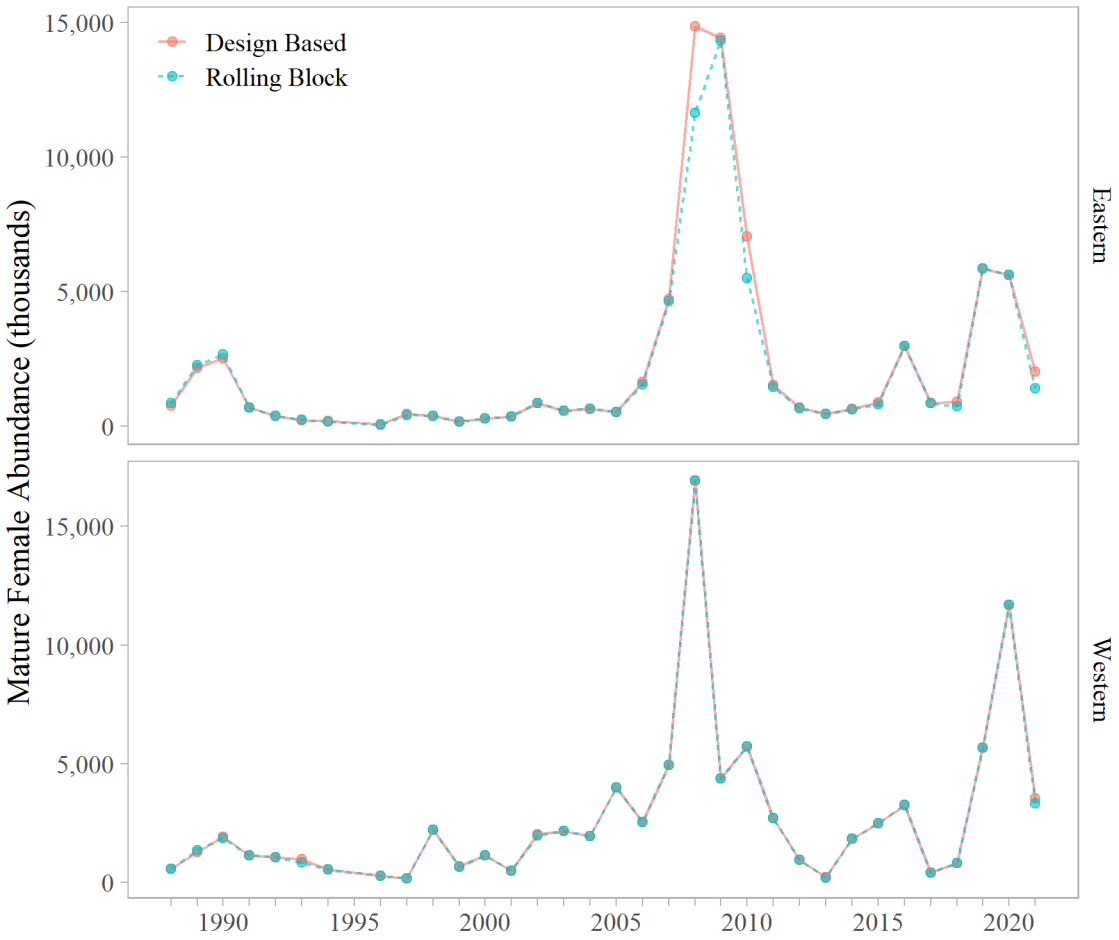

Figure 8: Timeseries of mature female abundance by section in the South Peninsula district for the rolling block (blue) and design based (pink) method of estimation.

