Data Deficiencies - Russian data

06/04/2023

Authors:

James Ounsley, Etienne Rivot, Geir Bolstad, Hugo Maxwell, Jonathan Gillson, Alan Walker

Introduction

In the absence of data from the Russian Federation being reported to ICES for 2021 or 2022, the Working Group investigated alternative published sources of data and developed an approach to make those data usable for the assessment model.

The national total catch weights for fisheries in coastal waters, estuaries and in-river, the numbers of salmon caught and released, and this number expressed as a percentage of the total catch retained and released, are annually reported to NASCO in the Russian Federation's Annual Progress Report (APR). These reports are published on the NASCO website (at https://nasco.int/conservation/third-reporting-cycle-2/) and therefore the Working Group used these data to collate catch summaries for the North Atlantic, as reported in section 2 of the Working Group report, and the draft 'sal.other.all' advice.

In addition, however, the Working Group requires catch numbers by stock unit (4 stock units considered in Russia) and sea age class, to conduct the pre-fishery abundance and run reconstruction analyses. Data disaggregated to these levels are not reported to NASCO and therefore the Working Group developed an approach to derive estimated values for 2021 and 2022. The following text describes that approach, considers the strengths and weaknesses of this approach, makes suggestions for alternative approaches that might be examined in the future, and outlines issues with all of these.

Absence of Russian data

There are four regional stock units (SU) within Russia: Pechora River (RP), Archangel / Karelia (AK), Kola / White Sea (KW) and Kola / Barents Sea (KB). This split in the Russian stock is based on biological characteristics and the resolution of catch statistics reporting.

For each of the four SU, the NEAC Run Reconstruction model requires the following annual input data: catches by sea age (and additionally catches on delayed spawners for KW); declared returns for RP by sea age; exploitation rates and associated error by SU and sea age and unreported catch rates and associated error by SU and sea age.

WGNAS agreed upon an approach for accounting for the deficiency by constructing estimated values for the affected years (2021 and 2022) based on a set of assumptions given historic data.

Exploitation rates and unreported catch rates

For all four regional stock units, the exploitation rates and unreported catch rates, together with their associated errors, have been unchanged for at least the last ten years for which they have been provided to WGNAS. These values were assumed unchanged for the 2021 and 2022 stock years.

Estimating the catch

For the three stock units AK, KB and KW, the estimated catches for 2021 and 2022 were based on the five years mean of the most recent reported catches (i.e. catches for the period 2016 to 2020). Total catches for the entire Russian stock are available for 2021 and 2022 (NASCO, 2023), and provide information on the aggregate trend in catches at the country level. This information is incorporated into the estimated catches by scaling the five years mean for each stock unit by the relative change in catches observed in the total catch between 2021 and 2022 and the five year mean of total catch for the period 2016 to 2020.

Given total catch for Russia T_y in years $y = 2021{,}2022$, we derive the scaling factor α_y for year y as follows.

$$\alpha_y \quad = \frac{T_y}{\frac{1}{5}\sum_{i=2016}^{2020} T_i}.$$

The catches $C_{s,a,y}$ for each stock unit (s) and sea age (a) are then estimated by:

$$C_{s,a,y} \quad = \alpha_y \frac{1}{5} \sum_{i=2016}^{2020} C_{s,a,i}\,.$$

For RP, the declared returns $R_{a,y}^{\text{dec}}$ were estimated using the same method

$$R_{a,y}^{\text{dec}} = \alpha_y \frac{1}{5} \sum_{i=2016}^{2020} R_{a,i}^{\text{dec}}.$$

The resultant estimates can be seen in Figure 1.

The catches on delayed spawners in KW for 2021 and 2022 were estimated using the same approach. These values are used in the derivation of spawners for this region, but are not influential on the variance in any of the derived values of the NEAC run-reconstruction and are not considered in the following analysis.

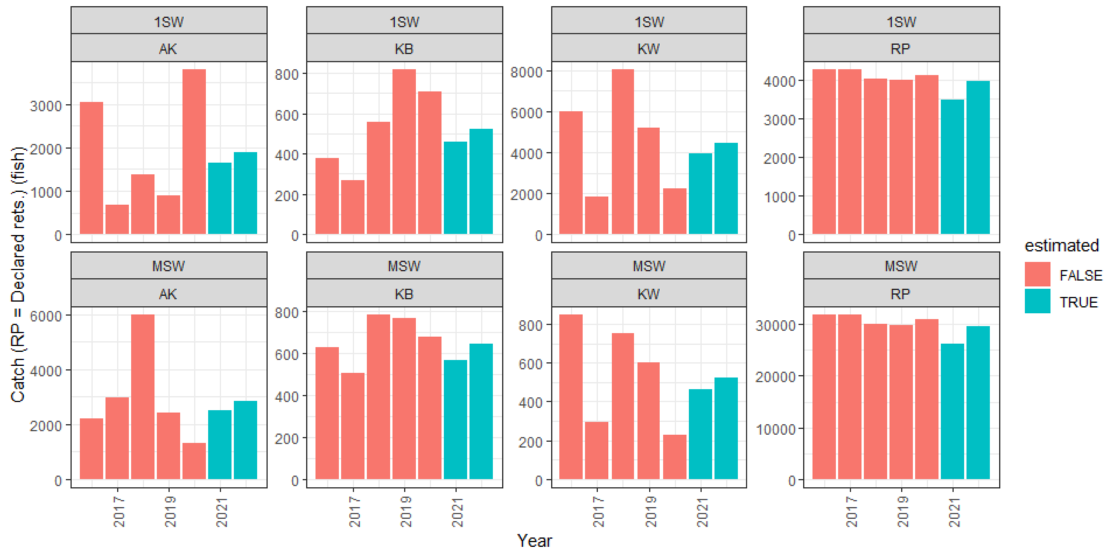

Figure 1. Reported (2016 - 2020) and estimated (2021, 2022) catches for the regional stock units AK, KB and KB and declared returns for RP

Accounting for additional uncertainty

The NEAC run-reconstruction model uses the catches (and declared returns for RP) to derive returns to home-waters and thereafter spawning abundances and PFA. Uncertainty in these values is introduced by integrating over uncertainty in the exploitation rates and unreported catch rates when deriving returns to home-waters. Uncertainty is integrated out using Monte Carlo numerical simulations.

For AK, KB & KW, the returns ($R_{s,a,y}$) are derived as follows

$$\begin{aligned} R_{s,a,y} &= \frac{C_{s,a,y}}{E_{s,a,y}\left(1 - U_{s,a,y}\right)} \\ E_{s,a,y} &\sim \text{unif}\left(\mu^E_{s,a,y} - \epsilon^E_{s,a,y}, e + \epsilon^E_{s,a,y}\right) \\ U_{s,a,y} &\sim \text{unif}\left(\mu^U_{s,a,y} - \epsilon^U_{s,a,y}, e + \epsilon^U_{s,a,y}\right). \end{aligned}$$

Where $E_{s,a,y}$ and $U_{s,a,y}$ are the distributions of the exploitation rates and unreported catch rates as defined by uniform distributions defined by their respective means $(\mu^E_{s,a,y}, \mu^U_{s,a,y})$ and half range $(\epsilon^E_{s,a,y}, \epsilon^U_{s,a,y})$.

For RP, the returns are defined as

$$
\begin{aligned}
R_{s,a,y} &= R_{a,y}^{dec}(1+U) \\
U_{s,a,y} &\sim \mathrm{unif}\left(\mu_{s,a,y}^{U} - \epsilon_{s,a,y}^{U}, e + \epsilon_{s,a,y}^{U}\right).
\end{aligned}
$$

To account for the fact that the catches (or declared returns for RP) were not available as data in 2021 and 2022 but first derived from the total catch in weight based on the method describe above, an approach was developed to scale up the variance of the probability distribution of the returns (and by extension spawner abundances and PFA) for 2021 and 2022. This was to ensure that the confidence intervals around the returns estimates for these years were more likely to include the mean value of the returns based on the true data had it been available (Figure 2, a and b).

Let $R_{l,y}^{est}$ be the estimated returns for year y with lag l. The lag defines the number of years since the year of the most recent reported data used in the derivation of the estimated catches (or declared returns for RP). For example, for 2021 with a lag of $l = 1$, the returns are relative to the five years average of data for 2016 to 2020, and for 2022 with a lag of $l = 2$, the returns are relative to the five years average for the same period. For clarity, the following derivation is for a single stock unit and age class. The derivation is the same of all four stock units and sea ages.