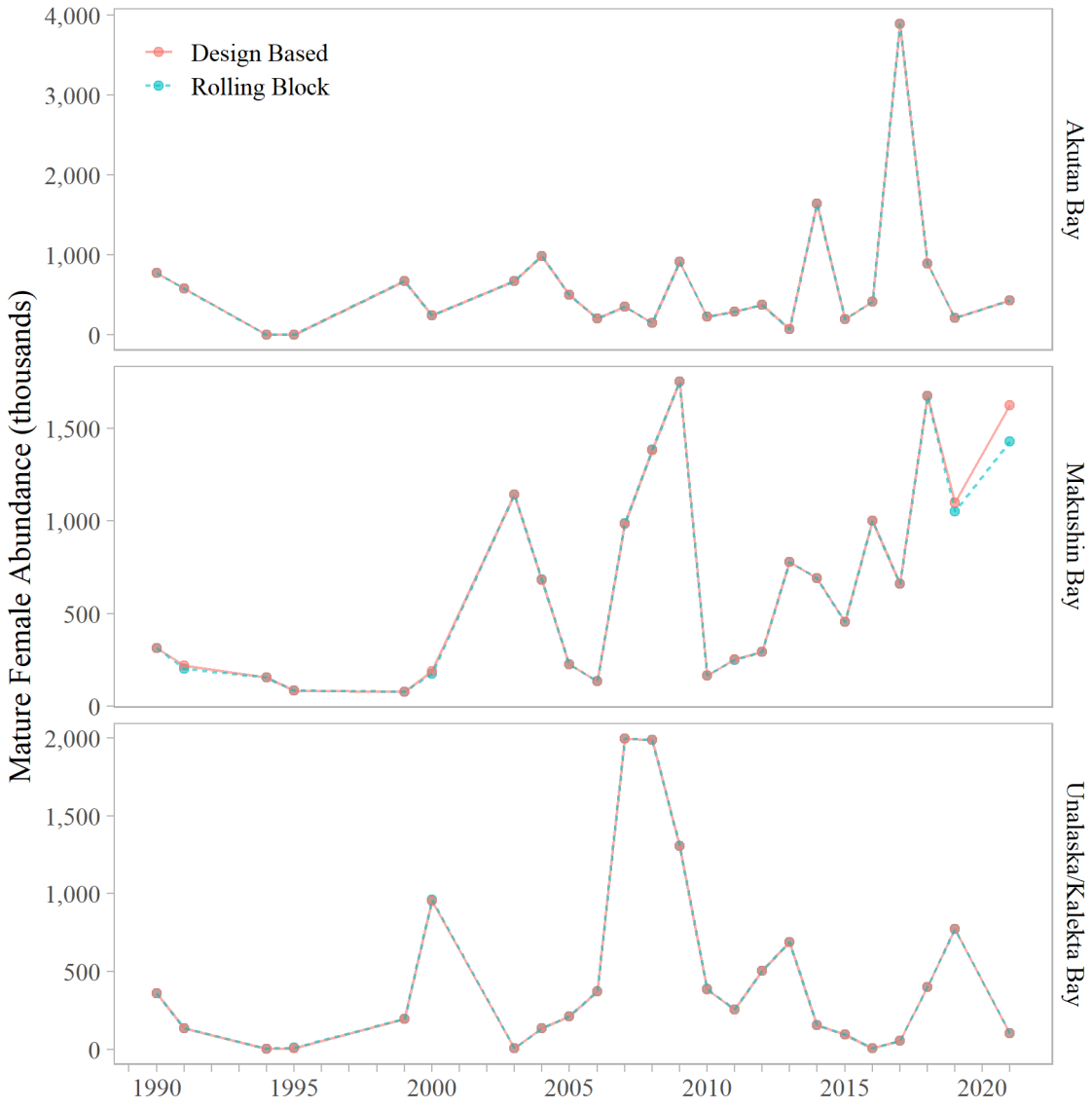

Figure 9: Timeseries of mature female abundance by section in the Eastern Aleutian district for the rolling block (blue) and design based (pink) method of estimation.

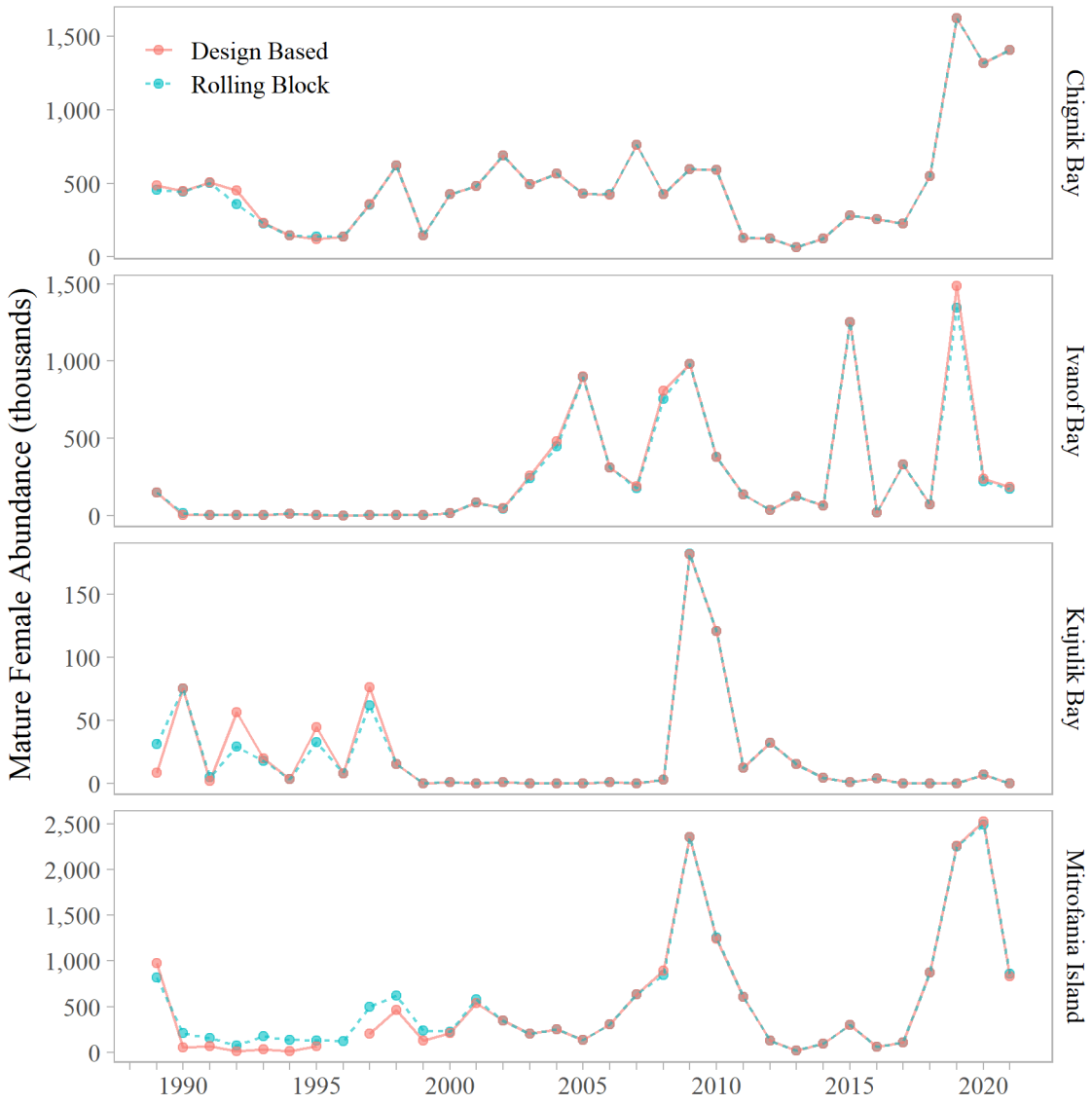

Figure 10: Timeseries of mature female abundance by section in the Chignik district for the rolling block (blue) and design based (pink) method of estimation.

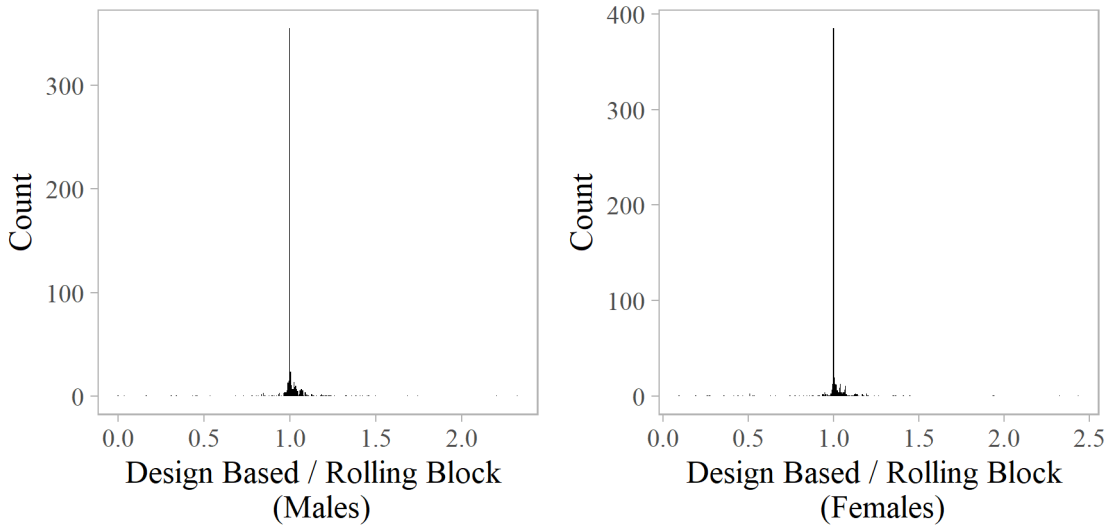

Figure 11: Histograms of the proportion difference between the design based estimate and that using the rolling block ANOVA to impute missing data among all survey years and sections for mature males (left) and mature females (right)