Let $R_{l,y}^{adj}$ be the adjusted returns after scaling up the variance of the estimated returns. The variance of the log returns can be scaled by multiplying the centered log returns by some scaling factor q as follows:

$$
\ln\left(R_{l,y}^{adj}\right) = q_l\left(\ln\left(R_{l,y}^{est}\right) - E\left[\ln\left(R_{l,y}^{est}\right)\right]\right) + E\left[\ln\left(R_{l,y}^{est}\right)\right]
$$

such that

$$
\mathrm{var}\left[\ln R_{l,y}^{adj}\right] = \mathrm{var}\left[q_l \ln R_{l,y}^{est}\right] = {q_l}^2 \mathrm{var}\left[\ln R_{l,y}^{est}\right]
$$

To capture the additional uncertainty resulting from the use of estimated data, it remains to find the scaling factor q such that

$$
\mathrm{var}\left[\ln R_{l,y}^{adj}\right] = r_l + \gamma_l
$$

where γ_l is the expected variance of the estimated log-returns and r_l is the expected mean squared error between the estimated log-returns and the observed log returns, i.e. the returns derived from observed catches (or declared returns for RP), denoted R_y^{obs} .

The required adjustment of the variance is then given by

$$
q_l = \sqrt{\frac{r_l + \gamma_l}{\gamma_l}}.
$$

In the absence of R_y^{obs} for the years 2021 and 2022, a one step ahead "cross-validation" approach was developed to numerically quantify the expected r_l and γ_l, denoted $\widehat{r_l}$ and $\widehat{\gamma_l}$, based on R_y^{obs} for the y in 2016 to 2020 and $R_{l,y}^{est}$ derived for the same y and for l=1 and l=2. Giving the numerically estimated $\widehat{q_l}$

$$\widehat{q_l} = \sqrt{\frac{\widehat{r_l} + \widehat{\gamma_l}}{\widehat{\gamma_l}}}.$$

Thus, $\widehat{r_l}$ was calculated as the mean of the squared difference between the means of the estimated and observed returns on the log scale, calculated over a 5 years window:

$$\widehat{r_l} = \frac{1}{5}\sum_{y=2016}^{2020}\left(E\left[\ln\left(R_{l,y}^{est}\right)\right] - E\left[\ln\left(R_y^{obs}\right)\right]\right)^2$$

Similarly $\widehat{\gamma_l}$, was calculated as the mean of the variance of the estimated returns on the log scale, calculated over a 5 year window:

$$\widehat{\gamma_l} = \frac{1}{5}\sum_{y=2016}^{2020} Var\left(\ln\left(R_{l,y}^{est}\right)\right)$$

A comparison of R_y^{obs}, $R_{l,y}^{est}$ and R_y^{obs} for the time period 2016 to 2020 are shown in Figure 2 a, with l = 1 and Figure 2 b. with l = 2.

Results. Adjusted returns compared to observed returns (years 2016-2020)

For the 1SW components of AK, KB and KW the observed returns are not well captured by the estimated returns. This discrepancy is mostly driven by large variability in observed 1SW returns for those stock units, which is not captured by the five year averages underpinning the estimated returns. The result of this is a large increase in the variance of the adjusted returns relative to the estimated returns, which successfully captures the observed returns. A similar dynamic is present in the MSW component of the KW stock unit.

For the RP stock unit, the variation in the observed returns is small. This is due to the declared returns being directly observed, and uncertainty in the returns being introduced by integration over the uncertainty in the unreported catch rate only. This results in a substantial increase in the variance when deriving the adjusted returns. Again the adjusted returns successfully capture the observed returns.

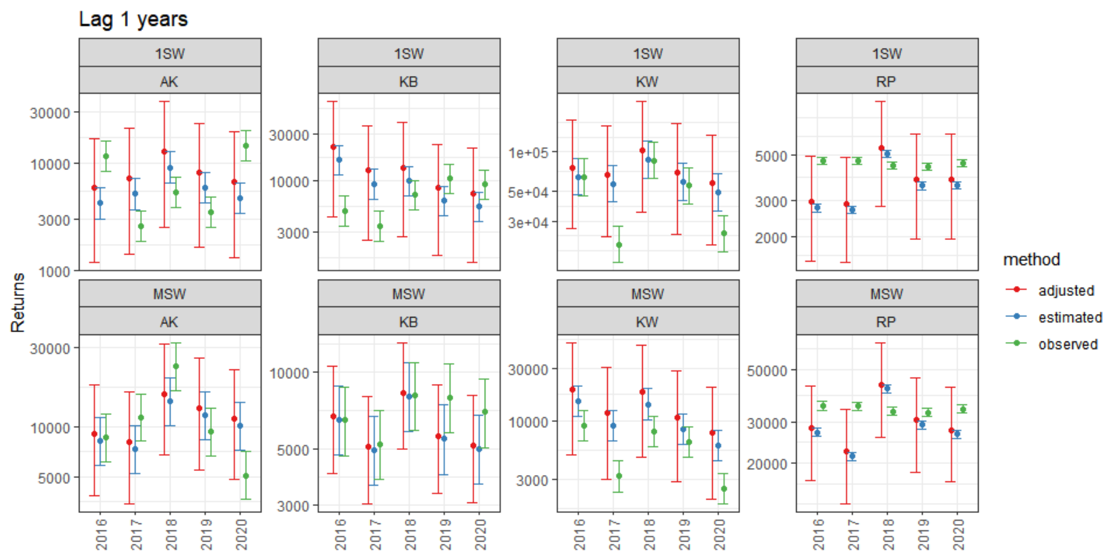

Figure 2 a. Distribution of 'observed' returns based on reported values for catches (declared returns for RP), returns based on estimated values and returns based on estimated values with adjusted variance for the four regional stock units of Russia and 1SW and MSW stocks. Estimated catches and declared returns are based on five years average lagged by 1 year. Points show the mean value, error bars show the 5th and 9th quantiles, y axis on the log scale.

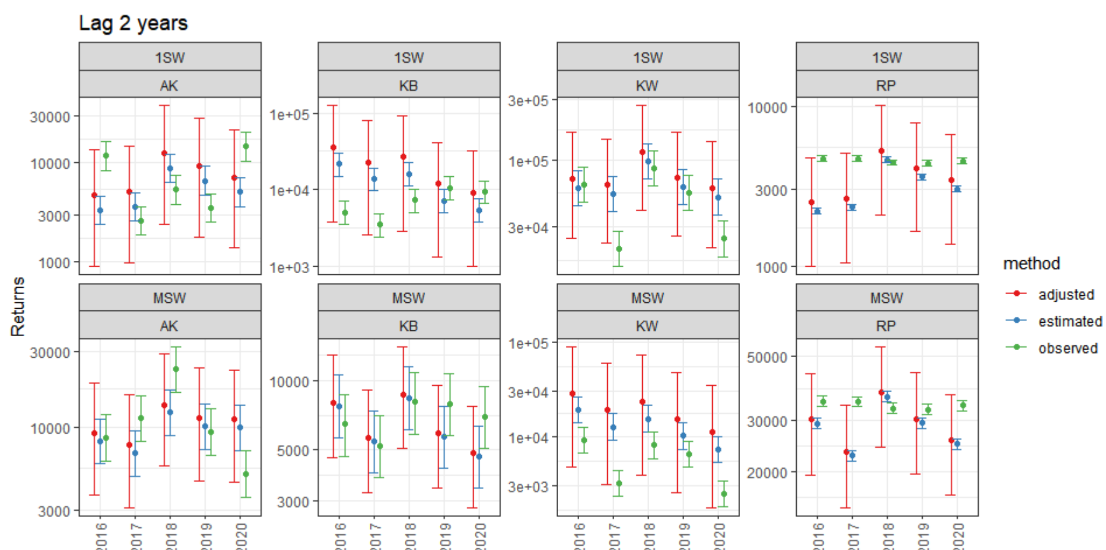

Figure 2 b. Distribution of 'observed' returns based on reported values for catches (declared returns for RP), returns based on estimated values and returns based on estimated values with adjusted variance for the four regional stock units of Russia and 1SW and MSW stocks. Estimated catches and declared returns are based on five years average lagged by 2 years. Points show the mean value, error bars show the 5th and 9th quantiles, y axis on the log scale..

Results. Adjusted prediction of returns, years 2021 and 2022

Figures 3 shows the returns for 2021 and 2022 based on estimated catches (or declared returns for RP) before and after adjusting the variance to account for the additional uncertainty, together with the historic estimates of returns based on reported catch (or declared returns for RP).

As expected from the prior analysis, the increase in the variance is most pronounced for the RP region and where historic returns estimates have high variability. Whilst the uncertainty adjustments are large, this is reflective of genuine additional uncertainty in the returns in the absence of data and represents a conservative approach.

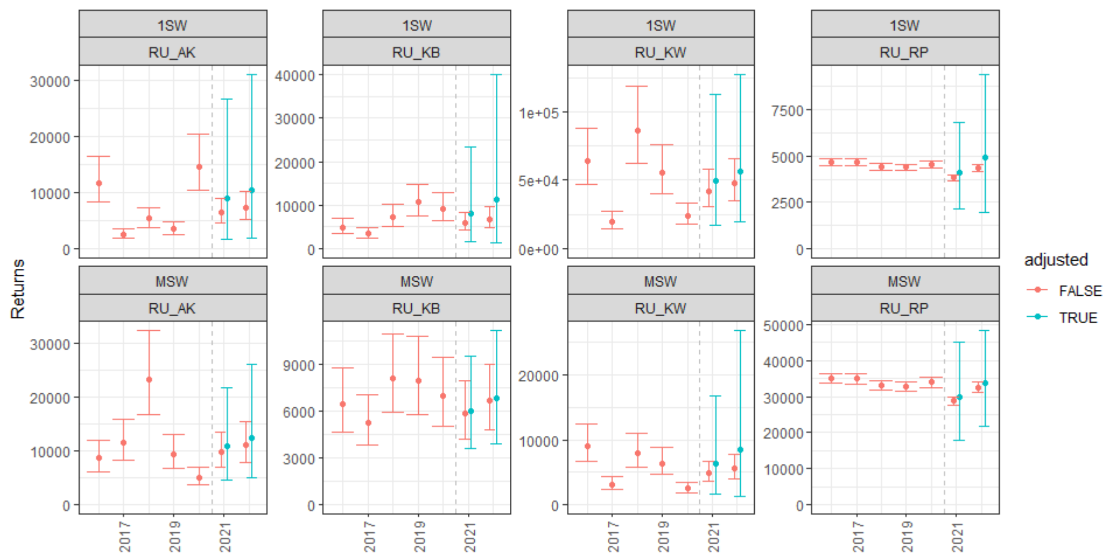

Figure 3. Returns based on reported values (2016 to 2020) and estimated returns with and without adjusted variance (2021, 2022) for the four regional stock units in Russia and the 1SW and MSW stocks. Error bars show the 5th and 95th quantiles

Discussion

The proposed method constitutes one approach for providing the input data needed for the run-reconstruction and PFA models based on the total catches in weights reported by Russia to NASCO. It was developed by the Working Group in 2023 for the purposes of assessing the Russian stocks in 2021 and 2022. However, the approach is based on strong hypotheses and has limitations. If this situation continues, a robust approach to handling this deficiency going forward is desired. The Working Group anticipates exploration of the following issues; some of which could be addressed during the WGNAS benchmark process (BWKSalmon).

- The method used to scale the variance relies on the last five years of available data, and all years in this five-year window have the same weight in the analysis. The choice of a five-year window was made based on the assumption that more recent data would be more representative of the present. Alternative methods could consider data from additional years in the time-series, weighting the influence of each year by recency. Using time-series-based statistical models to capture the influence of previous years while avoiding the strong hypothesis of a simple average could also be investigated.

- Implicit in this approach is the assumption that each stock unit covaries with the total catch in tonnage. Indeed, the approach scales the expected catches in each of the four regions using the same scaling factor (the ratio of the total catches in weight between the last years of data and the predicted years). To address this limitation, alternatives approaches could be developed to stochastically model the split between the four regions (and the same holds for the split between sea-ages within regions). Modelling the split using Multinomial or Multinomial-Dirichlet distributions would allow for stochasticity in the split while ensuring that estimated catches for each stock unit and age class sum to the total reported tonnage of fish at the scale of Russia.

- Any method used to disaggregate the catch in Russia that is based on historic data will become less applicable the more time passes since the data were last updated. Hence, to enhance robustness, an alternative method would be to modify the assessment model by aggregating the four stock units of Russia to a single stock unit. The implications of this approach, given biologic basis for the current split, should be considered.